

Université De Strasbourg

Thèse

Discipline

Aspect Moléculaire et Cellulaire de la Biologie mention Biologie structurale

Benoit Maillot

Fonction et structure  
du complexe entre l'intégrase du VIH-1  
et deux protéines cellulaires (LEDGF et INI1)

Présenté le 09 septembre 2010

pour obtenir le grade de Docteur de l'Université De Strasbourg

**Membres du jury**

Marc Ruff	Directeur de thèse
Jean-François Mouscadet	Rapporteur externe
Patrice Gouet	Rapporteur externe
Stéphane Emiliani	Examineur externe
Jean Cavarelli	Examineur interne
Patrick Schultz	Examineur interne

*« Mentir, baiser, mourir. Il venait d'être défendu d'entreprendre autre chose »*

Louis-Ferdinand Céline.

## Remerciements

*Le travail présenté dans ce mémoire a été réalisé à l'institut de génétique et de biologie moléculaire et cellulaire. J'exprime toute ma gratitude au professeur Dino Moras pour m'avoir accueilli dans son équipe.*

*Je tiens à exprimer mes remerciements à Messieurs Jean-François Mouscadet, Patrice Gouet, Stéphane Emiliani, Jean Cavarelli et Patrick Schultz pour leur patience et pour avoir accepté de juger ce travail.*

*Je remercie Marc Ruff qui a dirigé une grande partie de mon travail. Toujours d'une grande disponibilité, il a été pédagogue et patient.*

*Je remercie l'ANRS (Agence nationale de recherches sur le sida et les hépatites virales) pour leur soutien financier et leur confiance en moi. Je remercie le Sidaction et l'ANR qui ont participé au financement de ce projet.*

*Je remercie le Dr. Yves Mely et le Dr. Patrick Schultz pour m'avoir permis d'accéder aux ressources de leurs laboratoires.*

*Je remercie également :*

*Florence Granger pour son savoir-faire et ses mains expertes dans la manipulation des protéines ; Sylvia Eiler pour la pertinence des remarques et ses connaissances scientifiques ; Fabrice Michel pour son extrême patience et sa suprême pédagogie ; Corine Crucifix pour son savoir-faire et sa pédagogie en microscopie électronique, Karine Pradeau Aubreton ; Aurélie Schaetzel ; Yann Brelivet pour ses conseils éclairés en bio-informatique et sa collégialité ; et bien sûr Nicolas Levy (TIE).*

*Ludovic Richert pour sa patience et sa disponibilité, Pascal Didier pour sa pédagogie, Guy Duportail pour son savoir-faire et sa gentillesse, Julien Godet pour ses précieux conseils en anisotropie et Youri Arntz (M3).*

*Tous les membres du département, particulièrement : Carole Peluso-Iltis (l'as des BPL), Wassim (Le maire) Abdul Rahman, Chris Browning, Jean-Claude Thierry, Valérie Lamour, Yannick Beck ( $\mu$ ), Julie Papillon, Albin (sympathique) Simoni ; et également Benjamin Herquel (pour sa restriction), et Amin Choukrallah (La barre Amin™)...*

*Les services scientifiques communs de l'IGBMC, notamment Raymond Ripp, Pierre Poussin-Courmontagne, Loubna Salim, Edouard Troesch, Matthieu Stierlé, et Didier Busso pour leurs conseils et leurs disponibilités.*

*Le personnel des services communs de l'IGBMC.*

*Je remercie personnellement ma femme, ma fille, et ma famille.*

## Table des matières

Remerciements .....	1
Table des matières .....	3
Tables des abréviations .....	7
Introduction.....	12
Introduction générale.....	12
Le virus de l'immunodéficience humaine (VIH) .....	15
Le cycle viral du VIH.....	17
La phase précoce .....	17
La phase tardive .....	18
Les traitements antirétroviraux .....	21
Rôle de l'intégrase dans le VIH.....	23
Structure.....	23
Fonction.....	27
Le Complexe de PréIntégration (CPI) .....	30
Les inhibiteurs de l'intégrase.....	36
Projet de recherche .....	36
Étude Bioinformatique .....	38
Matériel et méthodes.....	38
La stratégie <i>in silico</i> .....	38
L'environnement bioinformatique .....	38
GScope.....	38
PipeAlign.....	39
Logiciel d'analyse des MACS.....	42
Définition automatique des domaines.....	42
Prédiction des structures secondaires .....	43
KoAnno .....	45
Résultats .....	47
Pourquoi une étude <i>in silico</i> ? .....	47
Le rôle central des alignements multiples de séquences.....	47
Analyse de MACS.....	48
Les domaines : pFAM, ProDom et Morozov.....	48
Les bornes.....	49

Clonage et tests d'expressions .....	54
Matériels et méthodes .....	54
La préparation de bactéries chimiocompétentes .....	54
Transformation.....	55
Purification de plasmide : mini-préparations d'ADN .....	55
Le clonage.....	56
Tests de mini-expression .....	62
Résultats .....	64
Production et purification .....	66
Matériels et méthodes.....	66
La production dans le système bactérien.....	66
Production à l'échelle du litre .....	66
Purification .....	67
Les différentes stratégies de formation de complexe.....	72
Fin purification .....	78
Résultats .....	78
Introduction.....	78
Purification du complexe Intégrase / LEDGF / INI1(173-290) .....	85
Études fonctionnelles.....	92
Matériel et methodes.....	92
Anisotropie de fluorescence .....	92
Résultats .....	98
Intégrase et liaison à l'ADN .....	98
Intégrase et 3' processing.....	102
Études structurales.....	105
Matériels et méthodes.....	105
Introduction.....	105
La stratégie de cristallisation.....	105
La stratégie de microscopie électronique .....	107
Résultats .....	117
Cristallogénèse .....	117
Microscopie électronique moléculaire.....	119
Discussions .....	124
Discussion de l'étude <i>in silico</i> .....	124

Résumé des résultats de l'étude in silico .....	124
Discussion .....	124
Perspectives.....	124
Discussion tests d'expression et production.....	125
Résumé des résultats des tests d'expression.....	125
Discussion .....	125
Perspectives.....	125
Discussion purification .....	126
Résumé des résultats de purification.....	126
Discussion .....	126
Discussion Fonction.....	127
Liaison à l'ADN.....	127
3'processing.....	128
Les autres activités de l'intégrase .....	129
Discussion structure .....	131
Annexes .....	132
Composition des milieux .....	132
Composition de milieu LB.....	132
Composition de milieu LB agar.....	132
Composition de milieu SOB (pour bactéries chimiocompétentes).....	132
Composition de milieu TB (pour bactéries chimiocompétentes).....	133
Composition de milieu ZYP-5052 (pour tests mini-expression).....	133
Electrophorèse sur gel de polyacrylamide en conditions dénaturantes (SDS-PAGE).....	134
Composition du gel de séparation .....	134
Composition du gel de concentration .....	134
Composition du tampon de dénaturation .....	134
Composition du tampon de migration 5x concentré .....	135
Composition de la solution de coloration .....	135
Composition de la solution de décoloration .....	135
Composition de la solution de conservation.....	135
Protocole .....	135
Hybridation des oligonucléotides.....	136
Bibliographie .....	137
Résumé.....	148



## Tables des abréviations

### A

ADDA	Automatic Domain Decomposition Algorithm
ADN	Acide désoxyribonucléique
ADNc	ADN cible
ADNv	ADN viral
ARN	Acide ribonucléique
ARNm	ARN messenger
ARNt	ARN de transfert
ASV	Avian sarcoma virus

### B

$\beta$ MEt	$\beta$ mercaptoethanol
-------------	-------------------------

### C

CA	Capsid protein
CCR5	Chemokine (C-C motif) receptor 5
CPI	Complexe de PréIntégration = PIC Preintegration complex
C-ter	Carboxy-terminal
CXCR4	Chemokine (C-X-C motif) receptor 4
Cy5	Cyanine 5

### D

DIFP	Inhibiteurs classiques de protéases à sérine
DKA	Dérivés des dicéto-acides
DOxxxnm	Densité Optique à XXXnm
dpi	“dot per inch” point par pouce

## E

e.g. "exempli gratia" = "par exemple"

env envelope

## F

FAM Fluorescéine

FDA Food and Drug Administration

## G

gag group-specific antigen

GST Glutathione S-Transferase

## H

HAART Highly Active Antiretroviral Therapy

HMGA1 High mobility group AT-hook 1

HTHi Helix-Turn-Helix inverse

## I

i.e. "id est" = "c'est à dire"

IGBMC Institut de Génétique et de Biologie Moléculaire et Cellulaire

IMAC Immobilized Metal Affinity Chromatography = Chromatographie d'affinité cations divalents immobilisés

IN INtégrase

IN Intégrase

INI1 Intégrase interactor-1

IPTG IsoPropyl  $\beta$ -D-1-ThioGalactopyranoside

## K

Ka	Constante d'association
Kd	Constante de dissociation

## L

LEDGF	Lens epithelium derived growth factor
LMS	Local Maximum conserved Segments
LTR	Long terminal repeat

## M

m /v	masse/volume
MA	Matrix protein
MACS	Multiple Alignment of Complete Sequences
MBP	Maltose Binding Protein
MWCO	Molecular Weight Cut-Off

## N

NC	Nucleocapsid protein
Nef	“(Negative Regulatory Factor”
NLS	Nuclear localization signal
N-ter	Amino-terminal

## P

p/v	poids/ volume
P3C	Protease 3C Precision
PCR	Polymerase Chain Reaction Réaction de Polymérisation en Chaîne.
PDB	Proteine Databank Brookhaven

pGT(FAM)	phospho-guanine-thymine-(fluorescéine)
PML	Protein promyelocytic leukaemia
pol	Polymerase
PR	Protease
PSI-BLAST	Position-Specific Iterative BLAST
BLAST	Basic Local Alignment Search Tool
px	pixel
R	
RMN	Résonance Magnétique Nucléaire
RNase H	Ribonucléase H
RSV	Rous sarcoma virus
RT	Reverse transcriptase
S	
SCOP	Structural Classification of Proteins
SDS	Sodium Dodecyl Sulfate = laurylsulfate de sodium
SDS-PAGE	Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis = électrophorèse en gel de polyacrylamide contenant du laurylsulfate de sodium
SFH1	SNF Five Homologue 1
SIDA	Syndrôme d'immuno-déficience acquise
SIV	Simian immunodeficiency virus
SNF5	Sucrose NonFermentable protein 5
SU	Protéine de surface (gp120)
SW/SNF	SWItch/Sucrose NonFermentable
T	
Tat	TransActivator of Transcription

TE	Tris EDTA
TEV	
TM	Protéine transmembranaire
U	
U3	Unique sequence element at the 3' end of the viral RNA
U5	Unique sequence element at the 5' end of the viral RNA
V	
v/v	volume/ volume
VC	Volume Colonne
Vif	Viral infectivity factor
VIH-1	Virus de l'immunodéficience humaine de type 1
VIH-2	Virus de l'immunodéficience humaine de type 2
Vpr	Viral protein R
Vpu	Viral protein U
vx	voxel

## Introduction

### Introduction générale

Le virus de l'immunodéficience humaine (VIH) est la cause du syndrome de l'immunodéficience acquise (SIDA) (N.I.A.I.D. 2010). Le VIH a été découvert en 1983 (Barré-Sinoussi, Chermann et al. 1983). Vingt-huit ans plus tard, aucun traitement curatif n'est capable d'éliminer définitivement le VIH de l'organisme. Cependant, la compréhension des mécanismes de ce rétrovirus a permis de développer un important panel de médicaments (De Clercq 2009) inhibant les étapes clés de la prolifération et de la réplication du VIH. Ainsi, il est possible de traiter les conséquences (le SIDA) mais pas la cause (le VIH). La meilleure arme à disposition reste encore la prévention. Les risques de transmission du VIH-1 sont relatifs à la voie d'exposition, principalement la transfusion sanguine, l'accouchement, le partage de seringue chez des consommateurs de drogues injectables, les rapports sexuels non protégés. Le VIH est présent dans de nombreux fluides organiques principalement dans le sang, le lait maternel, la cyprine, et le sperme.

Selon le rapport sur l'épidémie mondiale de SIDA (ONUSIDA 2008), depuis le début de l'épidémie, près de 60 millions de personnes ont été infectées par le VIH et 25 millions de personnes sont décédées de causes liées au VIH. En 2008, l'épidémie du VIH est mondiale et l'infection concerne quelque 33,4 millions d'individus (31,1 millions-35,8 millions) (Figure 1). Il y a chaque année 2,7 millions (2,4 millions-3,0 millions) de nouvelles infections et 2 millions (1,7 million-2,4 millions) de décès liés au SIDA. Environ 430000 (240000-610000) enfants sont nés avec le VIH, portant à 2,1 millions (1,2 million-2,9 millions) le nombre total d'enfants de moins de 15 ans vivant avec le VIH. Les jeunes représentent environ 40% de l'ensemble des infections à VIH chez l'adulte (15+) dans le monde. L'Afrique subsaharienne est la région la plus touchée et abrite 67% de l'ensemble des personnes vivant avec le VIH dans le monde et 91% du total des nouvelles infections parmi les enfants.

La phase de primo-infection est symptomatique ou asymptomatique. La primo-infection symptomatique se manifeste chez 50 % des sujets infectés par des signes généraux peu spécifiques (ressemblant à ceux observés lors d'affections virales tel que la mononucléose), et risque souvent de passer inaperçue. Les premiers signes de primo-infection apparaissent en moyenne 20 jours après la contamination et durent de 1 à 3 semaines. Cette phase correspond à une multiplication virale intense et à la dissémination du virus. L'évolution vers la phase

SIDA est plus rapide après une primo-infection symptomatique. Le système immunitaire réagit à l'infection et, en quelques semaines, des anticorps apparaissent dans le sérum, dirigés contre l'ensemble des protéines du VIH, c'est la séroconversion : le sujet infecté devient séropositif.

La deuxième phase est asymptomatique et peut être très prolongée (10 à 15 ans). Le virus se réplique continuellement dans les gâtes lymphoïdes : les lymphocytes TCD4 vont lentement, mais inexorablement, diminuer. La phase clinique correspond au SIDA proprement dit, lorsque le nombre de lymphocytes TCD4 diminue au-delà d'un certain seuil : le syndrome d'immunodéficience apparaît. Les manifestations les plus fréquentes sont les infections opportunistes (virus, bactéries, parasites) et/ou l'apparition de cancers. Dans la majorité des cas, les signes cliniques de déficit immunitaire apparaissent pendant les dix années suivant la séroconversion.

La mortalité due aux conséquences de l'infection du VIH a nettement diminuée dans les régions du monde où l'accès aux médicaments antirétroviraux n'est pas un problème. Cependant, le développement de souches résistantes limite l'efficacité des traitements. Cette résistance peut mener jusqu'à l'échec du traitement et augmenter la mortalité (Hogg, Bangsberg et al. 2006). Le traitement antirétroviral hautement actif (HAART: *Highly Active Antiretroviral Therapy*) est une multi-thérapie destinée à obtenir une diminution maximale de la charge virale. Le mode d'action de la multi-thérapie consiste à combiner différentes classes d'inhibiteurs antirétroviraux ciblant différentes étapes du cycle de réplication du VIH. L'action synergique des inhibiteurs permet de réduire le dosage pour limiter les effets secondaires, et diminuer l'émergence de virus résistant. En effet, un inhibiteur exerce une pression de sélection sur sa cible, favorisant l'apparition de mutation de résistance de la cible. En multipliant les pressions sur différentes cibles, toutes les cibles inhibées doivent muter simultanément pour que le virus entier devienne résistant et puisse se répliquer. La multi-thérapie permet de limiter les possibilités du virus de s'adapter à son milieu. Plus le nombre d'antirétroviraux à disposition est important, plus le nombre de combinaisons de différents inhibiteurs sera important, diminuant ainsi l'apparition de multi-résistance. Six classes d'antirétroviraux sont couramment utilisées pour inhiber la réplication du VIH à différentes étapes du cycle viral. Le développement de nouvelles classes d'antirétroviraux ayant d'autres cibles du mécanisme rétroviral permet d'obtenir d'autres leviers pour inhiber la réplication du VIH, et augmenter le nombre de combinaison d'inhibiteurs et ainsi réduire le nombre de patients en échec thérapeutiques.

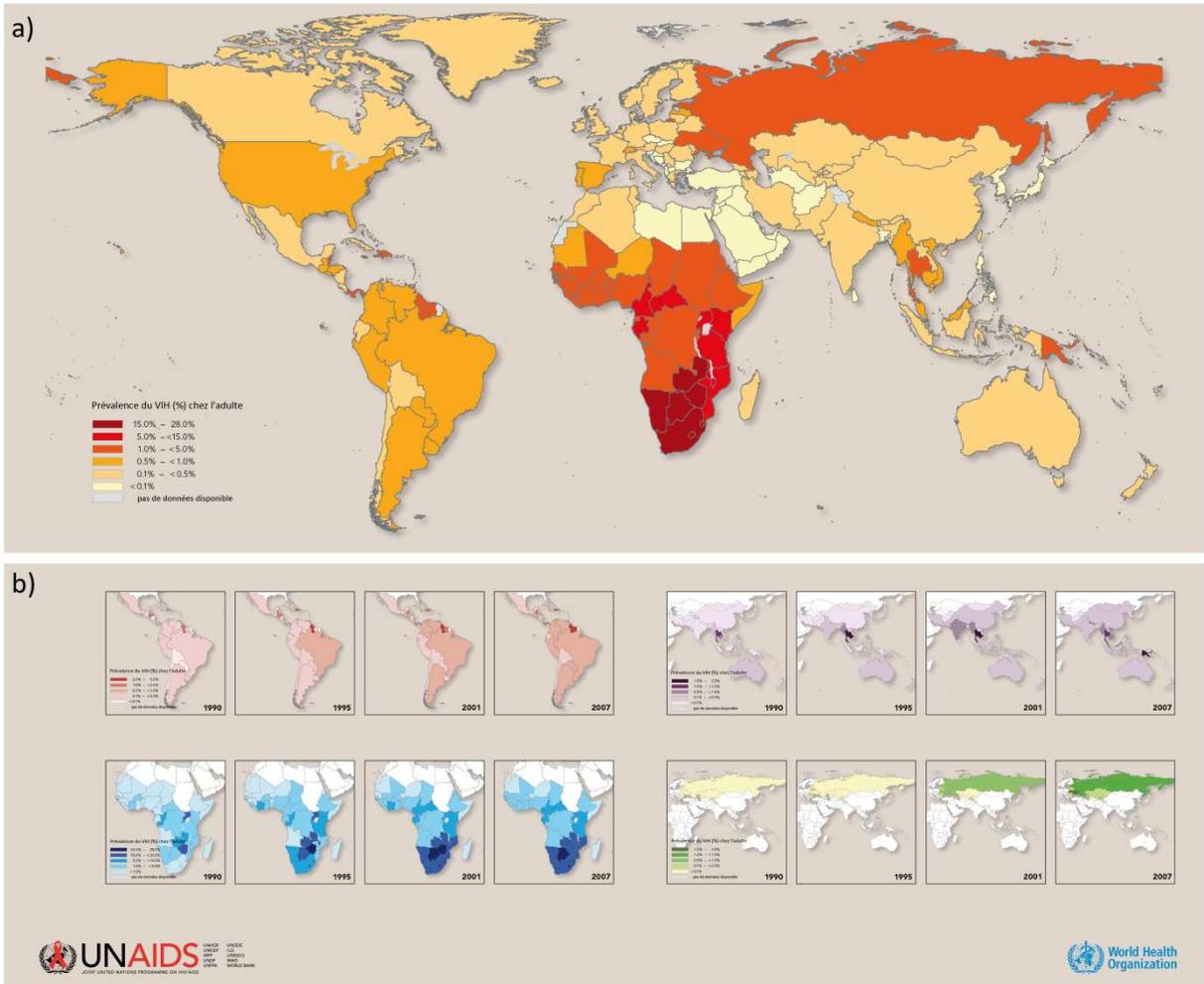


Figure 1 Vue d'ensemble de l'infection à VIH à l'échelle mondiale. a) estimation de la prévalence du VIH (%) chez l'adulte en 2007. b) évolution de la prévalence du VIH (%) chez l'adulte entre 1990 et 2007. (source ONUSIDA)

## Le virus de l'immunodéficience humaine (VIH)

Le virus de l'immunodéficience humaine (VIH) appartient à la famille des Retroviridae, genre des Lentivirus. Les rétrovirus sont des virus à ARN monocaténaire, de polarité positive, infectant les vertébrés. Ils se distinguent notamment par la présence d'une enzyme virale : la transcriptase inverse qui rétro-transcrit leur génome d'ARN en ADN pour être intégré par la suite dans le génome de la cellule. La transcriptase inverse a la particularité de commettre relativement facilement des erreurs, ce qui fait que certains rétrovirus ont une grande variabilité génétique. Les Lentivirus ont pour caractéristiques essentielles d'avoir une longue période d'incubation et d'être cytopathogène. Les Lentivirus ont la particularité de pouvoir se répliquer dans des cellules qui ne se divisent pas.

Il existe plusieurs type de VIH : les VIH de type 1 dont la distribution est mondiale, et les VIH de type 2, retrouvés essentiellement en Afrique de l'ouest. La variabilité génétique très importante du VIH est notamment à l'origine de l'émergence des résistances aux antirétroviraux.

L'aspect de la particule virale est globalement sphérique d'un diamètre variant de 90 à 120 nm (Figure 2). L'enveloppe est composée des restes de la membrane de la cellule infectée et est recouverte de deux types de glycoprotéines : la première est la gp41 qui traverse la membrane (TM ectodomaine), la seconde est la gp120 (SU) qui recouvre la partie de la gp41 qui sort de la membrane. La bicouche lipidique qui provient de la membrane de la cellule hôte contient plusieurs protéines cellulaires (principalement les antigènes d'histocompatibilité, l'actine et l'ubiquitine (Arthur, Bess Jr et al. 1992).

À l'intérieur de l'enveloppe, se trouve une matrice protéique (MA) sur la surface intérieure de la membrane virale, une capsid (CA) qui forme un noyau conique et enfin une nucléocapside (NC) qui abrite deux exemplaires d'un ARN simple brin positif : le génome du VIH. Ce génome est accompagné des trois enzymes indispensables à la réplication et n'ayant pas d'équivalent cellulaire (la transcriptase inverse (RT), l'intégrase (IN), et la protéase virale (PR)). Elles sont spécifiques aux rétrovirus et une cinquantaine de molécules de chaque enzyme sont dans chaque virion. Les virions empaquettent également les protéines accessoires : Nef, Vif et Vpr. Trois protéines accessoires additionnelles ayant des rôles durant l'infection de la cellule hôte, Rev, Tat et Vpu, ne semblent pas être empaquetées

Le génome du VIH est composé de neuf gènes. Les trois principaux sont gag, pol et env, qui définissent la structure du virus et sont communs à tous les rétrovirus. Les six autres gènes sont tat, rev, nef, vif, vpr et vpu (ou vpx pour le VIH-2), qui codent des protéines régulatrices. Toutes ces protéines virales ont un rôle plus ou moins important dans le cycle de réplication du VIH-1.

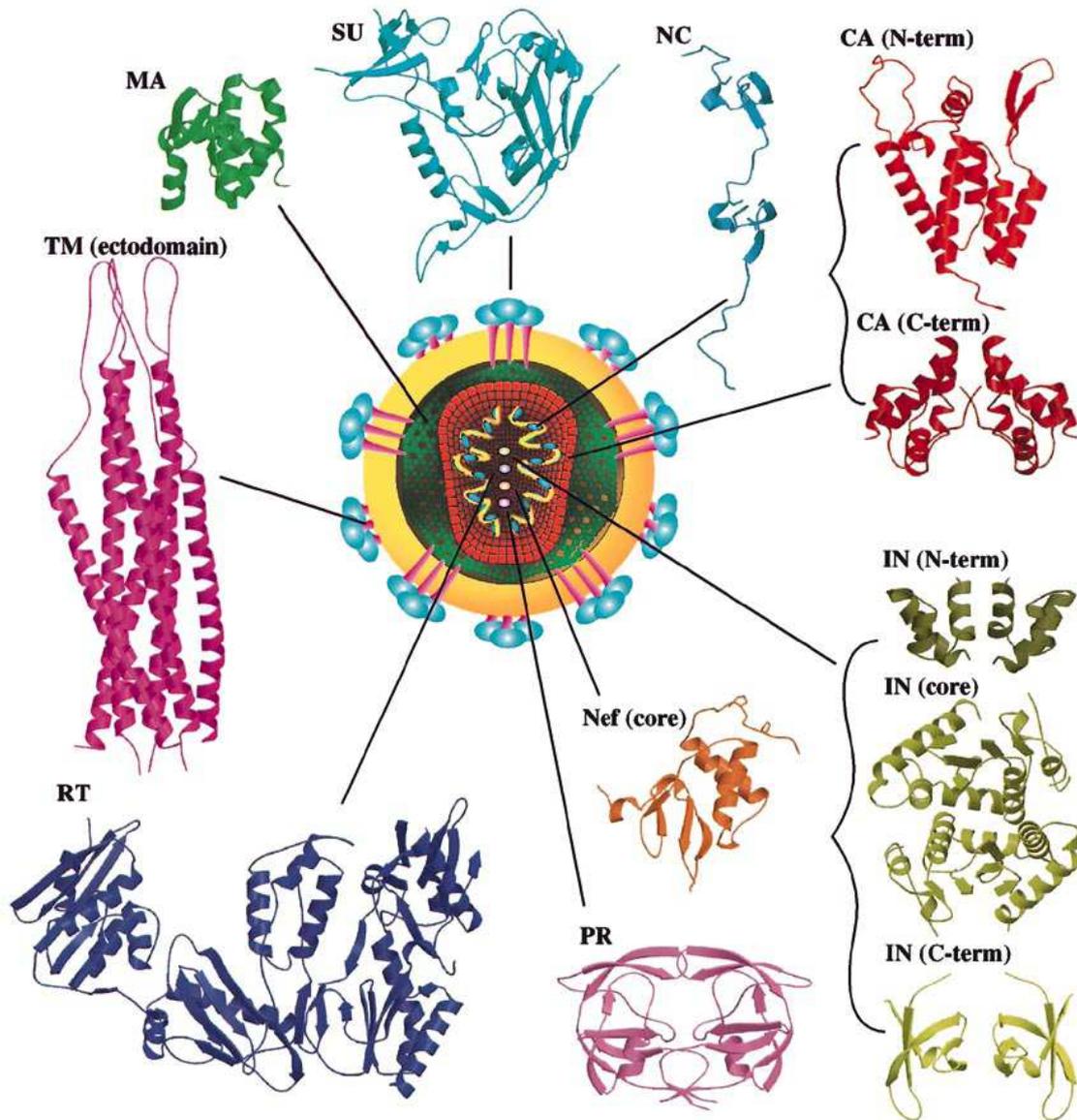


Figure 2 Schéma d'un virion de VIH-1 mature et structures des protéines virales. Les structures des protéines sont à la même échelle (Turner and Summers 1999). La bicouche lipidique (en jaune) provient de la membrane de la cellule hôte. La glycoprotéine exposée (gp120 (SU) en cyan) interagit avec la protéine transmembranaire (gp41 (TM) en violet). Dans un virion il y a approximativement 2000 copies de la protéine de matrice (MA en vert). Au centre, la capsid a la forme d'un noyau conique en 2000 copies de la protéine de capsid (CA, p24 en rouge). Les deux copies ARN simple brin du génome viral (serpentin en jaune) sont stabilisées par la protéine de nucléocapsid (NC, p7 en bleu) en formant un complexe ribonucléoprotéique. Les trois enzymes virales : la protéase virale (PR en rose), la transcriptase inverse (RT en pourpre) et l'intégrase (IN en vert pale) sont essentielles. Les protéines Nef (orange), Vif et Vpr (non montrés) sont dites accessoires.

## Le cycle viral du VIH

Le cycle viral est divisé en deux phases : la phase précoce et la phase tardive (Figure 3).

### La phase précoce

#### *L'attachement et la fusion*

Les cellules cibles du VIH sont celles présentant des récepteurs CD4 à leur surface. Ainsi, les lymphocytes T CD4+, les macrophages, les cellules dendritiques et les cellules microgliales cérébrales peuvent être infectées par le VIH. Les particules virales du VIH-1 se lient spécifiquement aux cellules exposant à leur surface le récepteur CD4 dont la fonction normale est la reconnaissance de l'immunité. Par l'intermédiaire d'interactions spécifiques entre la glycoprotéine virale SU (gp120) et le domaine immunoglobuline de CD4, le virion se lie à la cellule. Ces interactions sont suffisantes pour l'attachement mais pas pour l'infection. À la différence d'autres rétrovirus, les lentivirus de primate exigent des protéines additionnelles de la surface cellulaire pour favoriser la fusion des membranes virales et cellulaires. Pour HIV-1, la fusion de la membrane est déclenchée par plusieurs récepteurs notamment CXCR4 et CCR5 (Doranz, Rucker et al. 1996; Feng, Broder et al. 1996; Clapham and Weiss 1997; Moore 1997).

#### *La transcription inverse*

La fusion de la membrane est suivie de la décapsidation encore mal comprise. S'en suit l'étape de transcription inverse qui est catalysée dans le cytosol par la transcriptase inverse. Cette étape consiste à transcrire un ARN simple brin (le génome viral) en ADN double brin (ADN proviral). Ce processus a été largement décrit dans la littérature (par exemple : (Goff 1990; Whitcomb and Hughes 1992; Katz and Skalka 1994; Telesnitsky and Goff 1997). La protéine accessoire Vif semble être importante pendant un ou plusieurs de ces événements, peut-être en facilitant les étapes initiales de la transcription inverse. La synthèse d'ADN dépendante de la transcriptase inverse est lancée par l'attache d'une amorce cellulaire de l'ARNt<sup>Lys</sup> et dépend également de la protéine virale de la NC (nucléocapside).

## *Le transport nucléaire et l'intégration*

Une fois que synthétisé, l'ADN proviral est transporté dans le noyau en tant que complexe de préintégration (CPI). La réaction de 3' *processing* s'effectue dans le cytoplasme. Le CPI inclue bon nombre de protéines cellulaires et virales dont, pour la plupart, le rôle au sein du CPI est toujours obscur. Rapidement, le CPI se constitue aussi bien avec des protéines virales (protéine de la matrice (Carteau, Batson et al. 1997), transcriptase inverse (Hehl, Joshi et al. 2004), et Vpr), qu'avec des protéines cellulaires (LEDGF, INI1, HMG-I, transportine,...) (Van Maele, Busschots et al. 2006). Les protéines virales de la capsid ne semblent pas faire partie du CPI, bien qu'elles contribuent à la structure d'autres complexes rétroviraux de préintégration (Bowerman, Brown et al. 1989). La localisation nucléaire du CPI est aidée par la protéine accessoire Vpr (Aida and Matsuda 2009) qui ne contient pas de signal de localisation nucléaire mais semble jouer un rôle dans l'importation du complexe de préintégration par la machinerie cellulaire. Vpr perturbe également le contrôle du cycle cellulaire en arrêtant la croissance des cellules infectées dans la phase G2 (Jowett, Planelles et al. 1995; Re, Braaten et al. 1995; Rogel, Wu et al. 1995). Après transport actif au noyau, l'ADN proviral est intégré dans le génome de la cellule hôte par la réaction de transfert de brin catalysée par l'intégrase.

## **La phase tardive**

### *Synthèse des protéines virales*

La phase tardive du cycle de réplication commence par la synthèse des transcrits viraux épissés et non épissés, qui sont transportées hors du noyau pour la traduction. Initialement, des ARN messagers codant pour les protéines de régulation Tat, Rev et Nef sont synthétisés. La protéine Tat est un activateur transcriptionnel essentiel liant des tiges-boucle d'ARN naissant et recrute des protéines cellulaires (cycline T, Cdk9, TAK). La cycline T se lie directement à la protéine Tat, augmentant son affinité et changeant sa spécificité pour l'ARN (Wei, Ghosh et al. 1995). Puis Cdk9 stimule la transcription et l'élongation par phosphorylation du complexe de transcription de l'ARN polymérase II (Herrmann and Rice 1995; Rice, Supko et al. 1995; Reines, Conaway et al. 1996; Wei, Garber et al. 1998). Habituellement, les ARNm cellulaires non épissés restent dans le noyau où ils sont traités ou dégradés. Dans le cas des ARNm du VIH entiers ou épissés contenant des introns fonctionnel, ceux-ci sont envoyés vers le cytoplasme par l'intermédiaire de la protéine accessoire Rev pour

la synthèse de Gag et de Gag-Pol. La protéine Rev se lie à région de l'ARN du gène *env* appelé *rre* (Rev Reponse Element) contenant un signal d'export nucléaire, recrute l'exportin-1 (XPO) et (Ohno, Fornerod et al. 1998) et le facteur d'export nucléaire Ran (Pollard and Malim 1998). Ce complexe est alors transporté à travers le pore nucléaire vers le cytosol. Le complexe se dissocie, et la protéine Rev est redirigé vers le noyau grâce à son signal nucléaire de localisation (NLS) (Malim, Tiley et al. 1990; Emerman and Malim 1998). De cette manière, la protéine Rev fonctionne comme commutateur entre la synthèse des ARNm fortement épissés précoces (Tat, Rev et Nef), tardives (ENV, Vpu, Vif et Vpr) et des ARNm non épissés (Gag et Gag-Pol). La polyprotéine gp160 (précurseur d'ENV) est synthétisée dans le réticulum endoplasmique.

### *Protéolyse virale et assemblage*

La protéine ENV subit des modifications post-traductionnelles dans le réticulum endoplasmique et dans l'appareil de Golgi puis est clivée en gp120 et gp41. Les deux protéines gp120-gp41 forment un trimère d'hétérodimère et sont transportées vers la membrane cellulaire pour l'assemblage du virus. Les protéines ENV et CD4 sont toutes les deux synthétisées dans le réticulum endoplasmique, et la liaison prématurée de CD4 à ENV dans l'ER peut empêcher la translocation de ENV vers la membrane cellulaire ou la formation du complexe fonctionnel gp120-gp41 (Hoxie, Alpers et al. 1986). Pour bloquer ces interactions, la protéine accessoire virale Vpu lie la protéine CD4 pour la transporter hors du réticulum endoplasmique pour sa dégradation (Crise, Buonocore et al. 1990; Margottin, Bour et al. 1998; Schubert, Antón et al. 1998). De même, les protéines CD4 lié à la protéine accessoire Nef sont dégradées de manière endosomale. Pour le bourgeonnement, la protéine Nef stimule la formation de puits recouverts de clathrines (Foti, Mangasarian et al. 1997; Greenberg, Bronson et al. 1997; Le Gall, Erdtmann et al. 1998; Piguet, Chen et al. 1998). La régulation de la quantité de CD4 à la surface des cellules infectées peut également empêcher une réaction immunitaire. La polyprotéine Gag est synthétisée à partir d'ARNm non épissé. Un déphasage du cadre de lecture permet la production d'un faible pourcentage de précurseur Gag-Pol. Le domaine N-ter de la matrice (MA) de Gag lie directement la membrane cellulaire et interagit avec la queue cytoplasmique de gp41 (Freed, Englund et al. 1995).

## Bourgeoisement et maturation

Approximativement 1200 à 2000 copies de Gag bourgeonnent pour former une particule immature, qui encapside deux copies du génome viral à ARN simple brin positif. À la suite du bourgeoisement, les polyprotéines sont clivées par la protéase virale (PR) pour produire les enzymes indépendantes (RT, IN, et PR) et les protéines structurales (MA, CA, et NC). Les protéines structurales se réarrangent pour finaliser la particule infectieuse et terminer le processus de maturation. D'autres facteurs peuvent également être importants pour l'infectiosité. Par exemple, le VIH-1 exige l'empaquetage de la cyclophiline A, contrairement au VIH-2 et à la plupart des autres lentivirus de primate (Franke, Yuan et al. 1994). En outre, la protéine virale Vif est nécessaire à la production des virions infectieux pour certaines variétés cellulaires (Sova and Volsky 1993).

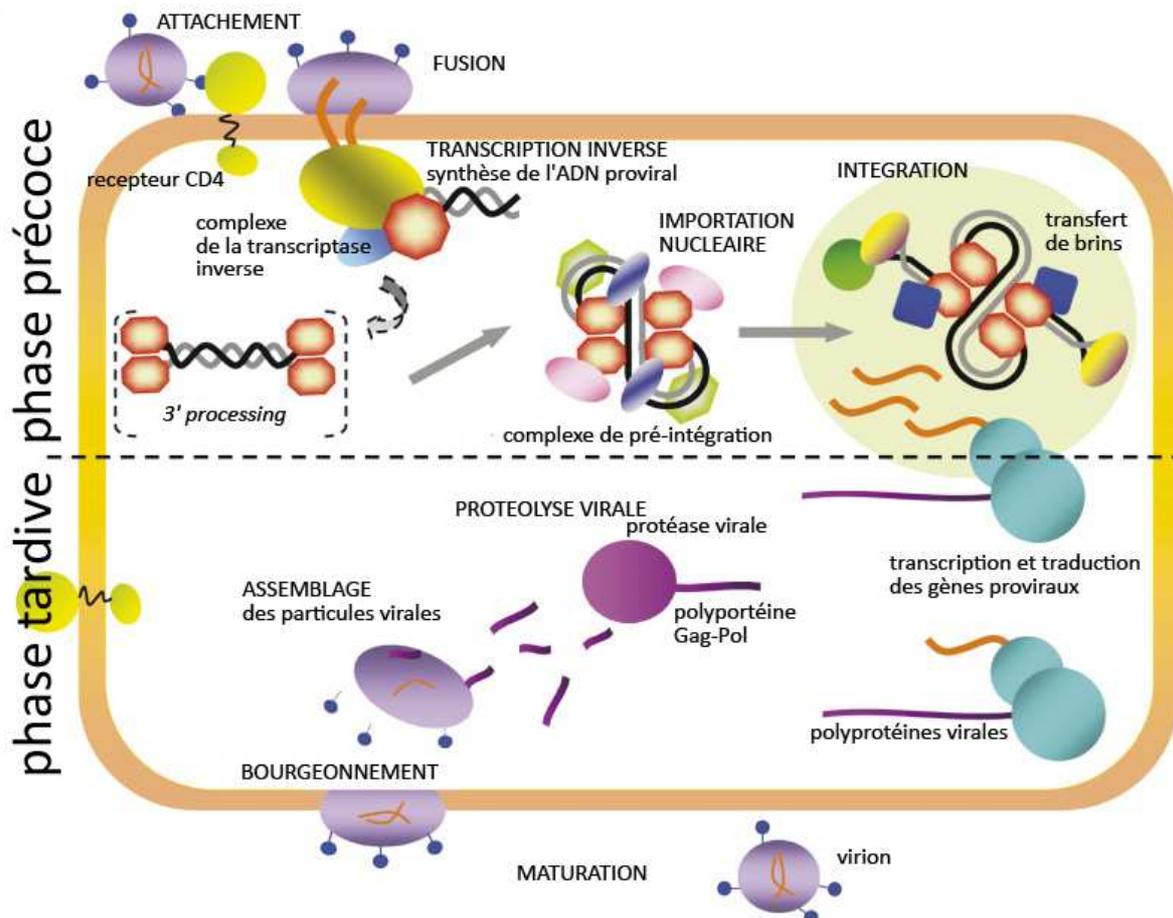


Figure 3 Cycle de réplication du VIH-1. (Al-Mawsawi and Neamati 2007) La réplication virale peut être divisée en deux principales étapes (précoce et tardive). La phase précoce inclut l'attachement, la fusion, la transcription inverse, le transfert nucléaire et l'intégration. La phase tardive commence avec la transcription/traduction des gènes proviraux par la machinerie cellulaire. Après la protéolyse virale, les particules virales s'assemblent, bourgeonnent et finissent leur maturation. Les virions produits sont capables d'infecter de nouvelles cellules immunitaires.

## Les traitements antirétroviraux

Depuis 1996, l'importance de la multi-thérapie pour contrer l'évolution du VIH est largement admise (De Clercq 2009). La combinaison de trois (ou plus) d'antirétroviraux permet d'obtenir une synergie entre différents inhibiteurs qui agissent sur différentes cibles moléculaires, d'abaisser les dosages pour réduire les effets secondaires toxiques, enfin, de diminuer la probabilité du développement de résistance. L'important est d'avoir à disposition un nombre suffisant d'inhibiteurs pour pouvoir obtenir un nombre très élevé de combinaisons possibles (Figure 4). Différentes formulations permettent de combiner, en une seule pilule plusieurs antirétroviraux, et ainsi faciliter la vie des malades. Dans le traitement du SIDA, 25 inhibiteurs différents (Figure 5) ont été approuvée par la FDA (Food and Drug Administration). Différentes classes d'inhibiteurs ciblent différentes étapes de la réplication virale. Les 13 inhibiteurs de la transcription inverse (RT) sont répartis en deux sous classes nucléosidiques et non nucléosidique. Les 9 inhibiteurs de la maturation ciblent la protéase virale (PR). L'inhibiteur de la fusion interagit avec la protéine virale gp41. L'inhibiteur de l'entrée de la particule virale dans la cellule bloque l'interaction entre la protéine virale gp120 et son co-récepteur cellulaire (CCR5 ou CXCR4). Enfin, un inhibiteur de l'intégrase (IN) cible l'étape de transfert de brin. Des trois enzymes virales (RT, PR, et IN) indispensables à la réplication virale, 22 inhibiteurs approuvés ciblent la RT ou la PR. L'intégrase représente aussi une cible alternative de choix pour le développement d'agents thérapeutiques sélectifs car il n'existe aucun analogue cellulaire direct (Anthony 2004).

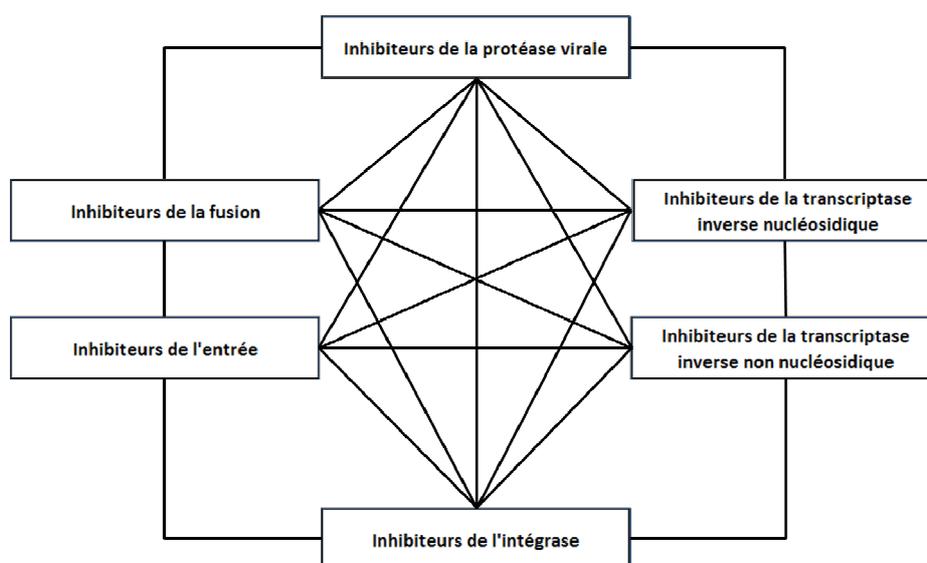


Figure 4 Rosace des combinaisons d'antirétroviraux dans le cadre d'une multi-thérapie.

	année	Nom générique	Fabriquant
<b>Inhibiteurs de la transcriptase inverse nucléosidique</b>			
	1987	zidovudine, azidothymidine, AZT, ZDV	GlaxoSmithKline
	1991	didanosine, dideoxyinosine, ddi	Bristol-Myers Squibb
	1992	zalcitabine, dideoxycytidine, ddC	Roche Pharmaceuticals
	1994	stavudine, d4T	Bristol-Myers Squibb
	1995	lamivudine, 3TC	GlaxoSmithKline
	1997	lamivudine + zidovudine	GlaxoSmithKline
	1998	abacavir sulfate, ABC	GlaxoSmithKline
	2000	abacavir + lamivudine + zidovudine	GlaxoSmithKline
	2000	enteric coated didanosine, ddi EC	Bristol-Myers Squibb
	2001	tenofovir disoproxil fumarate, TDF	Gilead Sciences
	2003	emtricitabine, FTC	Gilead Sciences
	2004	abacavir + lamivudine	GlaxoSmithKline
	2004	emtricitabine + tenofovir disoproxil fumarate	Gilead Sciences
<b>Inhibiteurs de la transcriptase inverse non nucléosidique</b>			
	1996	nevirapine, NVP	Boehringer Ingelheim
	1997	delavirdine, DLV	Pfizer
	1998	efavirenz, EFV	Bristol-Myers Squibb
	2008	etravirine	Tibotec Therapeutics
<b>Inhibiteurs de la protéase virale</b>			
	1995	saquinavir mesylate, SQV	Roche Pharmaceuticals
	1996	ritonavir, RTV	Abbott Laboratories
	1996	indinavir, IDV	Merck
	1997	nelfinavir mesylate, NFV	Pfizer
	1997	saquinavir (retiré)	Roche Pharmaceuticals
	1999	amprenavir, APV	GlaxoSmithKline
	2000	lopinavir + ritonavir, LPV/RTV	Abbott Laboratories
	2003	atazanavir sulfate, ATV	Bristol-Myers Squibb
	2003	fosamprenavir calcium, FOS-APV	GlaxoSmithKline
	2005	triplanavir, TPV	Boehringer Ingelheim
	2006	darunavir	Tibotec Therapeutics
<b>Inhibiteurs de la fusion</b>			
	2003	enfuvirtide, T-20	Roche Pharmaceuticals & Trimeris
<b>Inhibiteurs de l'entrée</b>			
	2007	maraviroc	Pfizer
<b>Inhibiteurs de l'intégrase</b>			
	2007	raltegravir	Merck

Figure 5 : Thérapie anti-SIDA approuvé. D'après (Marchand, Maddali et al. 2009)

## Rôle de l'intégrase dans le VIH

L'intégrase est une protéine codée par l'extrémité du gène pol. Elle est produite dans le contexte du précurseur polypeptidique Gag-Pol dont elle est clivée par la PR. L'IN intervient dans plusieurs étapes du cycle de réplication du virus : intégration de l'ADN viral dans l'ADN hôte (LaFemina, Schneider et al. 1992; Wiskerchen and Muesing 1995), transport dans le noyau du complexe de préintégration (CPI) (Gallay, Hope et al. 1997; Bouyac-Bertoia, Dvorin et al. 2001), coopération avec la transcriptase inverse (Wu, Liu et al. 1999), et production de particules virales (Bukovsky and Gottlinger 1996).

### Structure

L'intégrase est longue de 288 acides aminés et comporte trois domaines indépendants (Figure 6) : un domaine central (50-212) responsable de l'activité catalytique, un domaine N-terminal (1-49) qui est nécessaire à l'oligomérisation de l'intégrase, enfin un domaine C-terminal (213-288) qui possède par lui-même une activité de fixation à l'ADN de caractère non spécifique.

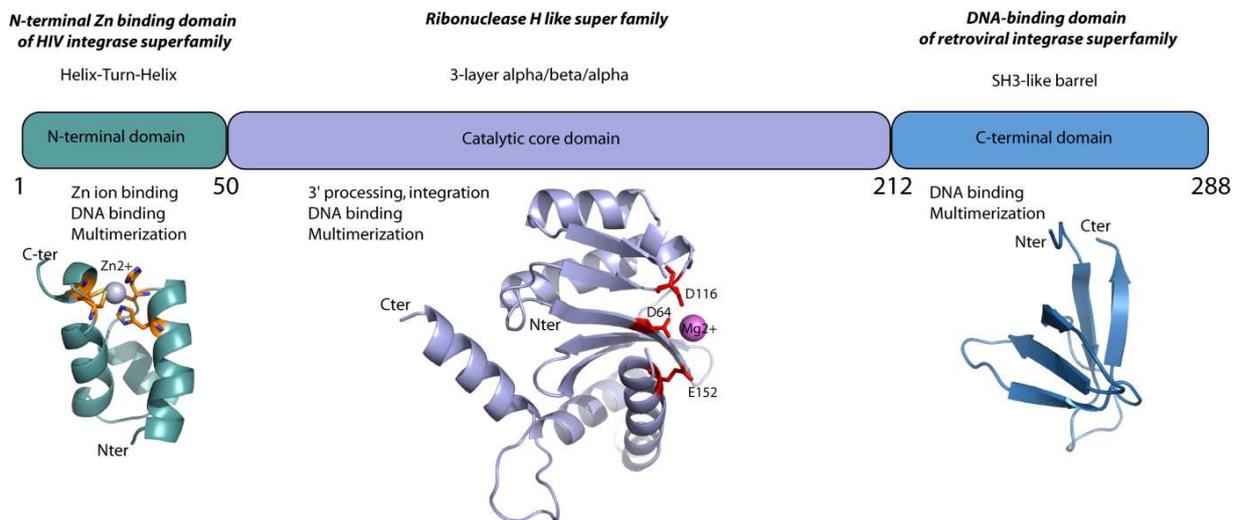


Figure 6 Organisation structurale de l'intégrase du VIH-1.

## Les structures connues des intégrases

Aucune structure complète d'intégrase (1-288) du VIH-1, ou d'un de ses complexes avec l'ADN n'a été résolue. En effet, la faible solubilité et la flexibilité inter-domaines sont un frein à la résolution de la structure. Cependant, plusieurs structures de domaines isolés ou bien encore de deux domaines consécutifs ont été élucidées (Figure 7).

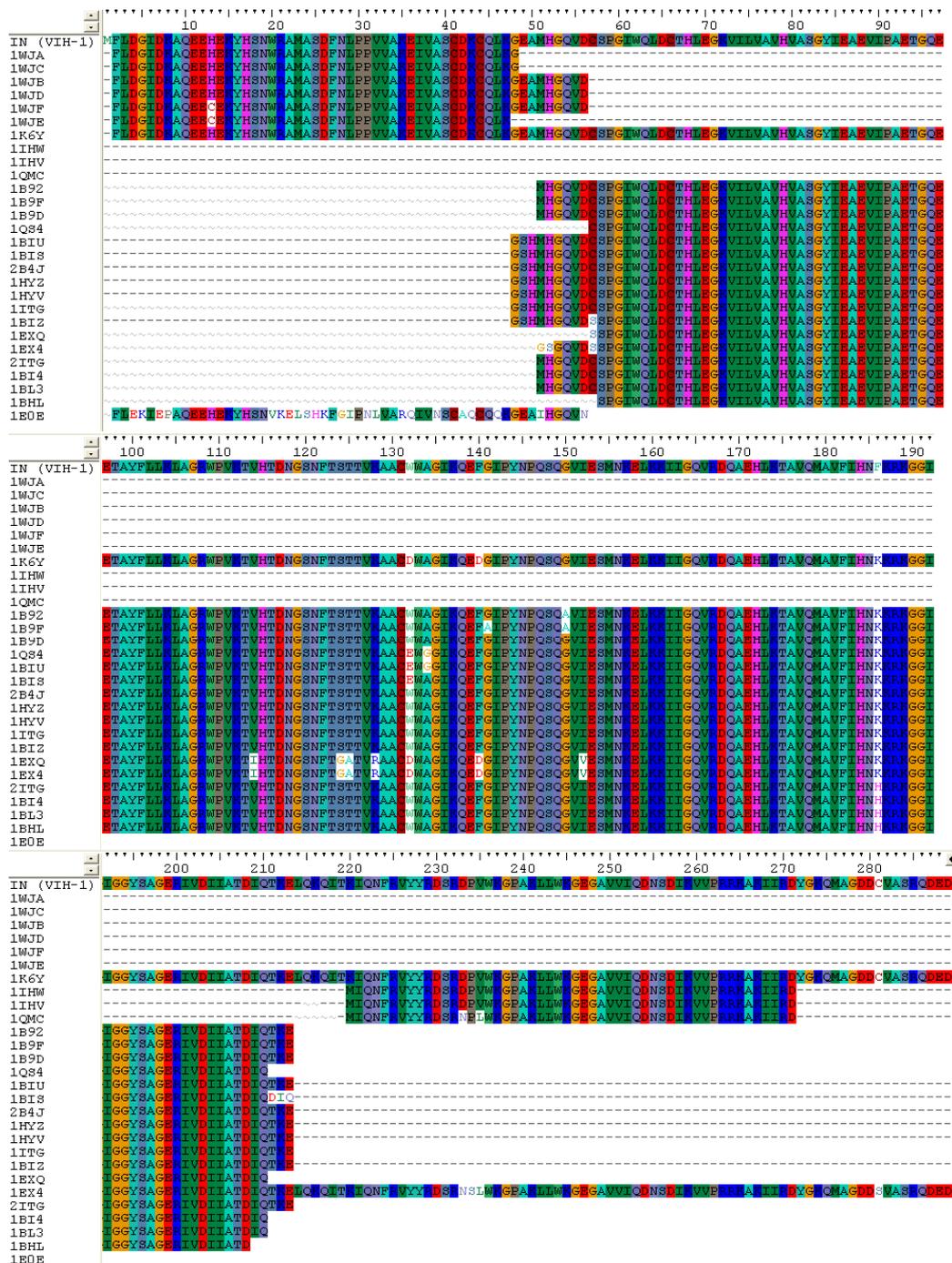


Figure 7 Alignement multiple des structures connues de l'intégrase du VIH-1. La première ligne de l'alignement montre la séquence pleine longueur sauvage. La dernière ligne montre la séquence N-ter de l'intégrase du VIH-2.

## Le domaine N-terminal

Le domaine N-terminal (N-ter) (1-49) comporte un motif HHCC (H12, H16, C40 et C43) analogue à un doigt à zinc permet de favoriser le processus de multimérisation de la protéine entière nécessaire à l'activité d'intégration. La fixation du zinc par le motif HHCC induit le repliement du domaine N-terminal (Zheng, Jenkins et al. 1996). Le domaine est structurée en quatre hélices  $\alpha$ , les hélices  $\alpha_2$  et  $\alpha_3$  formant un motif hélice-tour-hélice, stabilisées par la coordination d'un ion  $Zn^{2+}$  et par des interactions hydrophobes (Burke, Sanyal et al. 1992). Le domaine N-ter interagit avec l'ADN mais ne fixe pas l'ADN *in vitro*. Il est strictement nécessaire au processus de 3' *processing* et de transfert de brins.

## Le domaine Cœur Catalytique

Le domaine cœur catalytique de l'intégrase (50-212) contient le site actif constitué par une triade catalytique (D64, D116, E152) caractéristique des polynucléotididyl transférase. Les trois acides aminés de ce motif D-X(38-58)-D-X(35)-E sont strictement requis pour l'activité catalytique de la protéine. Les résidus importants pour l'interaction spécifique avec l'ADN viral sont également localisés dans le cœur catalytique (Q148, Y143, K156, K159). La mutation de deux d'entre eux, les K156 et K159, abolit l'activité spécifique. Par ailleurs, ce domaine permet le positionnement correct de la protéine sur l'ADN cible (Harper, Skinner et al. 2001). Une étude systématique de mutants dans le corps catalytique a permis d'identifier une mutation (F185K) qui augmente considérablement sa solubilité (Jenkins, Hickman et al. 1995; Jenkins, Engelman et al. 1996). Ce mutant F185K a permis de résoudre plusieurs structures dans diverses conditions physicochimiques (+/- cacodylate, +/- magnésium) (Dyda, Hickman et al. 1994; Goldgur, Dyda et al. 1998; Maignan, Guilloteau et al. 1998; Greenwald, Le et al. 1999). Un motif hélice tour hélice inversé (162-178) intervient dans la formation des dimères et tétramères, ainsi que dans la reconnaissance des ADN viraux et cellulaires (Femandjian, Maroun et al. 2001). Des structures ont également montré l'importance de la formation de tétramères d'intégrases pour la fonction d'intégration (Hare, Di et al. 2009). La structure du domaine catalytique se présente sous forme d'un feuillet  $\beta$  à cinq brins dont le second est antiparallèle aux autres et de six hélices  $\alpha$  (de  $\alpha_1$  à  $\alpha_6$ ). Ce repliement présente des similarités à celui de la RNase H. La boucle (140-149) flexible permet l'orientation du résidu E152 en direction des résidus D64 et D116. La présence des mutations G140A/G149A rigidifie la boucle qui entraîne une diminution de l'activité de l'enzyme sans affecter

l'interaction avec l'ADN. La flexibilité de la boucle est cruciale pour l'activité. En effet, l'apparition d'un contact entre le résidu G148 et l'extrémité de l'ADN viral suite au 3'-*processing* (Johnson, Sayer et al. 2006).

#### Le domaine C-terminal

Le domaine C-ter (212-288) est le domaine le moins conservé entre les différentes intégrases rétrovirales. Il interagit de façon non spécifique avec l'ADN et aurait un rôle stabilisateur des interactions intégrase-ADN. Il est constitué de 5 brins  $\beta$  antiparallèles formant une structure en tonneau similaire au domaine SH3 (*SH3-like barrel*). Les domaines SH3 (Src-Homology 3) sont présents dans de nombreuses protéines qui sont impliquées dans la transduction des signaux cellulaires mais aussi dans des protéines qui se lient à l'ADN. Sa structure a été obtenue par RMN (Lodi, Ernst et al. 1995) et montre que le domaine C-terminal est dimérique et présente un sillon, composé de nombreux résidus chargés. De par sa taille, ce sillon pourrait fixer une hélice d'ADN double brin. De plus, une mutation de la K264 entraîne une baisse importante de la liaison à l'ADN, suggérant ainsi que ce résidu est impliqué dans l'interaction ADN-intégrase (Lutzke, Vink et al. 1994). Sous forme dimérique, les domaines C-terminaux s'orientent de façon parallèle l'un par rapport à l'autre et sont également impliqués dans la multimérisation de la protéine (Lutzke and Plasterk 1998).

#### Structures à deux domaines

La structure du corps catalytique et du domaine C-terminal a été déterminée sur un mutant où 5 résidus ont été mutés (dont la mutation F185K) (Chen, Krucinski et al. 2000) ainsi que la structure du domaine N-ter plus le corps catalytique comprenant trois mutations dont toujours F185K (Wang, Ling et al. 2001). Toutes ces structures montrent une même organisation du corps catalytique en dimère, mais des positions pour les domaines N-ter et C-ter différentes (Jaskolski, Alexandratos et al. 2009) qui montre la flexibilité entre les différents domaines de l'intégrase. Ces structures isolées ne permettent pas de comprendre l'agencement des différents domaines ainsi que de définir l'unité fonctionnelle. De plus, l'espacement des sites actifs (30Å) n'est pas compatible avec la distance entre les sites d'insertions des deux brins de l'ADN cible (15Å) (Craigie 2001). Par ailleurs la mutation F185K est située à l'interface dimérique observée dans les structures et participe activement à la cohésion de celle-ci. Enfin une structure d'un mutant F185H a été résolue, elle est similaire à celle du mutant F185K. Il

est démontré par comparaison avec la structure du corps catalytique d'ASV que ces mutants ne peuvent être fonctionnels de par la conformation de leur site actif (Bujacz, Alexandratos et al. 1996)

### Structures apparentées

La structure du corps catalytique de l'intégrase d'un autre virus (Avian Sarcoma Virus), présentant des homologies fortes de séquences et de structure, ont été résolues (Bujacz, Jaskolski et al. 1995; Bujacz, Jaskolski et al. 1996) en présence d'un inhibiteur de l'intégrase du VIH (Lubkowski, Yang et al. 1998). La structure cristallographique de l'intégrase complète du Spumavirus en complexe avec de l'ADN viral à récemment été résolue (Hare, Gupta et al. 2010). L'organisation des domaines de l'intégrase du Spumavirus est différente de celle de HIV-1 décrite par microscopie électronique (Michel et al., 2009), La différence provient probablement du fait que l'intégrase du Spumavirus n'utilise pas de co-facteurs cellulaires comme LEDGF pour la stabilisation d'un tetramère.

### Fonction

L'intégrase du VIH-1 est une transposase (MeSH: D08.811.739.500.667.500) qui possède plusieurs activités intrinsèques : la liaison à l'ADN, le 3' *processing* et le transfert de brin (Lewinski and Bushman 2005).

#### *La liaison à l'ADN*

Dans un premier temps, l'IN se fixe sur une séquence spécifique (5'-GCAGT3') située aux deux extrémités (U5 et U3) de l'ADN proviral (ou LTR pour Long Terminal Repeat). En présence d'ADN viral on observe la formation d'un dimère à chaque extrémité (Gao, Butler et al. 2001). L'IN lie l'ADN avec deux sites fonctionnellement distincts, un spécifique pour les extrémités de l'ADN viral et une deuxième non spécifique pour l'ADN cible génomique. Une interaction entre les sites allostériques de liaison à l'ADN peut empêcher un site viral inoccupé de se lier à une extrémité de l'ADN viral lorsque le site cible liaison à l'ADN est rempli (Drelich, Haenggi et al. 1993; Engelman, Hickman et al. 1994).

## Le 3' processing

Le 3' processing est une maturation endonucléolytique des extrémités 3'-OH de l'ADN viral qui se traduit par l'élimination d'un dinucléotide (GT) à chaque extrémité (U5 et U3). La réaction catalysée par un dimère d'IN est une trans-estérification (Figure 8a) qui consiste en une attaque nucléophile sur une liaison phosphodiester de l'extrémité 3' de l'ADN viral. La réaction de 3' processing a lieu dans le cytoplasme

La réaction de 3' processing requière la présence de cations divalents soit  $Mg^{2+}$  soit  $Mn^{2+}$ . Le co facteur métallique *in vivo* est très certainement  $Mg^{2+}$ . Les cations  $Mn^{2+}$  rendent l'activité de l'intégrase plus tolérante aux variations de séquences aux extrémités de l'ADN viral à laquelle s'ajoute une activité nucléase non spécifique (Engelman and Craigie 1995). L'IN est également capable de couper la jonction d'une séquence palindromique lorsque l'ADN viral adopte une forme circulaire (Lee and Han 1996).

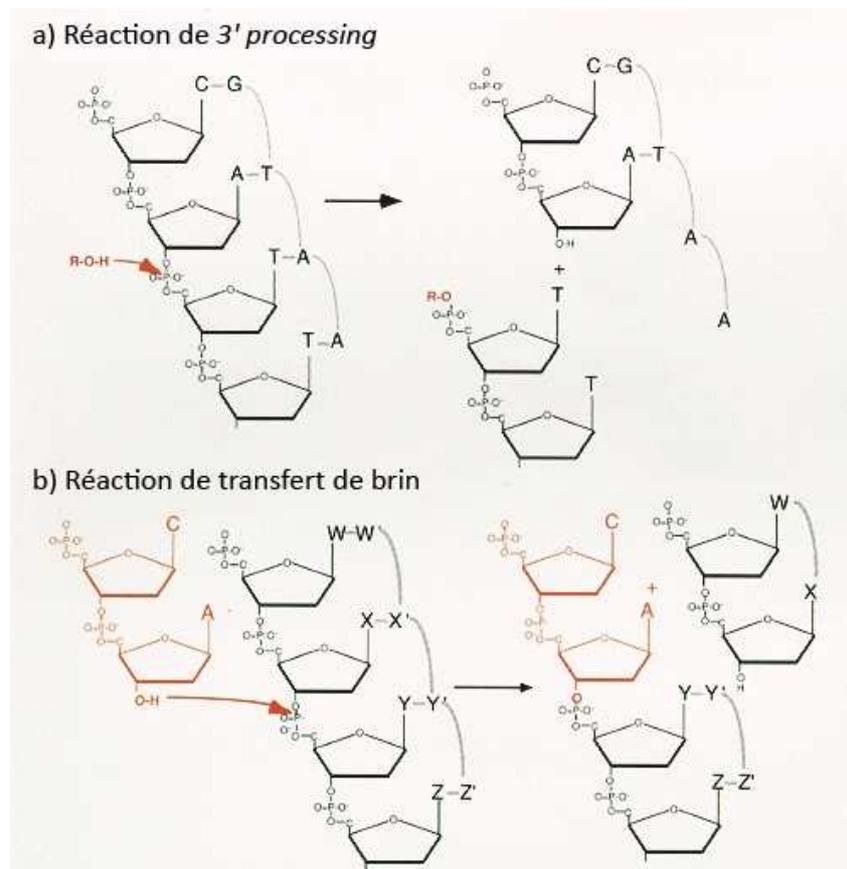


Figure 8 Réaction de trans-estérification catalysée par l'intégrase. (Retroviruses 1997) a) La réaction de 3' processing est une attaque nucléophile sur une liaison phosphodiester de l'extrémité 3' de l'ADN viral. b) La réaction de transfert de brin est également une attaque nucléophile sur une liaison phosphodiester de l'ADN cible par le -OH de l'ADN viral.

## Le transfert de brin

La réaction de transfert de brin a lieu dans le noyau et requière également un cofacteur métallique  $Mg^{2+}$ . La réaction est identique à celle du *3' processing* (Figure 8b). L'attaque nucléophile du groupement hydroxyle de l'extrémité 3' de l'ADN viral sur la liaison phosphodiester de l'ADN cible génomique. La différence d'activité observée en présence de  $Mg^{2+}$  ou de  $Mn^{2+}$  ont permis la conception rationnelle d'inhibiteurs de l'intégrase notamment le Raltégravir et l'Elvitégravir (Marchand, Maddali et al. 2009). Ces inhibiteurs ciblent la réaction de transfert de brin.

Dans la cellule, le transfert de brin s'effectue de manière concerté (Figure 9). La présence de partenaire protéiques viraux (NC) ou cellulaires de l'intégrase augmente l'efficacité (Carteau, Gorelick et al. 1999). L'intégration concertée est difficile à mettre en évidence *in vitro* (Sinha and Grandgenett 2005).

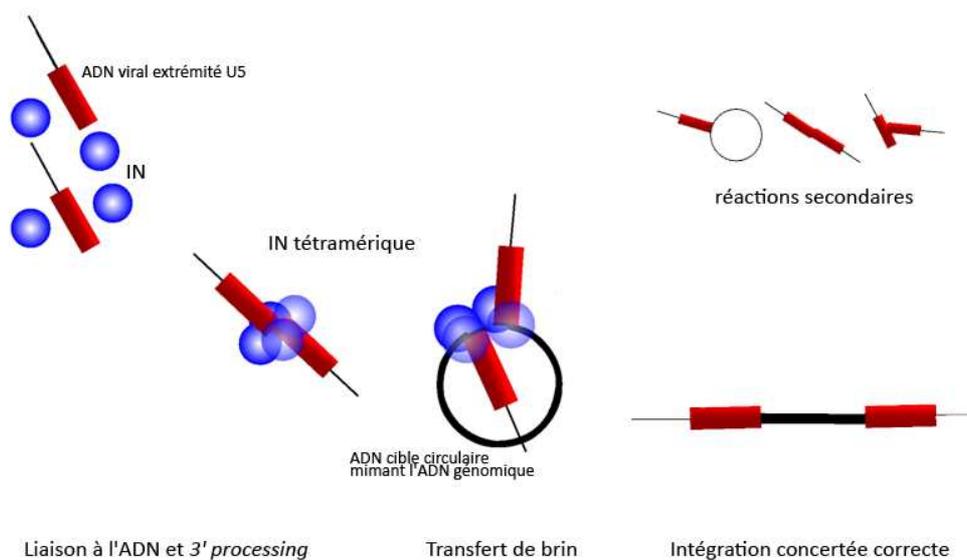


Figure 9 Représentation schématique de l'intégration concertée pour des tests *in vitro*. D'après (Zahm, Bera et al. 2008)

La réparation de l'ADN après le transfert de brin fait appel à des protéines cellulaires (Craigie, Mizuuchi et al. 1991) mais n'est pas effectué par l'intégrase. Un effet collatéral de l'activité de l'intégrase du VIH-1 entraîne une duplication d'une courte séquence chromosomique de part et d'autre de la séquence d'ADN proviral intégrée. Il n'existe pas de réelle sélectivité de séquence cible pour l'intégration, contrairement aux rétrotransposons Ty

de levure. L'intégration semble se produire de façon aléatoire, même si les structures locales de l'ADN (courbure, flexibilité) jouent probablement un rôle.

La réaction finale aboutit donc à la présence d'un ADN viral intégré, le provirus, colinéaire de l'ADN génomique avec duplication d'une courte séquence chromosomique de part et d'autre de la séquence virale. Les tests d'activité *in vitro* reproduisent le processus d'intégration d'une seule extrémité en utilisant des oligonucléotides double-brin mimant la séquence des extrémités virales des LTR U5 ou U3 en présence d'enzyme recombinante.

### **La désintégration**

La désintégration a été mise en évidence *in vitro* (Chow, Vincent et al. 1992). A l'inverse du 3'*processing* et du transfert de brin qui requièrent les trois domaines de l'IN, la réaction de désintégration peut être catalysée par le seul domaine catalytique. Cette réaction peut donc être considérée comme une activité endonucléolytique non spécifique. De façon générale, cette réaction permet de tester le mécanisme compétitif de certains inhibiteurs puisque le domaine central isolé peut, seul, produire cette réaction. Aucune donnée expérimentale n'indique que cette réaction puisse se produire *in vivo*, il serait intéressant de favoriser cette réaction pour diminuer l'efficacité d'intégration.

### **Le Complexe de PréIntégration (CPI)**

Plusieurs protéines cellulaires ont été identifiées en association avec l'intégrase, mais pour la plupart leur rôle dans l'activité de l'intégrase n'a pas été élucidé (Mulder, Chakrabarti et al. 2002). Les laboratoires de Richard Benarous en collaboration avec Hybrigenics, et indépendamment celui de Zeger Debyser, ont identifié un nouveau partenaire d'interaction de l'intégrase du VIH-1, LEDGF (Cherepanov, Maertens et al. 2003; Maertens, Cherepanov et al. 2003; Cherepanov, Devroe et al. 2004; Emiliani, Mousnier et al. 2005) dont le rôle dans l'activité de l'intégrase a pu être au moins partiellement élucidé (Ciuffi, Llano et al. 2005; Emiliani, Mousnier et al. 2005; Busschots, Voet et al. 2007; Hare and Cherepanov 2009). Un domaine de LEDGF se fixant à la chromatine a été mis en évidence et a confirmé que l'une des fonctions de LEDGF est de transporter et de fixer l'intégrase à la chromatine (Ciuffi, Llano et al. 2005; Emiliani, Mousnier et al. 2005; Llano, Saenz et al. 2006; Busschots, Voet et al. 2007; Hare and Cherepanov 2009).

L'interaction de l'intégrase avec d'autres protéines cellulaires a été mise en évidence par le laboratoire de R. Benarous en collaboration avec la société Hybrigenics par un criblage double hybride. VBP1, protéine chaperonne qui se fixe à des polypeptides en cours de synthèse pour favoriser leurs repliements et peut être transporté dans le noyau (Tsuchiya, Iseda et al. 1996). Récemment, il a été montré que VBP1 était impliqué dans le contrôle de la réplication virale (Mousnier, Kubat et al. 2007). La transportine (connue aussi sous le nom d'importin ou de karyopherin), protéine de transport nucléaire essentielle pour l'importation du complexe de préintégration dans le noyau (Bukrinsky and Haffar 1997). La liste des protéines cellulaires hôtes interagissant avec le CPI ou directement avec l'intégrase ne cesse de croître (Turlure, Devroe et al. 2004; Llano, Saenz et al. 2006). L'intégration du génome viral s'effectue préférentiellement près de gènes transcriptionnellement actifs. Il a été montré que p300 une acetyltransférase qui régule la conformation de la chromatine par l'acétylation des histones, acétyle également des lysines dans la partie C-terminale de l'intégrase et contrôle son activité (Cereseto, Manganaro et al. 2005)

### **LEDGF**

La protéine LEDGF (Lens Epithelium Derived Growth Factor) est une protéine 62 KDa (Dietz, Franken et al. 2002) qui a été identifiée comme un partenaire de liaison de la PC4 co-activateur transcriptionnel, suggérant un rôle dans la régulation transcriptionnelle (Ge, Si et al. 1998). La protéine est étroitement associée à la chromatine condensée dans le noyau (Nishizawa, Usukura et al. 2001). La présence de LEDGF dans le milieu de culture de certains types cellulaires stimulent leur croissance et prolonge la survie cellulaire (Singh, Ohguro et al. 2000). Par conséquent, LEDG/p75 a probablement un rôle clé dans l'équilibre entre la survie et la mort cellulaire. Comme facteur de survie (Sharma, Singh et al. 2000), LEDGF favorise la croissance des cellules de mammifères et protège la cellule contre l'apoptose induite par le stress. Le rôle normal de LEDGF est exercée par l'activation transcriptionnelle de protéines liées au stress ou anti-apoptotiques (Shinohara, Singh et al. 2002). La culture de cellules sous stress oxydatif conduit à des niveaux élevés de LEDGF (Sharma, Singh et al. 2000). Il a été démontré que LEDGF liait des séquences d'ADN relatifs au stress (Singh, Fatma et al. 2001). LEDGF est également impliquée dans certaines maladies auto-immunes. Dans le cancer de la prostate, la protéine fonctionne comme un auto-antigène nucléaire pour générer une réponse auto-anticorps (Ganapathy and Casiano 2004; Daniels, Zhang et al. 2005).

La protéine LEDGF contient 530 acides aminés et plusieurs domaines fonctionnels. Dans la région N-ter, un domaine PWWP est présent qui fonctionne comme un domaine d'interaction protéine-protéine (Stec, Nagl et al. 2000) et/ou comme un domaine de liaison à l'ADN (Qiu, Sawada et al. 2002). Un signal de localisation fonctionnelle nucléaire (NLS), (résidus 148-156), est présent (Maertens, Cherepanov et al. 2004; Vanegas, Llano et al. 2005).

LEDGF joue des rôles multiples au cours de l'intégration du VIH. L'interaction avec l'intégrase est spécifique aux lentivirus (Llano, Vanegas et al. 2004; Busschots, Vercammen et al. 2005), et LEDGF est une composante fonctionnelle du CPI (Vandegraaff, Devroe et al. 2006). L'interaction avec LEDGF protège également IN de l'ubiquitination et la dégradation par le protéasome *in vivo* (Llano, Vanegas et al. 2004; Llano, Saenz et al. 2006). La protéine LEDGF stimule l'activité *in vitro* de transfert de brins (Cherepanov, Maertens et al. 2003) et augmente l'affinité pour l'ADN de l'IN (Busschots, Vercammen et al. 2005). L'interaction entre LEDGF et IN a été validé par double-hybride (Emiliani, Mousnier et al. 2005) et par co-immunoprécipitation (Turlure, Devroe et al. 2004). Le domaine d'interaction de l'intégrase de LEDGF est conservé et se situe entre les résidus 347-429 (Cherepanov, Devroe et al. 2004). Une structure RMN du domaine d'interaction de LEDGF avec l'intégrase ainsi qu'une structure cristallographique de ce domaine avec le corps catalytique mutant ont été résolues (Cherepanov, Ambrosio et al. 2005; Cherepanov, Sun et al. 2005). Ce domaine est composé de cinq hélices dont les résidus I365, N366 et F406 ont été identifiés comme importants pour l'interaction avec l'intégrase du VIH-1. Une structure cristallographique des domaines N-ter et corps catalytique de l'intégrase de HIV-2 en complexe avec l'IBD de LEDGF a récemment été résolue (Hare, Shun et al. 2009)

### ***Complexe Intégrase/LEDGF***

Différents modèles théoriques d'organisation structurale et de mécanismes d'intégration ont été proposés (Wiskerchen and Muesing 1995; Gao, Butler et al. 2001; Podtelezhnikov, Gao et al. 2003; Karki, Tang et al. 2004; Chen, Weber et al. 2006) ainsi qu'un mécanisme d'intégration effectué par l'intégrase de Spumavirus à partir de la structure cristallographique de cette intégrase en complexe avec de l'ADN (Hare, Gupta et al. 2010). Au laboratoire, nous avons résolue la structure de l'intégrase sauvage entière de HIV-1 en complexe avec une protéine partenaire, LEDGF et de l'ADN (Michel et al., 2009) (Figure 10). Des structures montrant l'organisation supramoléculaire de multimères d'intégrase en complexe avec des protéines cellulaires seront indispensables pour élucider les mécanismes de transport et

d'intégration de l'ADN viral dans le génome et pour la conception d'inhibiteurs sélectifs efficaces (Jaskolski, Alexandratos et al. 2009). La structure de l'intégrase est modulable et dépend des partenaires avec lesquelles elle interagit.

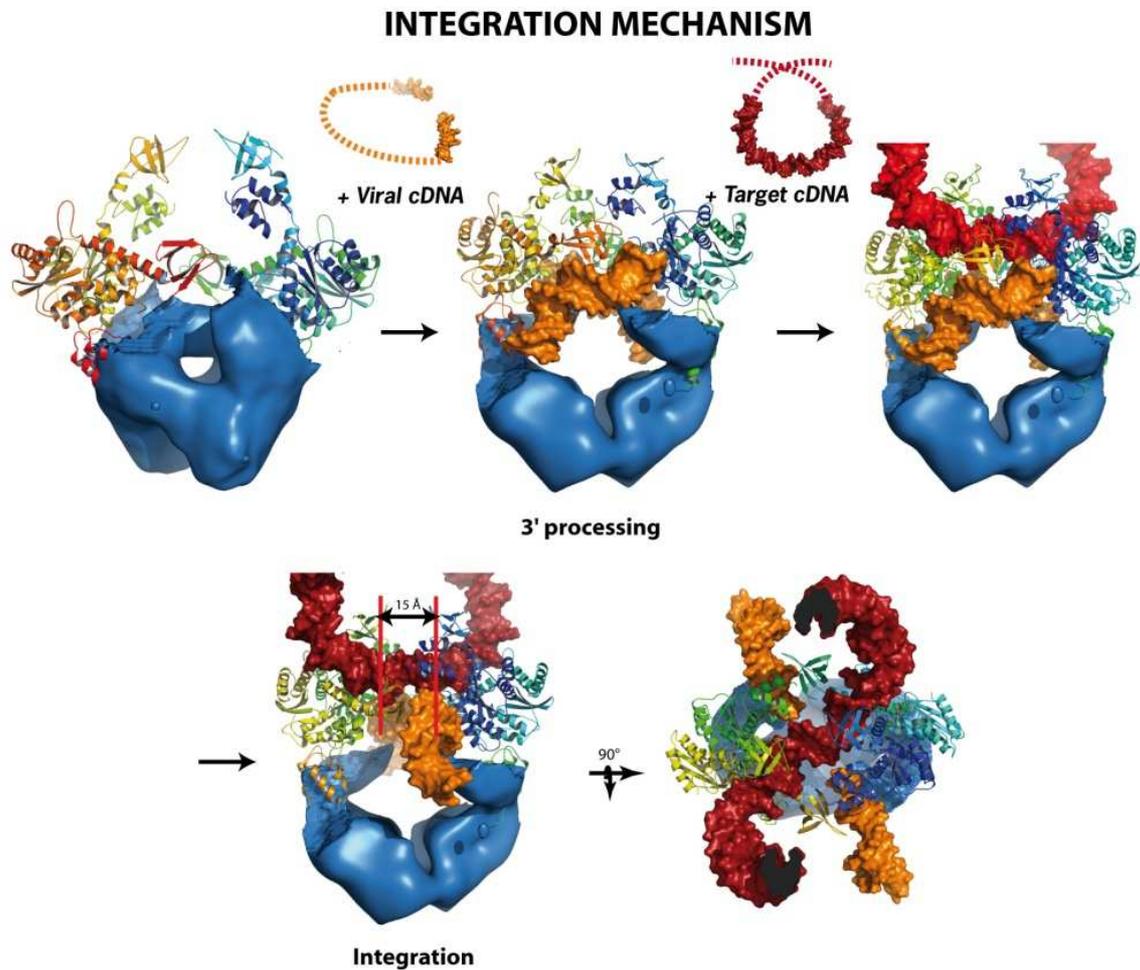


Figure 10 Modèle de la réaction d'intégration. La distance de 15 Å (5 paires de bases) correspond à la distance d'intégration entre l'ADN viral (en jaune) et l'ADN cible (en rouge). Le complexe se compose d'un tétramère d'intégrase (en cartoon) et d'un dimère de LEDGF (en bleu).

### *INI1*

INI1 est l'homologue humain de la protéine SNF5 de la levure. La protéine INI1 appartient au complexe SWI/SNF de mammifère (Wang, Cote et al. 1996), qui est impliquée dans le remodelage de la chromatine ATP-dépendante, c'est un activateur de la transcription, qui en décompactant la chromatine, facilite l'accès de la machinerie transcriptionnelle à l'ADN (Carlson and Laurent 1994). Le complexe SWI/SNF conservé au cours de l'évolution restructure la chromatine en décompactant les nucléosomes. Neuf protéines dont INI1 composent le cœur invariant du complexe et un nombre variable de sous-unités qui contribue

à l'activation ou la répression de la transcription (Roberts and Orkin 2004). La transactivation du promoteur tat du VIH-1 est aidée par SWI/SNF (Pumfery, Deng et al. 2003). INI1 est une protéine nucléaire de 385 acides aminés. La protéine INI1 est composée de trois régions conservées, dont deux répétitions directe (Morozov, Yung et al. 1998). Les résidus 106-183 constituent un domaine de liaison à l'ADN non spécifique potentiel.

#### INI1 dans la phase précoce

En 1994, INI1 a été découverte comme interagissant avec l'intégrase du VIH-1 et stimulant le transfert de brin *in vitro* (Kalpana, Marmon et al. 1994). Cette interaction est spécifique (Yung, Sorin et al. 2004). L'analyse par délétion d'INI1 montre que la région 183-243 est nécessaire et suffisante pour interagir avec l'intégrase du VIH-1 et que les résidus T213, D224, D226 et S246 sont essentiels pour l'interaction avec l'intégrase du VIH-1 (Yung, Sorin et al. 2001). En outre, une séquence d'exportation nucléaire (NLS) masquée (266-276) a été identifiée au début de la région répétée 259-319 (Craig, Zhang et al. 2002). Le domaine minimum d'interaction à l'intégrase d'INI1 comprend les résidus 183 à 294. La région minimale de l'intégrase du VIH-1 interagissant avec INI1 est 53-153. Les résidus K71 et E69 sont essentiels pour cette interaction (Maroun, Delelis et al. 2006).

Plusieurs résidus importants pour la multimérisation d'INI1 ont été mis en évidence (REF XX : Kalpana 2009 multimerisation) au début du domaine répété coïncidant avec le signal d'export nucléaire (266-276). Ces résidus sont hydrophobes ce qui suggère que la multimérisation d'INI1 s'effectue par les régions hydrophobes. La multimérisation d'INI1 permettrait de conserver sa co-localisation avec l'intégrase dans le noyau. La formation de dimère, tétramère, octamère et plus est dépendante de la concentration d'INI1. Les régions permettant la multimérisation et le domaine d'interaction avec l'intégrase se recouvrent.

INI1 interagit également avec le petit sillon de l'ADN ce qui pourrait stimuler la réaction de transfert de brin.

La relocalisation cytoplasmique d'INI1 nucléaire est détectée 30 min après l'infection. Jusqu'à 4 h après l'infection, INI1 est encore principalement localisée dans le noyau. Ce recrutement cytoplasmique nécessite l'entrée du virus, mais est indépendante de l'enveloppe du VIH-1. Par visualisation des particules virales à l'intérieur de la cellule, la co-localisation des complexes de pré-intégration du VIH avec INI1 a été montrée. Des premières études ont montrés que le rôle d'INI1 est de stimuler l'activité de l'intégrase *in vitro* (Kalpana, Marmon et al. 1994).

La relocalisation cellulaire d'INI1 dans le cytoplasme est observée au cours de l'infection du VIH-1. Dans le cytoplasme, INI1 pourrait s'associer au complexe de préintégration pendant les étapes précoces de l'infection par le VIH-1.

#### *INI1 dans la phase tardive*

L'expression d'INI1 pleine longueur montre une nette augmentation de la production de particules virales infectieuses. En outre, INI1 est présent dans les virions VIH-1. La contribution potentielle d'INI1 au cours de l'infection par le VIH-1 a également été étudiée en association avec une autre protéine nucléaire, la protéine de la leucémie promyélocytaire (PML). Tôt dans l'infection du VIH-1, une redistribution subcellulaire par l'intermédiaire de l'exportine d'INI1 et de PML est observé (Turelli, Doucas et al. 2001). Il est fort probable qu'INI1 joue également un rôle au cours des étapes post-intégration (Van Maele, Busschots et al. 2006).

#### *Le fragment S6*

Un fragment (183-294) d'INI1 réduit la production virale à des niveaux indétectables (Yung, Sorin et al. 2001). Quand le fragment d'INI1 est surexprimé, il agit comme un inhibiteur spécifique dans les étapes tardives de la réplication du VIH-1. La présence de ce fragment dans les cellules diminue la libération du virus de façon spectaculaire, alors que des mutations dans ce fragment qui perturbent les interactions de l'intégrase abrogent cet effet inhibiteur. La localisation cellulaire de ce fragment est cytoplasmique au lieu d'être nucléaire. Ces expériences avec les fragments d'INI1 ne précisent pas le rôle d'INI1 pleine longueur au cours de la réplication rétrovirale.

#### *Le rôle controversé d'INI1*

Les résultats sont contradictoires concernant le rôle que joue INI1 pendant l'infection du VIH-1. Une étude (Boese, Sommer et al. 2004) a suggéré qu'INI1 n'avait aucun effet sur l'intégration du virus dans des cellules déplétives en INI1. Tandis que d'autres rapports ont montrés qu'INI1 participe à une réponse antivirale cellulaire en interférant avec les étapes précoces de la réplication virale (Turelli, Doucas et al. 2001; Maroun, Delelis et al. 2006). A l'inverse, une autre étude (Ariumi, Serhan et al. 2006) a montrée qu'INI1 renforce indirectement la transcription provirale. INI1 a également été jugée essentielle pour la production de particules virales infectieuses (Yung, Sorin et al. 2001; Sorin, Yung et al. 2006). Les particules virales produites en l'absence d'INI1 présentent une réduction drastique de la synthèse de l'ADN proviral, ce qui suggère qu'INI1 est également essentiel pour les événements au début de la transcription inverse. L'incorporation d'INI1 dans le virion est

médiée par un contact direct avec l'intégrase dans le contexte de la polyprotéine Gag-Pol et est nécessaire à l'infection. La compréhension des interactions entre l'intégrase et INI1 permettrait d'affiner cette cible potentielle pour la création d'inhibiteurs bloquant ces interactions.

## Les inhibiteurs de l'intégrase

Plusieurs familles de molécules ont pu être identifiées comme des inhibiteurs potentiels de l'intégrase (Marchand, Maddali et al. 2009) dont les composés de la famille des acides dicétoniques (Hazuda, Felock et al. 2000; Grobler, Stillmock et al. 2002). Ces molécules sont développées par Merck & Co et Shion/GlaxoSmithKline. En octobre 2007, une molécule de cette famille, le Raltegravir a été approuvé par la FDA. C'est un agent anti HIV très efficace mais avec des effets secondaires importants (Temesgen, Warnke et al. 2006). D'autres familles de molécules comme les styrylquinoléines (Ouali, Laboulais et al. 2000; Zouhiri, Mouscadet et al. 2000), des peptides (DeSoultrait, Caumont et al. 2002), des oligonucléotides (Jing and Xu 2001; Jing, Xiong et al. 2002) représentent des agents thérapeutiques potentiels. Des structures du domaine catalytique de l'intégrase en présence d'inhibiteurs potentiels : 5-CITEP (dicétoniques) (Goldgur, Craigie et al. 1999) ou des tetraphenyl arsenates (Molteni, Greenwald et al. 2001) ont été déterminées. L'amélioration de ces composés nécessite impérativement que l'on dispose de plus d'informations structurales sur la reconnaissance entre l'intégrase et ses ligands (Pommier, Johnson et al. 2005).

## Projet de recherche

Le but de ce projet est de comprendre les mécanismes moléculaires de l'intégration de l'ADN du virus du VIH dans le génome des cellules hôtes. Pour cela il est indispensable de connaître les structures atomiques des diverses protéines impliqués ainsi que leur organisation supramoléculaire. Les cristaux de l'intégrase entière sauvage obtenus jusqu'à présent ne sont pas de qualité suffisante pour déterminer la structure. L'hypothèse principale est la non structuration d'une partie de la protéine ou une importante mobilité inter domaine. Afin de stabiliser la structure de l'intégrase sauvage, nous préparons des complexes avec des protéines partenaires comme LEDGF et INI1.

Une étape intermédiaire pour une meilleure compréhension de l'oligomérisation de l'intégrase et de son interaction avec les partenaires cellulaires est la détermination de structures basses

résolutions par microscopie électronique, étape qui s'avère indispensable pour la définition de complexes stable en vue de leurs cristallisations.

Au cours de ma thèse j'ai résolue une structure basse résolution du complexe IN-LEDGF-INI1 par des techniques de cryo microscopie électronique. Cette structure montre une organisation de l'intégrase en tetramère, avec une interface de tetramérisation impliquant les domaines C- et N-terminaux, en complexe avec un dimère de LEDGF et un dimère de INI1.

Mon projet a consisté à : a) cloner avec l'aide de la plate forme de génomique structurale INI1 dans divers vecteurs d'expressions pour *E. coli* (système pET), préparer une banque de fragments d'INI1 contenant le domaine de liaison à l'intégrase avec différentes bornes déterminées par une analyse bioinformatiques et parmi ces fragments, utiliser les plus solubles pour des études structurales; b) purifier ces protéines à homogénéité c) préparer des complexes stables intégrase-LEDGF-INI1 pour des études par microscopie électronique et cristallographie ; d) résoudre et exploiter les structures des complexes pour déterminer un modèle supramoléculaire de l'organisation de ce complexe ; e) faire une analyse fonctionnelle de ce complexe ternaire et le comparer au complexe IN/LEDGF pour comprendre le rôle de INI1 dans l'étape précoce de l'infection par le VIH-1.

## Étude Bioinformatique

### Matériel et méthodes

#### La stratégie *in silico*

L'analyse des séquences protéiques commence par la réalisation d'un alignement multiple. L'alignement multiple de séquences protéiques trahit les homologies de séquences. Ces homologies sont révélatrices de domaines conservés et/ou de conservation de résidus ayant le plus souvent une signification fonctionnelle et/ou structurale. Il est possible également d'établir une phylogénie à partir de l'alignement multiple. Couplée à l'alignement multiple, une analyse des prédictions des structures secondaires et intrinsèquement désordonnées affine les limites des domaines. En règle générale, un domaine structuré a de meilleures chances d'être soluble quand les limites ne sectionnent pas de structures secondaires, de séquences hydrophobes ou une zone intrinsèquement désordonnée. La délimitation des bornes précises des domaines structuraux s'effectue grâce à l'utilisation conjointe de ces différentes analyses et est pondérée en fonction de l'importance et de la convergence des différentes prédictions.

#### L'environnement bioinformatique

Au cours de ma thèse, j'ai bénéficié des conseils avisés du Laboratoire de Biologie et de Génomique Structurales (LBGS) et de la Plateforme de Bioinformatique de Strasbourg (BIPS) (<http://bips.u-strasbg.fr/>). BIPS est une plateforme à haut débit, localisée à l'IGBMC, dédiée à l'analyse de séquences ainsi qu'à la génomique comparative et structurale. Elle a été labellisée Plateforme Nationale RIO en 2003 et constitue la Plateforme de Bioinformatique du Génopôle Grand-Est « du Gène au Médicament ».

#### GScope

GScope est une plateforme de génomique dédiée à la manipulation, le stockage et l'analyse de données massives englobant aussi bien des génomes complets qu'un ensemble de gènes ou de protéines. Ecrite en Tcl/Tk, et gérant des fichiers ascii, la plateforme GScope est développée au LBGI, principalement par Raymond Ripp qui maintient la cohérence générale de cet outil. GScope contient des programmes propres et des programmes externes, en particulier des programmes développés au sein du groupe autour de l'alignement multiple. Son architecture modulaire permet l'utilisation de modules spécifiques dans d'autres applications ainsi que l'intégration de nouveaux modules d'analyse. Une interface web est également disponible

(WScope). Au cours de ce travail, GScope a été utilisée pour rechercher des séquences protéiques et des structures dans les bases de données ainsi que pour construire les différents alignements multiples de séquences protéiques. De plus, une composante de GScope, dédiée au clonage, m'a permis de créer, étudier, stocker et gérer toutes les informations nucléotidiques ou protéiques relatives aux gènes d'intérêt en réalisant, de manière automatique, le dessin des oligonucléotides nécessaires au clonage ou à la mutagenèse, la commande d'oligonucléotides, la création virtuelle des produits de PCR et des vecteurs d'expression, la vérification du séquençage.

## PipeAlign

PipeAlign (Plewniak, Bianchetti et al. 2003) est un outil d'analyse de famille de protéines développé au sein du Laboratoire de Bioinformatique et de Génomique Intégratives (LBGI). Ce programme permet de réaliser des alignements multiples de séquences protéiques de manière automatique et regroupe les séquences en sous-familles. Ces alignements multiples de séquences protéiques complètes (MACS, Multiple Alignment of Complete Sequences) (Lecompte, Thompson et al. 2001) sont effectués à partir d'une recherche initiale dans les bases de données UniProt et PDB par BlastP à partir d'une séquence protéique unique ou d'un jeu de séquences protéiques. PipeAlign consiste en une cascade de sept programmes couplés qui peuvent également être utilisés de manière indépendante (

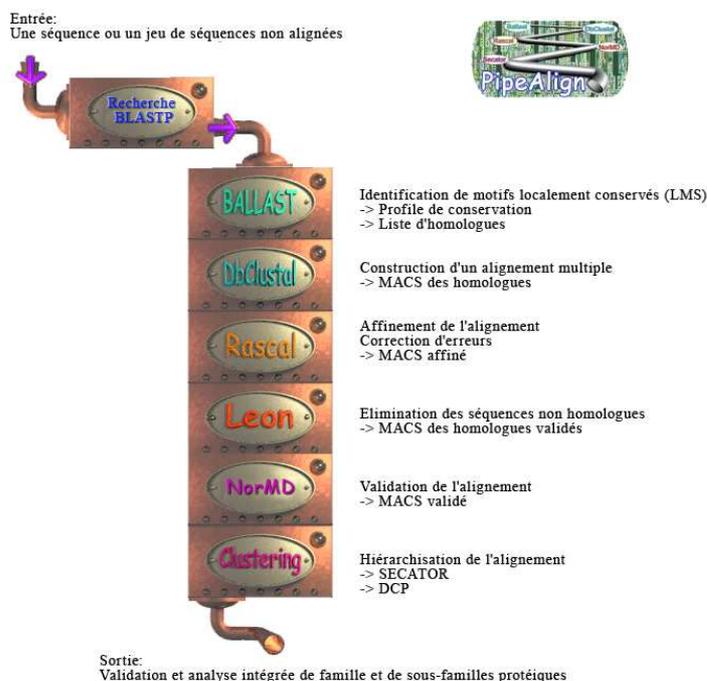


Figure 11).

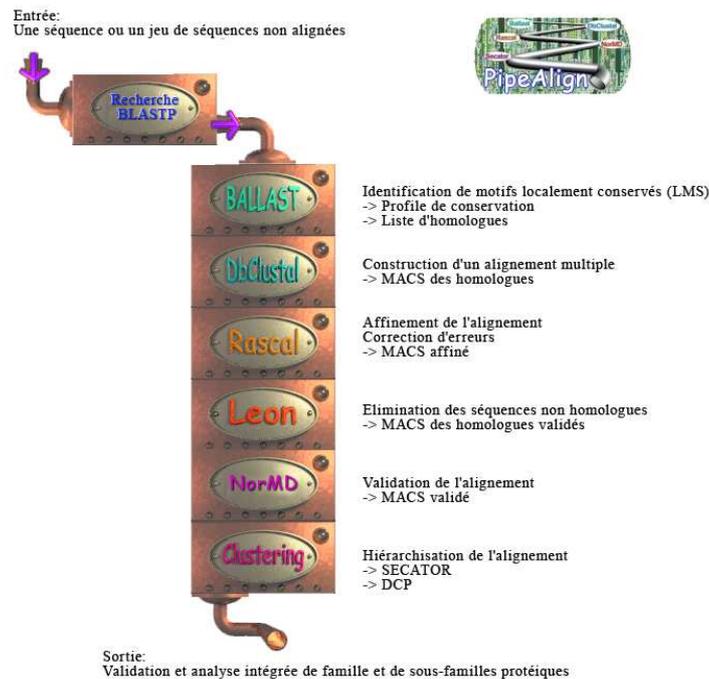


Figure 11: Aperçu de la cascade de programmes de PipeAlign

## **BALLAST**

La procédure BALLAST (*Blast post-processing based on locally conserved segments*) consiste dans la recherche des séquences similaires et homologues à partir du programme BlastP (*Basic Local Alignment Search Tool Protein*) et ne retient que les résultats significatifs pour les séquences similaires détectées. La contribution de chaque touche dans le profil est proportionnelle à sa significativité. Les LMSs (*Local Maximum conserved Segments*) correspondent aux segments de séquences les mieux conservés entre la séquence initiale et les touches détectées par BlastP. Ces segments de conservation maximale permettent au module BALLAST de retenir aussi les séquences homologues divergentes par les segments de séquences relativement conservés. Les positions des LMSs dans chaque séquence sont identifiées et conservées dans un fichier comme une liste d'ancres locales utilisées par le module suivant : DbClustal.

## **DbClustal**

La procédure DbClustal est une version modifiée de ClustalW (Thompson, Higgins et al. 1994) qui permet de combiner les approches globale et locale. En effet, DbClustal est basé sur le même algorithme d'alignement global (Needleman and Wunsch 1970) que ClustalW pour construire des MACS ce qui présente des inconvénients pour les séquences contenant des

insertions ou des extensions N-terminales ou C-terminales (Thompson, Plewniak et al. 1999). Pour pallier ces aléas, les alignements locaux LMSs sont utilisés comme points d'ancrage et pondèrent l'alignement global. Le MACS obtenu en sortie du programme DbClustal est divisé horizontalement (blocs de forte conservation) et verticalement (présence de sous-familles) pour former un « quadrillage » au sein duquel les régions bien alignées, et donc fiables, peuvent être identifiées et différenciées selon leur degré de conservation ou leur appartenance à une sous-famille.

### **RASCAL**

Le programme RASCAL (*RApid Scanning and Correction of multiple sequence ALignments*) (Thompson, Thierry et al. 2003) est utilisé dans le cadre de la détection et de la correction des erreurs subreptices au sein de l'alignement multiple. En effet, DbClustal est basé sur un algorithme utilisant des approximations ce qui implique l'introduction d'erreurs potentielles. Elles sont détectées par comparaison des profils des régions fiables (« *core blocks* »). RASCAL réaligne chaque région mal alignée en utilisant un algorithme proche de celui implémenté dans ClustalW. La correction de l'alignement est restreinte aux régions les moins fiables, permettant une stratégie de ré-affinement plus performante.

### **LEON**

Le module LEON (*multiple aLignment Evaluation Of Neighbours*) (Thompson, Prigent et al. 2004) vérifie l'homologie des séquences *i.e.* leur appartenance à une famille d'intérêt. En effet, dans un alignement multiple les séquences doivent être homologues. En se basant sur les régions fiables délimitées par RASCAL, LEON profite de la nature transitive des relations d'homologie : l'information des séquences intermédiaires est mise à profit pour mettre en évidence les régions conservées des séquences les plus divergentes. Les blocs de conservation de chaque sous-famille du MACS sont ensuite reliés, afin de former des régions contiguës de conservation, considérées comme homologues à la séquence initiale. La composition en acides aminés des séquences de l'alignement est également prise en compte par l'incorporation d'un certain nombre d'algorithmes de détection de segments, dont la composition est biaisée. Finalement, les séquences qui ne contiennent que des blocs disjoints et aucune région homologue sont retirées du MACS. En sortie de LEON, nous disposons donc d'un MACS de bonne qualité qui ne contient que des séquences partageant au moins une région homologue à la séquence initiale.

## **NorMD**

NorMD (*Normalized Mean Distance*) (Thompson, Plewniak et al. 2001), est une fonction objective utilisée pour évaluer la qualité d'un MACS, résultat des quatre étapes précédentes. NorMD combine les avantages des techniques basées sur les scores de colonnes avec la sensibilité des méthodes introduisant des scores de similarité de résidus. Le score assigné à l'alignement est indépendant du nombre de séquences que contient le MACS, de leur pourcentage d'identité, de leur longueur, etc. Ceci permet de comparer des scores NorMD entre alignements indépendants. Le score assigné par NorMD sera, de manière générale, compris entre 0 et 1. Plus le score est proche de 1, plus la qualité de l'alignement peut être considérée comme satisfaisante. Après analyse du comportement de NorMD sur de nombreux cas référencés, un seuil de 0,3 a été fixé, en dessous duquel la qualité du MACS est considérée comme insatisfaisante.

## **Secator et DPC**

Les programmes Secator (Wicker, Perrin et al. 2001) et DPC (*Density of Points Clustering*) (Wicker, Dembele et al. 2002) permettent de classifier, ou hiérarchiser, des séquences d'un MACS dans de potentielles sous-familles de séquences fonctionnelles. Le nombre de sous-familles créées étant déterminé automatiquement par ces programmes. Le résultat de PipeAlign est un MACS validé de haute qualité, dans lequel les séquences sont classées en sous-familles. PipeAlign est mis à disposition de la communauté scientifique par l'intermédiaire du site web de la plateforme de bioinformatique de Strasbourg (<http://bips.u-strasbg.fr/PipeAlign/>).

## **Logiciel d'analyse des MACS**

### ***JalView 2 & MACSIMS***

JalView version 2 (Waterhouse, Procter et al. 2009) est un programme de visualisation, d'analyse et d'édition de MACS. MACSIMS (Thompson, Muller et al. 2006) est un programme qui facilite la gestion et l'extraction d'informations pertinentes au sein de l'alignement. Le programme combine les informations (structurale et fonctionnelle) extraites des banques de données et des méthodes d'analyse de séquences *ab initio* comme l'identification des régions homologues. Ces programmes ont été utilisés lors du traitement et l'analyse des MACS.

## Définition automatique des domaines

Plusieurs programmes sont disponibles pour prédire automatiquement des domaines. La combinaison de plusieurs méthodes permet de confronter différentes prédictions et ainsi d'en augmenter la perspicacité.

### *ProDom*

Prodom (<http://prodom.prabi.fr/prodom/current/html/home.php>) est une base de données référençant toutes les familles de domaines de protéine générées automatiquement à partir des séquences complètes de SWISS-Prot et TrEMBL (Corpet, Servant et al. 2000). Chaque entrée de ProDom produit un alignement multiple de séquences de domaines homologues et une séquence consensus représentative de la famille. ProDom-SG (Bru, Courcelle et al. 2005) est une base de données orientée vers la structure tridimensionnelle des protéines qui permet de recruter par PSI-BLAST de nouvelles séquences homologues à partir de domaines bien définis de la base de données SCOP (Lo Conte, Brenner et al. 2002).

### *pFAM*

La base de données pFAM (Finn, Mistry et al. 2010) (<http://pfam.sanger.ac.uk/>) est divisée en deux niveaux de qualité : pFAM-A générée manuellement et pFAM-B générée automatiquement. Les domaines présents dans pFAM-A sont déterminés à partir d'un alignement contenant un petit nombre de séquences représentatives de la famille et sont caractérisés par un modèle de Markov caché qui permet son application à la séquence d'intérêt. Les domaines de la base de données pFAM-B sont créés automatiquement à partir d'un groupe de séquences non redondantes produit par l'algorithme ADDA (Heger, Wilton et al. 2005) et contient un grand nombre de familles mais qui ne sont pas annotées.

## Prédiction des structures secondaires

Plusieurs programmes sont disponibles pour prédire les structures secondaires à partir d'une séquence primaire. De la même manière que pour la définition des domaines, en compilant plusieurs résultats de différents programmes, la combinaison des prédictions augmente la véracité des recouvrements.

## PredictProtein

Le serveur PredictProtein (<http://www.predictprotein.org/>) (Rost, Yachdav et al. 2004) regroupe plusieurs programmes détaillés ci-après (Figure 12). Les résultats compilent des prédictions recoupant plusieurs programmes dont certains utilisés par ailleurs (prédictions de domaines). Les résultats obtenus par PredictProtein seront surtout focalisés sur les prédictions de structures secondaires et l'analyse de la séquence primaire.

Méthodes	Taches	Références	Sites
<b>Base de données</b>			
Swiss-Prot	Séquences protéiques annotées	(Bairoch and Apweiler 2000)	<a href="http://expasy.cbr.nrc.ca/sprot/">http://expasy.cbr.nrc.ca/sprot/</a>
TrEMBL	Séquences protéiques brutes	(Bairoch and Apweiler 2000)	<a href="http://www.ebi.ac.uk/swissprot/">http://www.ebi.ac.uk/swissprot/</a>
PDB	Séquences de structures déterminées expérimentalement	(Berman, Westbrook et al. 2000)	<a href="http://www.rcsb.org/pdb/">http://www.rcsb.org/pdb/</a>
BIG	Combinaison non-redondante Swiss-Prot, TrEMBL, PDB	(Przybylski and Rost 2002)	
<b>Alignement</b>			
MaxHom	Alignement multiple	(Sander and Schneider 1991)	
BLASTP	Alignement par comparaison appariée	(Altschul and Gish 1996)	<a href="http://www.ncbi.nlm.nih.gov/BLAST/">http://www.ncbi.nlm.nih.gov/BLAST/</a>
PSI-BLAST	Alignement basé sur un profil	(Altschul, Madden et al. 1997)	<a href="http://www.ncbi.nlm.nih.gov/BLAST/">http://www.ncbi.nlm.nih.gov/BLAST/</a>
HMMer	Recherche par un modèle de Markov caché	(Eddy 1998)	<a href="http://hmmer.wustl.edu/">http://hmmer.wustl.edu/</a>
TOPITS	Prediction-based threading	(Rost 1995; Rost, Schneider et al. 1997)	
<b>Domaine</b>			
ProDom	Régions de domaine similaire	(Corpet, Servant et al. 2000)	<a href="http://protein.toulouse.inra.fr/prodom.html">http://protein.toulouse.inra.fr/prodom.html</a>
Pfam-A	Famille de protéine	(Bateman, Coin et al. 2004)	<a href="http://www.sanger.ac.uk/Software/Pfam/">http://www.sanger.ac.uk/Software/Pfam/</a>
CHOP	Fragments de domaine similaire	(Liu and Rost 2004)	
SEG	Régions faiblement complexes	(Wootton and Federhen 1996)	<a href="http://trex.musc.edu/manuals/unix/seg.html">http://trex.musc.edu/manuals/unix/seg.html</a>
NORSp	Régions mobiles	(Liu, Tan et al. 2002; Liu and Rost 2003)	
<b>Structure</b>			
PHDsec	Structure secondaire	(Rost and Sander 1993; Rost and Sander 1994; Rost 1996)	<a href="http://www.predictprotein.org">http://www.predictprotein.org</a>
PHDacc	Accessibilité au solvant	(Rost and Sander 1994; Rost 1996)	<a href="http://www.predictprotein.org">http://www.predictprotein.org</a>
PHDhtm	Hélices membranaires	(Rost 1996; Rost, Fariselli et al. 1996)	<a href="http://www.predictprotein.org">http://www.predictprotein.org</a>
PROFsec	Structure secondaire	(Rost 2001)	<a href="http://www.predictprotein.org">http://www.predictprotein.org</a>
PROFacc	Accessibilité au solvant	B. Rost Unpublished	<a href="http://www.predictprotein.org">http://www.predictprotein.org</a>
GLOBE	Caractéristique globale	B. Rost Unpublished	<a href="http://www.predictprotein.org">http://www.predictprotein.org</a>
COILS	Régions enroulées	(Lupas 1996)	
CYSPRED	Ponts disulfure	(Fariselli, Riccobelli et al. 1999)	<a href="http://cassandra.dsi.unifi.it/cysteines/index.html">http://cassandra.dsi.unifi.it/cysteines/index.html</a>
ASP	Régions de conformation ambivalente	(Young, Kirshenbaum et al. 1999)	<a href="http://prion.biocomp.unibo.it/cyspred.html">http://prion.biocomp.unibo.it/cyspred.html</a>
PROFcon08	Contact inter-résiduel	(Punta and Rost 2005)	<a href="http://www.predictprotein.org">http://www.predictprotein.org</a>
PROFtmb	Structure membranaire de type barrel	(Bigelow and Rost 2006)	<a href="http://www.predictprotein.org">http://www.predictprotein.org</a>
<b>Fonction</b>			
PredictNLS	Signal de localisation nucléaire (NLS)	(Cokol, Nair et al. 2000; Liu and Rost 2003)	<a href="http://cubic.bioc.columbia.edu/predictNLS">http://cubic.bioc.columbia.edu/predictNLS</a>
PROSITE	Motifs fonctionnels	(Hofmann, Bucher et al. 1999)	<a href="http://www.ncbi.nlm.nih.gov/BLAST/">http://www.ncbi.nlm.nih.gov/BLAST/</a>
LOCnet	Prédiction de localisation sub-cellulaire	(Nair and Rost 2003)	

Figure 12: Récapitulatif des tâches effectuées par le programme PredictProtein adapté de (Rost, Yachdav et al. 2004).

## GARNIER

Garnier appartient à la suite EMBOSS (*European Molecular Biology Open Software Suite*) (<http://bips.u-strasbg.fr/EMBOSS/>). Il permet de prédire la structure secondaire d'une

séquence protéique à partir de la méthode GOR (*Garnier Osguthorpe Robson*) (Garnier, Osguthorpe et al. 1978). Cette méthode est basée sur les probabilités de trouver un résidu donné formant des structures secondaires. La probabilité est calculée pour chaque position dans une fenêtre de 17 acides aminés. Ces paramètres probabilistes sont dérivés de l'étude empirique de structures connues résolues par cristallographie aux rayons X. Quatre résidus contigus sont nécessaires à la prédiction d'une hélice  $\alpha$  et deux pour un feuillet  $\beta$ .

## KoAnno

Un grand nombre de régions protéiques et de protéines peuvent exercer leur fonction sans un repliement bien défini (Wright and Dyson 1999). Ces protéines sont nommées « intrinsèquement désordonnées ». La composition en acides aminés, la complexité de la séquence, l'hydrophobicité, la charge nette sont autant de paramètres qui sont utilisés pour déterminer *in silico* le désordre intrinsèque. Les caractéristiques des régions désordonnées sont généralement une faible hydrophobicité, une charge nette élevée, une très faible variabilité dans la composition en acides aminés.

De nombreux logiciels de prédiction des régions désordonnées dans les protéines, s'appuyant sur des algorithmes différents et ayant chacun leur propre définition du désordre, sont actuellement disponibles. Pour améliorer l'efficacité de prédiction, une des stratégies est de combiner plusieurs méthodes performantes. KoAnno est un serveur web (<http://www-bio3d-igbmc.u-strasbg.fr/~brelivet/KoAnno/>) développé au sein du LGBI par Yann Brelivet et maintenu par Raymond Ripp. Le programme KoAnno a été écrit en Tcl/Tk, fait parti GScope. Ce programme regroupe quatre programmes de prédiction des régions désordonnées dans les protéines.

## RONN

RONN est une extension du réseau de neurones BBFNR (*Bio-Basis Functional Neural Networks*) (Yang, Thomson et al. 2005) destiné à l'origine à la caractérisation de sites de clivage protéolytique et maintenant adapté à la détection de régions désordonnées. Des bibliothèques d'alignements de séquences de repliements connus (ordonnés, désordonnés ou un mélange des deux) ont été réalisées manuellement d'après des données structurales. Si l'on veut connaître les zones ordonnées ou désordonnées d'une séquence quelconque, une

méthode consiste à réaliser l'alignement de cette séquence par rapport aux séquences de ces bibliothèques. Suivant le degré de similitude, les zones ordonnées ou désordonnées peuvent être prédites. Cette prédiction est réalisée par un réseau de neurones entraîné sur ces bibliothèques de séquences connues.

### *GlobPlot2*

GlobPlot2 (Linding, Russell et al. 2003), outil d'identification de régions globulaires et désordonnées dans les protéines, utilise une approche simple basée sur les propensions des acides aminés à se trouver soit dans des structures secondaires régulières (hélices  $\alpha$  ou feuillets de brins  $\beta$ ) soit dans d'autres structures de type «Random Coil», boucles, « turns », etc.) ; trahissant un état ordonné ou désordonné.

### *DisEMBL1.5*

Le programme DisEMBL1.5 (Linding, Jensen et al. 2003) est basé sur un réseau de neurones. Pour assigner le désordre, le programme passe en revue trois critères différents. Les résidus présents dans les hélices  $\alpha$ , les hélices 3.10 ou les feuillets  $\beta$  sont considérés comme ordonnés. Le premier critère est « LOOPS/COILS » (boucle/enroulée). L'appartenance des résidus à des états de type « turn » (liaison H de type  $CO_{(i)}$  et  $NH_{(i+n)}$ ), « bend » (chaîne des  $C\alpha$  présentant une forte courbure d'au moins  $70^\circ$ ) et les hélices  $\pi$  est nécessaire mais pas suffisant pour qualifier leur état désordonné. Le deuxième critère « HOTLOOPS » représente des boucles avec un degré de mobilité élevé (facteur B). Pour caractériser le désordre dans les protéines, DisEMBL1.5 utilise une technique dérivée basée sur une échelle de propension de mobilité des résidus en fonction des facteurs d'agitation. Les coordonnées absentes dans une structure résolue par cristallographie aux rayons X définissent le troisième critère de définition du désordre par le terme « REM465 ». En effet, les densités électroniques non assignées reflètent le plus souvent un désordre intrinsèque. Des réseaux de neurones ont été entraînés sur des jeux de données correspondant aux trois définitions du désordre (« LOOPS/COILS », « HOTLOOPS », « REM465 »).

### *FoldIndex©*

FoldIndex© (Prilusky, Felder et al. 2005) prédit si une séquence protéique donnée est intrinsèquement désordonnée. Ce calcul est basé sur la moyenne des résidus hydrophobes  $\langle H \rangle$  (la somme de tous les résidus hydrophobes échelonnée de 0 à 1, divisée par le nombre total

de résidus) de la protéine et sur la valeur absolue de sa charge nette  $\langle R \rangle$  (différence entre le nombre de résidus chargés positivement et négativement à pH7, divisée par le nombre total de résidus).

L'équation utilisée par le programme est :  $I_F^{KD} = 2.785 \langle H \rangle - |\langle R \rangle| - 1.151$ . Toutes les valeurs positives représentent des domaines repliés et les valeurs négatives des domaines intrinsèquement désordonnés.

## Résultats

### Pourquoi une étude *in silico* ?

L'étude *in silico* a pour but de prédire les bornes des constructions qui seront utilisées par la suite. Deux principales raisons motivent l'élaboration de ces constructions. Sans étiquette de solubilité (GST ou MBP) et/ou sans détergent, la protéine complète INI1 de type sauvage est insoluble. Les constructions d'INI1 contenant séparément un ou plusieurs domaines permettent de déterminer les rôles des différents domaines d'INI1. Les bornes des constructions sont choisies en fonction des informations obtenues par l'analyse bioinformatique. La difficulté réside dans la détermination précise à l'acide aminé près, de ces bornes. Plusieurs règles empiriques (détaillées ci-après) permettent de prédire les limites de fragments qui auront une meilleure probabilité d'être solubles.

### Le rôle central des alignements multiples de séquences

Au cours de l'évolution, des événements de mutation surviennent aléatoirement modifiant les séquences des gènes qui codent pour des ARNs et des protéines. Les mutations sont susceptibles d'engendrer une cascade de modifications pouvant entraîner des altérations importantes. Plusieurs types de mutations existent, notamment, les mutations ponctuelles définies par la substitution d'un nucléotide de l'ADN par un autre (mutations faux-sens, non sens, silencieuses), ou par l'insertion/délétion d'un à plusieurs centaines de nucléotides. La plupart des mutations sont dites neutres. Elles ne modifient pas le fonctionnement de la protéine et n'ont pas de conséquences phénotypiques macroscopiques. Cependant, certaines mutations, par exemple la mutation de résidus du site actif d'une enzyme ou la mutation d'un résidu empêchant le repliement structural correcte de la protéine pourraient être à l'origine de conséquences pour le moins dramatiques en engendrant des protéines non fonctionnelles. Si ces protéines sont impliquées dans des processus biologiques nécessaires à la survie de la cellule, leur mutation peut être responsable de la mort de la cellule. Ainsi, au cours de l'évolution, les résidus fonctionnellement et structuralement importants pour la protéine subissent une forte pression de sélection et doivent être conservés. En comparant les séquences proches les unes des autres et en regardant les résidus qui restent les mêmes (conservés) dans tous les membres de la famille protéique, on peut ainsi déduire quels sont les résidus déterminants pour le bon fonctionnement de la protéine.

## Analyse de MACS

L'analyse *in silico* débute par la réalisation d'un MACS. La séquence protéique d'INI1 du plasmide initial est répertoriée dans la base de données UniProt sous le numéro d'accès Q12824 (<http://www.uniprot.org/uniprot/Q12824>). Cette séquence Q12824 est soumise à la cascade de programme PipeAlign. Le MACS brut établi par PipeAlign contient 195 séquences. L'optimisation du MACS a nécessité des ajustements manuels telle que l'élimination de séquences protéiques putatives non caractérisées et des processus cycliques tel que le rechargement de la cascade de programme PipeAlign à partir de l'étape DBClustal. Au final, le MACS est constitué de 82 séquences de longueur variable, avec une longueur minimum de 139 aa et maximum de 1109 (Figure 13). Les séquences protéiques SFH1 (SNF Five Homologue 1) homologues ont été conservées au sein du MACS.

Des études (Morozov, Yung et al. 1998; Yung, Sorin et al. 2001) ont montrées que les résidus T213, D224, D226 et S246 d'INI1 sont cruciaux pour l'interaction avec l'intégrase du VIH-1. Le domaine d'interaction a été défini entre P183 et Q243, ce qui exclu le résidu crucial S246. Le résidu D224 est conservé dans plus de 90 % des séquences du MACS. Pour T213 et D226, le consensus est établi à 59 %, enfin le résidu S246 n'est présent que dans 37,5 % des séquences.

## Les domaines : pFAM, ProDom et Morozov

Trois domaines distincts ont été prédits par le programme ProDom dont les limites sont 4-182 pour PD017359, 184-262 pour PD228116 et 266-385 pour PD007647. Le programme PRODOM génère automatiquement des commentaires pour chaque domaine. Les prédictions ne sont pas toutes de même consistance. Ainsi, l'alignement de séquences qui définit le domaine 4-182 comporte des séquences hétérogènes. A l'opposé, les deux autres domaines ont été déterminés à partir de séquences plus homogènes impliquant des limites de domaines mieux définies. La prédiction automatique de domaines pFAM-B détecte un domaine de 28-154 (domaine 4996). L'annotation manuelle pFAM-A donne deux domaines (178-255) et (250-373) et une région faiblement complexe (366-377). Les limites 106-183 représentent potentiellement un domaine de liaison à l'ADN (annotation SwissProt). D'après la littérature (Morozov, Yung et al. 1998; Cheng, Davies et al. 1999), les domaines, intimement entremêlés, de liaison à l'intégrase et à c-MYC ont été définis expérimentalement et les

limites respectives sont (183-243) et (186-245). Dans la même étude, une répétition imparfaite de ce domaine (259-319) a été mise en évidence. Trois domaines dont les limites varient sont définis :

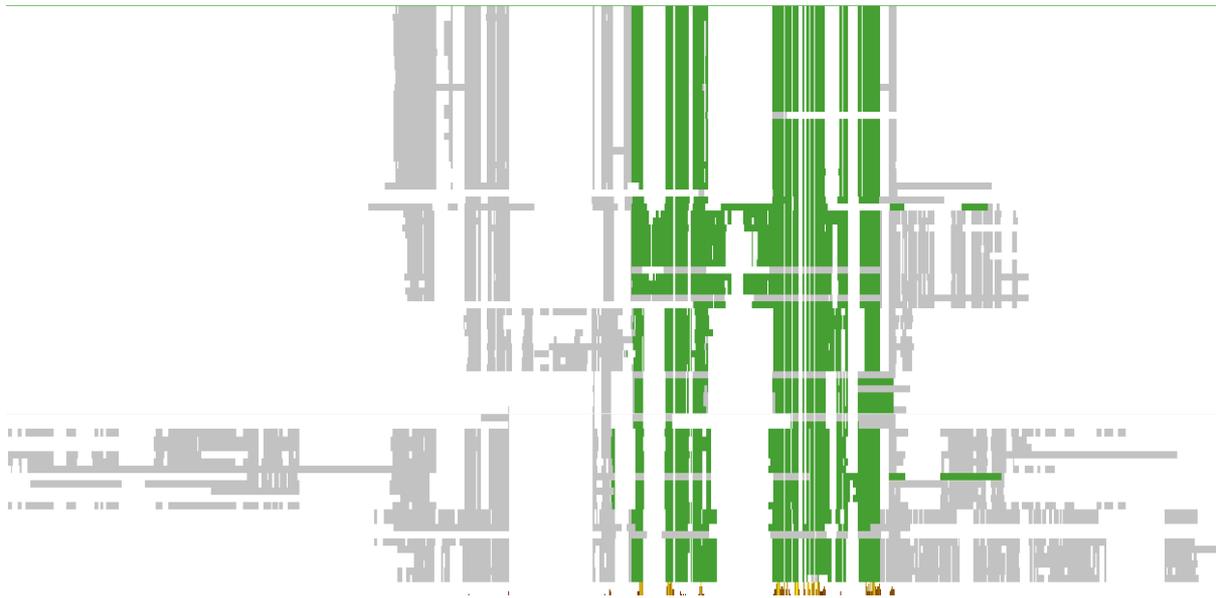


Figure 13 : Vue schématique du MACS d'INI1 (par Jalview). Le MACS a été construit à partir de la séquence requête Q12824 (SwissProt). Le domaine pFAM SNF5 (domaine de liaison à l'intégrase) est coloré en vert et étendu à l'ensemble du MACS par le programme MACSIMS. Les résidus conservés sont listés sous forme d'histogramme.

### Les bornes

Les bornes des domaines ont été définies à partir de l'analyse *in silico*. Le découpage d'INI1 m'a permis d'obtenir cinq extrémités en N-ter et cinq en C-ter.

#### La borne P74

La première borne du fragment devait contenir le domaine potentiel de liaison à l'ADN (106-183). La région 69-77 est absente dans l'isoforme B d'INI1. Les résidus 74 à 78 sont prédits non structurés mais ayant une très bonne accessibilité au solvant (PROF(acc)). La région suivant la borne P74 est prédite hydrophile et structurée.

Une alternative aurait été de commencer le fragment à G102 au milieu d'une région hydrophile, juste avant le domaine de liaison à l'ADN, ayant des prédictions de désordre moindre. Cependant la région suivante (110-117) est prédite comme désordonnée (KoAnno).

### *La borne P165*

La borne P165 a été définie comme une limite alternative contenant le domaine de liaison à l'intégrase du VIH-1 (183-243). La limite a été déterminée en fonction de la prédiction de désordre (KoAnno : 3 sur 4 prédisent une région ordonnée et hydrophile entre 165 et 173). Cependant les prédictions de structures secondaires ne dégagent pas de consensus. La limite a été définie au niveau de la P165, acide aminé connu pour être présent au début des structures secondaires. Cette proline est conservé dans 31 % des séquences du MACS.

Au sein du MACS dans la famille d'INI1, la protéine de numéro d'accès Q9H836 (SwissProt) comporte une insertion de 17 résidus à cette position. La séquence Q9H836 est tirée de la collection « FLJ » d'ADNc humain (Ota, Suzuki et al. 2004). La séquence Q9H836 diffère par rapport à Q12824 par : deux mutations ponctuelles (N154D, C168H), la caractéristique de l'isoforme B (région 69-77 manquante) et la présence en P165 d'une insertion de 17 aa.

### *La borne P173*

La borne P173 a été définie comme contenant le domaine de liaison à l'intégrase du VIH-1 (183-243). Cette proline se retrouve dans 30 % des séquences du MACS juste après une région hydrophobe et avant une région hydrophile. Toutes les prédictions de structures secondaires indiquent le commencement d'une hélice  $\alpha$  de même longueur (173-178). A la fin de cette hélice  $\alpha$  prédite, débute, en position 178, le domaine pFAM-A SNF5 (178-255), celui-ci sera contenu par le fragment commençant en P173. La limite P173 donne la meilleure prédiction.

### *La borne P183*

La borne P183 est tirée d'études préalables (Morozov, Yung et al. 1998) et se situe au début d'une région hydrophile. Les prédictions de structures secondaires conjecturent un feuillet beta dont la limite n'est pas unanime (Garnier E184 ; ASSP L186 ; PROF L186) et une structure secondaire non régulière (NORS). La limite P183 se situe juste avant une région (186-194) ayant une forte conservation et un fort consensus (>60 %). Les domaines définis à partir de P183 sont trois domaines entremêlés : le domaine de liaison à l'intégrase du VIH-1 (183-243), le domaine de répétition imparfaite (Repeat Domain) (186-245) et le domaine de liaison à c-MYC (186-245).

### *La borne P188*

Comme alternative à la limite P183, la borne P188, constitue en termes de prédiction du désordre une alternative importante. En effet, l'ensemble des prédictions de KoAnno détermine une région structurée. De plus, le programme Blocks propose une région d'alignement local (187-227). Le consensus du résidu P188 est le troisième plus élevé du MACS avec plus de 96 % indiquant une forte probabilité d'un résidu important. Le résidu P188 se situe juste après la prédiction d'une structure secondaire non conventionnelle (NORS).

### *La borne Q243*

La borne Q243 constitue la fin du domaine 183-243 de liaison à l'intégrase du VIH-1. Cette limite exclue cependant le résidu S247 important pour l'interaction (Yung, Sorin et al. 2001). En effet, l'interaction avec l'intégrase est plus faible lorsque la construction est limitée au domaine d'interaction 183-243 (Morozov, Yung et al. 1998). Les limites 183-243 définissent comme le domaine minimal d'interaction mais pas l'optimal. La borne Q243 termine une région prédite hydrophile et structurée par une structure secondaire en hélice. L'utilité de la limite Q243 réside essentiellement dans la possibilité de moduler l'affinité d'interaction avec l'intégrase et d'en mesurer les conséquences structurales et fonctionnelles.

### *La borne Q257*

La borne Q257 comprend tous les résidus définis comme importants pour l'interaction avec l'intégrase (Yung, Sorin et al. 2001). Cette borne a été définie à l'extrémité d'une région hydrophobe et prédite comme structurée (KoAnno). La limite Q257 contient le domaine pFAM SNF5 (178-255) ainsi que les domaines de répétition imparfaite (Repeat Domain) (186-245) et de liaison à c-MYC (186-245). Enfin, dans le groupe 2 du MACS, une insertion d'un total de 105 résidus est présente après le résidu 257.

### *La borne P290*

La borne P290 se situe en bordure d'une région hydrophobe prédite comme accessible au solvant (PROF(acc)). Le programme Blocks situe une région d'alignement local fortement conservée entre 259 et 290. Le consensus de P290 est de 78 % et est conservé au sein du

MACS. Cette borne contient, en plus du domaine de liaison à l'intégrase, la moitié du domaine « *approximate tandem repeats* » (répétition approximative en tandem).

### *La borne G302*

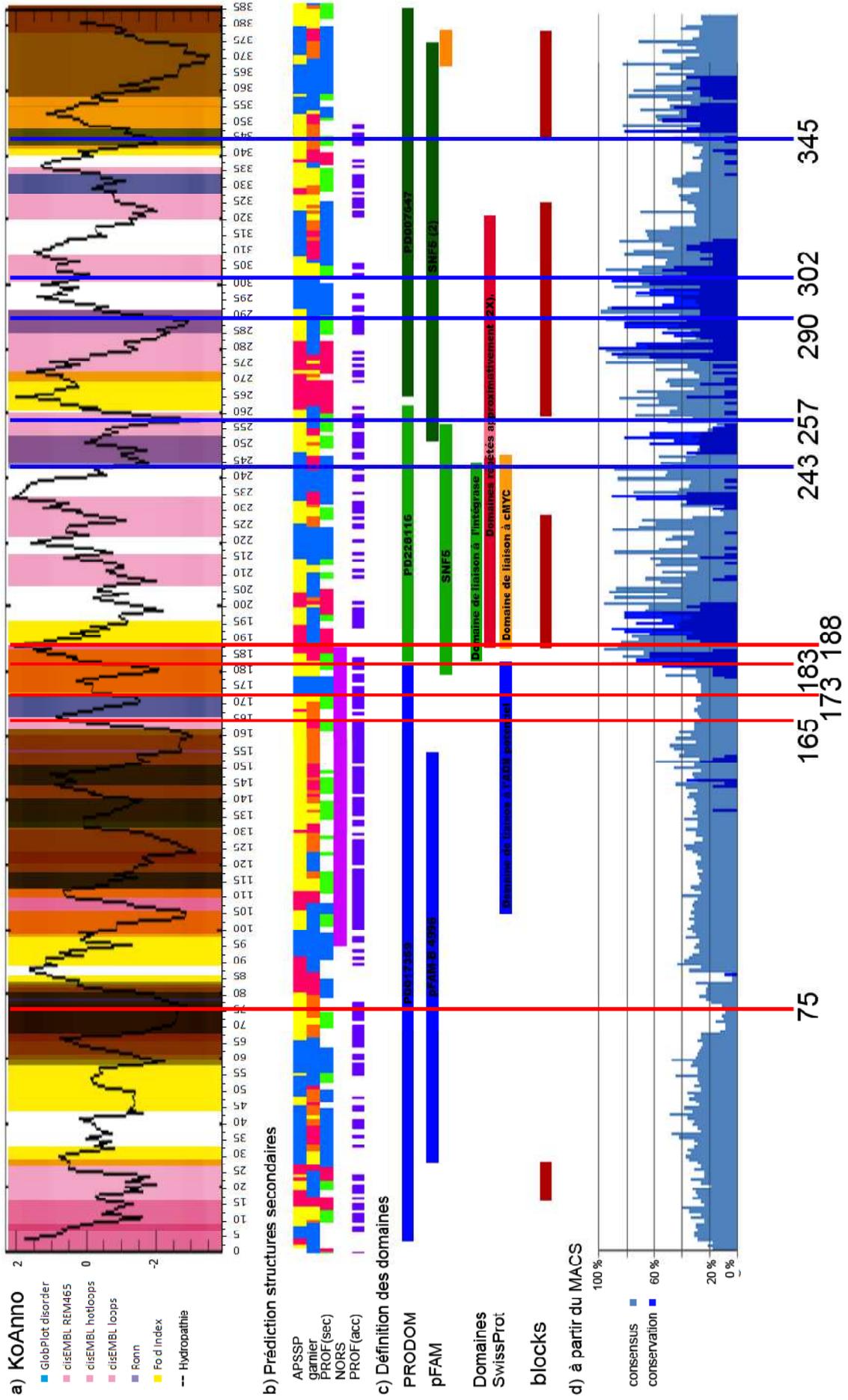
La borne G302 se situe à la fin de prédiction de structure secondaire, dans une région ordonnée (KoAnno) et hydrophile. Cette borne contient les deux tiers du domaine « *approximate tandem repeats* ».

### *La borne G345*

La borne G345 se situe après la prédiction d'une région ordonnée et avant une région prédite désordonnée. Cette borne contient la totalité du domaine 186-319 « *approximate tandem repeats* ». Les trois résidus précédant G345 sont prédits désordonnés mais accessibles au solvant, évitant de s'achever par une partie trop hydrophobe.

La figure 14 sur la page suivante montre la compilation des résultats de l'analyse *in silico*.

Figure 14 : Résultats de l'analyse *in silico* de la séquence d'INI1 (Q12824) : prédictions de désordre, de structures secondaires d'accessibilité au solvant, de domaines. \_a) Pour KoAnno (prédiction du désordre), la représentation graphique montre la corrélation des différentes prédictions en différentes couleurs (bleu : GlobPlot Disorder ; rose : disEMBL ; lavande : Ronn ; jaune : Fold Index). Plus les régions sont sombres, plus le nombre de prédictions de désordre concordantes est important. La courbe noire représente le degré d'hydropathie des acides aminés ramené à une échelle entre 2 et -2 (pepwindowall EMBOSS). \_b) Les prédictions de structures secondaires sont représentées en fuchsia : feuillet  $\beta$  ; en bleu : hélice  $\alpha$  ; en jaune : coil ; en vert : loop ; en orange : turn ; en violet : régions mobiles. \_c) Sur cette figure sont représentés par des rectangles les domaines ProDom, pFAM, les domaines définis dans la littérature, ainsi que les alignements locaux (blocks). L'histogramme montre les pourcentages de conservation (en bleu foncé) et de consensus (en bleu ciel) de la séquence d'INI1 extraite du MACS. Les limites des différentes constructions sont représentées par des traits verticaux rouges pour en N-ter et bleus en C-ter.



## Clonage et tests d'expressions

### Matériels et méthodes

#### La préparation de bactéries chimiocompétentes

Deux souches d'*Escherichia coli* (DH5 $\alpha$  et BL21-DE3) ont été utilisées pour effectuer les transformations. La première souche (DH5 $\alpha$  (Invitrogen™) : F<sup>-</sup>  $\Phi$ 80*lacZ* $\Delta$ M15  $\Delta$ (*lacZYA-argF*)U169 *deoR recA1 endA1 hsdR17*(r<sub>K</sub><sup>-</sup>,m<sub>K</sub><sup>+</sup>) *phoA supE44 thi-1 gyrA96 relA1*  $\lambda$ ) est destinée à l'amplification plasmidique a été utilisé lors des différentes étapes de clonage. La préparation de bactéries compétentes a été réalisée selon la méthode INOUE. La seconde souche (BL21 (DE3) (Novagen) : F<sup>-</sup> *ompT hsdS<sub>B</sub>*(r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>) *gal dcm* ( $\lambda$ *Its857, ind1, Sam7 nin5 lacUV5-T7 gene1*)) est destinée à l'expression de protéines recombinantes. Dans le génome bactérien de cette souche se trouve le lysogène du bactériophage  $\lambda$  DE3 qui contient le gène codant pour l'ARN polymérase du phage T7, sous le contrôle de l'opéron lactose inductible à l'IPTG. La préparation de bactéries chimiocompétentes s'est effectuée selon la methode CaCl<sub>2</sub>.

Méthode Inoue (Inoue, Nojima et al. 1990): bactéries chimio-compétentes DH5 $\alpha$ .

Une pré-culture de 100 ml de milieu LB est inoculée à partir d' un stock glycérol et incubée à 37°C sous agitation constante pendant la nuit. La pré-culture est refroidie pendant 8 heures à 4 °C. Une culture de 250 ml de milieu SOB estensemencée par 4 ml de la pré-culture (afin d'avoir une densité optique à 600 nm initiale 0,1) puis incubée à 18 °C sous agitation constante (190 rpm) jusqu'à atteindre une DO<sub>600nm</sub> de 0,4. La culture est ensuite refroidie 10 minutes sur la glace, puis centrifugée pendant 15 minutes à 2500 g à 4 °C. Le culot est resuspendu dans 40 ml de milieu TB froid. Ce mélange est incubé 10 minutes sur glace puis centrifugé pendant 15 minutes à 2500 g à 4 °C. Le culot est resuspendu dans 10 ml de tampon TB froid. Du DMSO est ajouté goutte à goutte à une concentration finale de 7 % (V/V). Ce mélange est laissé 10 minutes sur glace, puis aliquoté en fractions de 50  $\mu$ l. Les fractions aliquotes sont finalement congelées dans l'azote liquide et conservées a -80 °C.

Méthode au CaCl<sub>2</sub> (Dagert and Ehrlich 1979): bactérie chimio-compétentes BL21.

Une pré-culture de 100 ml de milieu LB est inoculée à partir d'un stock glycérol et incubée à 37 °C sous agitation constante pendant la nuit. Le lendemain, une culture d'un litre de milieu LB estensemencée avec 20 à 25 ml de cette pré-culture (afin d'avoir une DO à 600 nm

initiale comprise entre 0,05 et 0,1), incubée à 37 °C sous agitation constante (190 rpm) jusqu'à atteindre une DO à 600 nm de 0,4. La culture est ensuite refroidie 20 minutes sur glace puis centrifugée pendant 15 minutes à 2500 g à 4 °C. Le culot est resuspendu dans 250 ml d'une solution de NaCl à 10 mM. Ce mélange est incubé 10 minutes sur glace puis recentrifugé pendant 15 minutes à 2500 g à 4 °C. Le culot est repris dans 250 ml de CaCl<sub>2</sub> 75 mM et incubé sur glace pendant 35 minutes. Le mélange est centrifugé 15 minutes à 2500 g à 4 °C, le culot est repris dans 15 ml de CaCl<sub>2</sub> 75 mM. Après ajout de trois ml de glycérol 87 %, la solution est homogénéisée, incubée 20 minutes dans la glace puis dispensée en fractions aliquotes de 50 et 100 µl. Les fractions aliquotes sont finalement congelées à -80 °C.

### **Transformation**

Un aliquot de 50 ng de plasmide est ajouté de manière stérile à 50 µl de bactéries chimiocompétentes préalablement décongelées sur glace. Les bactéries sont laissées une minute sur glace, puis un choc thermique d'une minute à 42 °C permet l'ouverture des pores bactériens et l'entrée des plasmides dans les bactéries. La fermeture des pores se fait par une incubation d'une minute sur glace, puis les bactéries sont régénérées par ajout de 200 µl de milieu LB 2X. Une sélection sur milieu solide est ensuite réalisée. Pour cela, entre 100 et 200 µl du mélange sont étalés sur boîte de Pétri contenant du milieu LB agar avec l'antibiotique adéquat. La boîte de Pétri est mise à 37 °C pendant une nuit.

### **Purification de plasmide : mini-préparations d'ADN**

L'utilisation du kit de purification d'ADN plasmidique (NucleoSpin® Plasmid de chez Macherey-Nagel) permet, à partir de 5 mL de culture bactérienne spécifique (*E.coli* DH5α) transformée avec le plasmide d'intérêt, d'extraire 30 µg d'ADN plasmidique. A partir d'une colonie isolée prélevée sur une boîte de pétri (LB agar), on inocule 5mL de milieu LB supplémenté de la quantité d'antibiotique adéquat. Généralement, la solution stock d'antibiotiques est 1000x concentré, donc il est nécessaire d'ajouter 5 µL de cette solution. La culture est incubée à 37 °C 200 rpm pendant 12 à 14 heures. Les cellules sont récoltées par centrifugation 4000 g pendant 10 min à 4 °C. Le surnageant est éliminé et le culot bactérien est séché sur un papier absorbant.

## Le clonage

L'utilisation des ressources de la plateforme a permis d'opter pour une stratégie haut débit. Toutes les séquences définies par l'analyse bioinformatique ont été clonées (en utilisant la stratégie Gateway®) en parallèle dans des vecteurs d'entrée (pENTR). A partir des vecteurs d'entrée, ces séquences sont transférées par recombinaison dans différents vecteurs d'expression comportant différentes fusions pour la purification d'affinité et/ou pour augmenter la solubilité (GST, MBP et 6His).

### *La stratégie Gateway© (Invitrogen™ 2003)*

La technologie de clonage Gateway© (Invitrogen™) (Figure 15) repose sur des réactions de recombinaison conservatrice et directionnelle entre séquences ADN spécifiques du bactériophage lambda (sites *att*). Le système Gateway© est basé sur deux types de vecteurs : les vecteurs d'entrée et les vecteurs d'expression aussi appelé de destination. Le vecteur d'entrée obtenu par recombinaison spécifique à partir de la séquence nucléotidique d'intérêt et d'un vecteur donneur permet sa conservation au sein d'une banque de plasmides. Par la suite, il est possible de choisir parmi les différents vecteurs de destination pour générer (par recombinaison spécifique à partir du vecteur d'entrée contenant la séquence d'intérêt et un des différents vecteurs de destination) des vecteurs d'expression pour produire la protéine recombinée avec différent tag ou un autre système d'expression selon le vecteur de destination.

Un clonage Gateway© nécessite trois étapes. La création du fragment d'intérêt fusionné aux sites de recombinaison. Cette étape nécessite une ou deux PCR selon l'utilisation d'amorces comprenant un des sites de recombinaison *att*, un site de coupure d'une protéase et une des bornes de la séquence nucléotidique de la protéine à cloner. Ensuite, la recombinaison du produit de PCR obtenu dans un plasmide d'entrée. Enfin, la recombinaison entre le plasmide d'entrée et un ou plusieurs vecteurs d'expression.

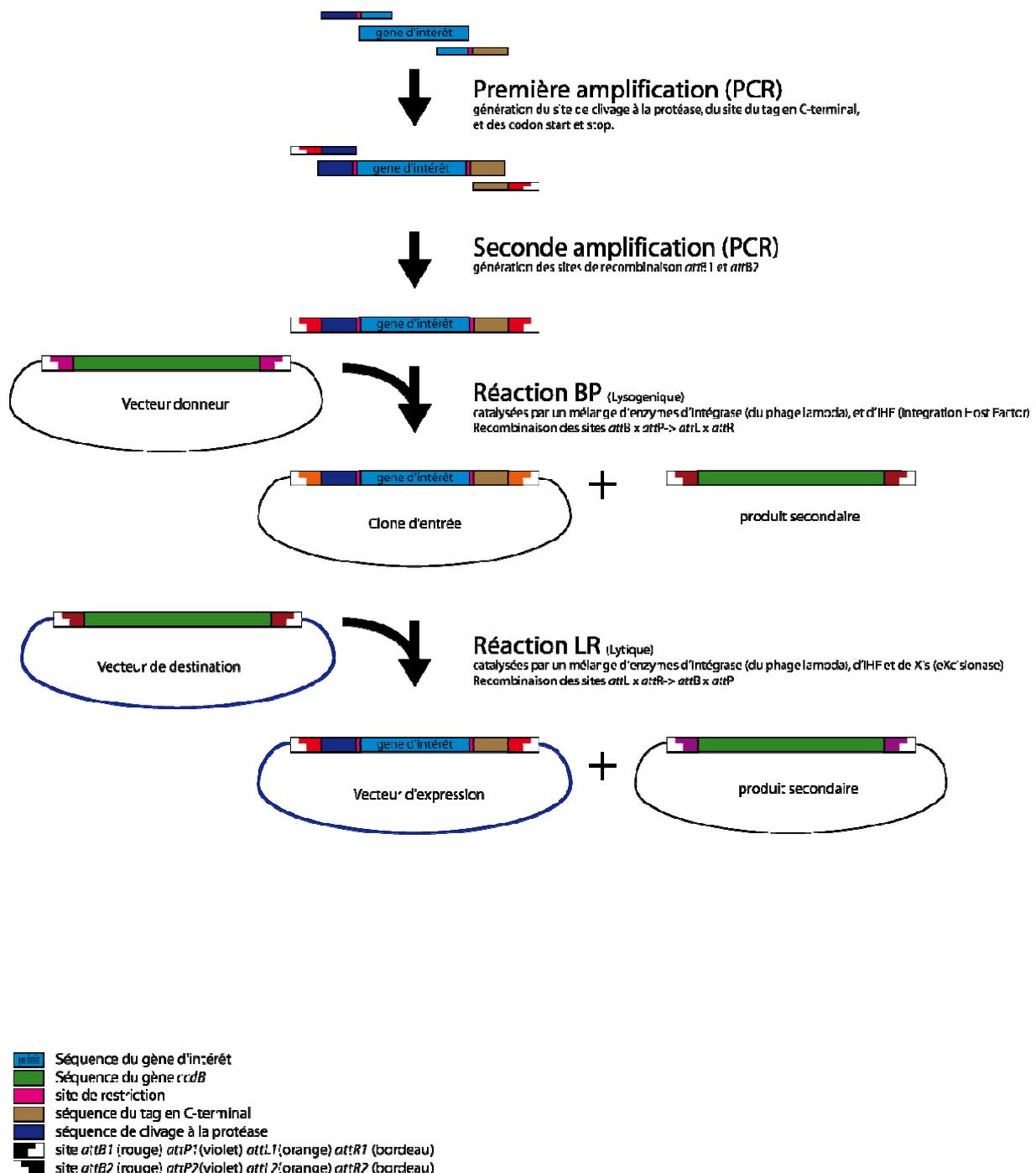


Figure 15 : Récapitulatif de la stratégie Gateway

Le gène *ccdB* permet la sélection négative des vecteurs de destination et donneur et sélection positive des vecteurs d'expression ayant intégrés le gène d'intérêt après la recombinaison et la transformation. Les produits secondaires et les plasmides n'ayant pas subi la réaction BP ou LR, expriment la protéine CcdB qui interfère avec l'ADN gyrase d'*E. coli* en inhibant la croissance des bactéries (Bernard and Couturier 1992).

## Le protocole Gateway

Design des amorces : le BrocOli

La réalisation d'un clonage Gateway® nécessite l'utilisation de quatre amorces : deux spécifiques du gène d'intérêt (sens et anti-sens), deux génériques (sens et anti-sens) indépendants de l'insert et recouvrant la séquence attB. Pour la première amplification, l'amorce spécifique sens comporte la séquence codant pour un site de clivage à une protéase, un site de restriction et 18 à 25 nucléotides spécifiques de la cible. L'amorce spécifique anti-sens comporte, dans l'orientation 5' vers 3', la séquence attB2, d'éventuels signaux et 18 à 25 nucléotides spécifiques de la séquence d'intérêt. Pour la seconde amplification, les amorces génériques sens et anti-sens contiennent respectivement les sites attB1 et attB2.

Séquences des amorces génériques utilisées :

GWS (« sens ») 5'-GTACAAAAAAGCAGGCTTC-3'

GWR (« anti-sens ») 5'-GTACAAGAAAGCTGGGTC-3'

### Amplification des domaines par PCR

Cette méthode enzymatique très fréquemment utilisée est efficace pour amplifier des fragments d'ADN *in vitro*. En général, une PCR réalisée dans un volume final de 20 µl suffit à obtenir suffisamment de matériel nécessaire à la réaction BP. L'amplification est réalisée par l'intermédiaire de deux PCR consécutives en utilisant principalement la Phusion ADN polymérase (2 U/µl) (Finnzymes) selon le protocole décrit ci-dessous. Une première PCR est réalisée dans un volume de 20 µl avec les amorces spécifiques sens anti-sens. A 1 µl d'ADN matriciel (10 ng) sont ajoutés 4 µl de tampon HF 5X, 0,25 µl de chaque oligonucléotide à 10 µM, 2 µl de dNTP+ (2 mM chacun), 0,4 U de Phusion ADN polymérase et de l'eau ultra-pure. Les conditions de la PCR sont les suivantes : une première dénaturation thermique des deux brins d'ADN durant 30 sec à 98°C, une amplification de 20 cycles comprenant une étape de dénaturation thermique à 98°C pendant 7 sec, suivie d'une étape d'hybridation des amorces à 55°C pendant 20 sec et une étape d'élongation des amorces à 72°C dont la durée dépend de la taille du fragment à amplifier (15 sec/kpb). Une élongation finale à 72°C pendant 7 min termine la PCR. La seconde PCR se déroule dans les mêmes conditions en utilisant les amorces génériques sens et anti-sens et 1µl d'ADN matriciel provenant de la

première PCR. Les conditions de la PCR sont identiques, excepté pour le nombre de cycles augmenté à 35. La réaction PCR est vérifiée en déposant 2 µl de la PCR (+1 µl de bleu de charge) sur gel d'agarose 1% (poids/vol) contenant 1 µl de BET (10 mg/ml) pour 100 ml de tampon TAE. Dans le cas d'un échec de l'amplification avec la Phusion ADN polymérase, l'Expand High Fidelity ADN polymérase (3,5 U/µl) de Roche Diagnostics® permet d'amplifier le fragment escompté. La PCR est relancée dans un volume réactionnel de 20 µl selon le protocole suivant. A 1 µl d'ADN matriciel (50 ng) sont ajoutés 2 µl de tampon Expand HiFi 10X, 0,7 µl d'amorce spécifique sens (10 µM), 0,7 µl d'amorce anti-sens (10 µM), 2 µl de dNTP (2 mM chacun), 0,5 µl d'Expand High Fidelity ADN polymérase et de l'eau ultra-pure. En option, 2 à 5 % final de DMSO (Dimethyl sulfoxide) peuvent être ajoutés. Le DMSO inhibe les hybridations non spécifiques et déstabilise les structures secondaires de la matrice ou des amorces. Les conditions de la PCR sont les suivantes : une première dénaturation thermique des deux brins d'ADN durant 4 min à 94°C. Une première amplification de 10 cycles comprenant une étape de dénaturation thermique à 94°C pendant 1 min, suivie d'une étape d'hybridation des amorces à 55°C pendant 45 sec et une étape d'élongation des amorces à 72°C pendant 1 min. Une élongation finale à 72°C pendant 7 min termine la PCR. Une deuxième amplification de 25 cycles comprenant une étape de dénaturation thermique à 94°C pendant 1 min est réalisée, suivie d'une étape d'hybridation des amorces à 55°C pendant 45 sec et une étape d'élongation des amorces à 72°C pendant 1 min. A chaque cycle, le temps d'élongation est augmenté de 5 sec. Une élongation finale à 72°C pendant 7 min termine la PCR.

#### Purification des produits PCR

Les produits PCR sont purifiés, notamment dans le but d'éliminer les amorces attB et les dimères d'oligonucléotides attB. Les dimères d'oligonucléotides peuvent parasiter la réaction BP en se recombinant avec le vecteur donneur. Le kit NucleoSpin® ExtractII (Macherey-Nagel) est utilisé directement à partir de la réaction PCR. Avec la méthode NucleoSpin ExtractII, l'ADN (1 vol) est fixé sur une membrane en silice d'une colonne NucleoSpin ExtractII en présence de sels chaotropiques (guanidine) par centrifugation à 11000 g pendant 1 min. Les contaminants sont éliminés par une simple étape de lavage avec 600 µl de tampon commercial NT3 contenant de l'éthanol et par centrifugation à 11000 g pendant 1 min. Une deuxième centrifugation à 11000 g pendant 2 min permet de bien sécher la membrane et d'éliminer toute trace d'éthanol. Finalement, l'ADN purifié est élué dans des conditions de

force ionique faible avec 15 à 50 µl de tampon NE (Tris-HCl 5 mM pH=8,5) par une centrifugation à 11000 g pendant 1 min. Les produits PCR purifiés sont quantifiés par mesure de la densité optique à 260nm avec un coefficient d'absorbance de 0,02 ng-1 cm-1 µL en utilisant le spectrophotomètre ND-1000 NanoDrop® (NanoDrop).

#### Réaction BP : Générer le clone d'entrée (pENTR)

La réaction BP se déroule dans un volume final de 5 µl en présence de 10 à 30 fmol de produit PCR, 100 ng de pDONR207 (vecteur donneur gentamycine résistant), 1 µl de tampon BP 5X, 1 µl de BP Clonase mix™ (Invitrogen) et de l'eau ultra-pure. Pour favoriser la réaction de recombinaison, le mélange est incubé toute la nuit à température ambiante. La réaction est stoppée par l'ajout de 1,2 µg de Protéinase K qui dégrade la BP Clonase en 10 min à 37°C. A ce stade, la réaction BP peut se conserver à -20°C avant transformation.

#### Transformation BP

Toutes les étapes suivantes sont effectuées en conditions stériles. La totalité de la réaction BP est ajoutée à 50µl de bactéries DH5α chimio-compétentes préalablement décongelées sur la glace. Le mélange est incubé durant 30min sur glace. Les bactéries sont transformées par choc thermique à 42°C pendant 1min. Après 2min d'incubation sur glace, 150µl de milieu LB sont ajoutés et les bactéries sont mises à incuber à 37°C pendant 1h sous agitation. Toute la réaction de transformation est étalée sur boîte de Pétri standard contenant du LB agar complétement de gentamycine (7 µg/ml final). La boîte est placée à 37°C toute la nuit.

#### Vérification de la présence du gène d'intérêt par PCR

Cinq colonies sont repiquées dans 5 ml de milieu LB complétement de gentamycine (7 µg/ml final). Après incubation à 37°C pendant la nuit sous agitation constante, ces minicultures pourront être utilisées pour les mini-préparations d'ADN. La PCR se déroule selon le protocole suivant en utilisant la Taq ADN polymérase (NEB) (5 U/µl). La réaction se déroule dans un volume final de 20 µl en présence de 1,5 µl de la mini-culture, 2 µl de tampon Thermo Pol 10X, 2 µl de dNTP (2 mM chacun), 0,5 µl de Taq ADN polymérase, 1 µl de chacun des amorces génériques GWS et GWR à 10 µM et de l'eau ultra-pure. Les conditions de la PCR sont les suivantes : une première dénaturation thermique des deux brins d'ADN durant 4 min à 94°C. Une amplification de 20 cycles comprenant une étape de dénaturation

thermique à 94°C pendant 1 min, suivie d'une étape d'hybridation des amorces à 55°C pendant 45 sec et une étape d'élongation des amorces à 72°C dont la durée dépend de la taille du fragment à amplifier (1 min/kpb). Une élongation finale à 72°C pendant 5 min termine la PCR. La réaction PCR est vérifiée en déposant 8 µl de la PCR (+2 µl de bleu de charge) sur gel d'agarose 1% (poids/volume) contenant 1 µl de BET (10 mg/ml) pour 100 ml de tampon TAE. Si la présence de l'insert est confirmée, la mini-culture est utilisée pour une mini-préparation d'ADN suivant le protocole présenté auparavant.

#### Séquençage du clone d'entrée

Les bornes des produits PCR sous-clonés, au niveau des sites de recombinaison, ainsi que le début et la fin de la séquence du gène d'intérêt, sont vérifiés par un criblage PCR en utilisant les amorces A1 (sens : 5'-TCGCGTTAACGCTAGCATGGATCTC-3') et A2 (anti-sens : 5'-GTAACATCAGAGATTTTGAGACAC-3') pour déterminer les clones positifs ayant intégrés le gène d'intérêt (la taille du produit PCR est vérifié sur gel d'agarose 1%). Les clones positifs sont séquencés par l'entreprise GATC basée à Constance, Allemagne (<http://www.gatc-biotech.com/en/index.html>). Pour une réaction de séquençage, 20 µl d'ADN plasmidique à une concentration de 100 ng/µl et 10 µl des amorces à 2 pmol/µl sont nécessaires.

#### Réaction LR : générer les vecteurs de destination (pDEST)

La réaction LR se déroule dans un volume final de 5µl et requiert 100ng de pDEST, 100ng de pENTR, 1µl de tampon LR™ (Invitrogen) 5X, 1µl de LR Clonase mix™ (Invitrogen) et du tampon TE pH=8. Le mélange est incubé au moins 1h à température ambiante. Puis la réaction est stoppée par l'ajout de 1,2µg de Protéinase K qui dégrade la LR Clonase. Ce mélange est incubé à 37°C pendant 10min. A ce stade, la réaction LR peut se conserver à -20°C pendant une semaine avant transformation.

#### Transformation LR

Le protocole est exactement le même que celui précédemment décrit pour la réaction BP. On ajoute 1 µl de la réaction LR à 50 µl de bactéries DH5α chimio-compétentes. Si peu, ou aucune colonie, n'est obtenue, il est conseillé de transformer le reste de la réaction LR.

## Vérification de la présence de la fusion en N-ter et du gène d'intérêt

Deux colonies sont repiquées dans 5 ml de milieu LB complété de l'antibiotique adéquat et incubées à 37°C pendant la nuit sous agitation constante. Ces mini-cultures pourront être utilisées pour les mini-préparations d'ADN. La PCR se déroule selon le même protocole que précédemment en utilisant la Taq ADN polymérase (NEB) (5 U/μl) et les amorces UP1 (sens : 5'-GATGCGTCCGGCGTAG-3') et GWR (anti-sens : 5'-GTACAAGAAAGCTGGGTC-3'). La réaction PCR est vérifiée en déposant 8 μl de la PCR (+2 μl de bleu de charge) sur gel d'agarose 1% (poids/volume) contenant 1 μl de BET (10 mg/ml) pour 100 ml de tampon TAE.

## Tests de mini-expression

Les tests d'expression et le criblage des conditions de solubilité sur les vecteurs d'expression bactériens permettent de repérer les meilleurs vecteurs d'expression en termes d'expression de protéine et de solubilité de celle-ci. Un des principaux inconvénients de cette technique est de produire une proportion significative de faux positif.

## Inoculation

Après transformation des vecteurs d'expression dans des bactéries chimio-compétentes, la sélection s'opère pour chaque transformation par étalement sur milieu solide (LB agar et antibiotique) dans un puits de plaques 6 puits. Puis les plaques sont incubées une nuit à 37 °C. Dans une plaque 96 puits, la pré-culture s'effectue dans 500 μL de milieu LB avec antibiotique inoculé par une colonie isolée piquée dans l'agar des plaques 6 puits. Cette microplaque 96 puits est recouverte d'une membrane respirant (« AirPore Sheet ») et incubée à 37 °C. Après une nuit, la DO595nm est mesurée (en diluant au 1/5<sup>ème</sup>) pour chaque puits à l'aide du lecteur de plaque Genios™ (TECAN® Ltd).

## Production

Dans des plaques 24 puits, 2 mL de milieu auto-inductible ZYP-5052 sont dispensés dans chaque puits. Le principe du milieu auto-inductible est d'avoir deux sources d'énergies différentes (le glucose et le lactose) qui vont être utilisées séquentiellement. La première, le glucose plus facilement métabolisable, permet aux bactéries de se multiplier. Après

l'appauvrissement du milieu en glucose, le lactose induit la production des gènes sous dépendance de l'opéron lactose notamment le gène de la protéine d'intérêt. On inocule tous les puits par un volume de pré-culture permettant d'obtenir une DO595nm de 0,1 dans le chaque puits. Pour contrôle certains puits sont dupliqués. Les plaques 24 puits sont incubés à 37 °C pendant 3 h jusqu'à atteindre une DO595nm de 1 à 1,5 puis à 25 °C pendant 20 h. Après une dilution au dixième, la DO595nm est mesurée pour chaque puits à l'aide du lecteur de plaque Genios™ (TECAN® Ltd). Les plaques sont centrifugées à 4000 rpm pendant 20 min, puis le culot est stocké à -20 °C.

### *Lyse*

Les culots sont resuspendus dans un volume de tampon de lyse. Le volume à ajouter est fonction de la mesure de la densité optique à 595nm ( $(\text{Volume de tampon de lyse en } \mu\text{L}) = (\text{DO595nm}/20) * (\text{Volume culture en } \mu\text{L})$ ). Afin de lyser les bactéries, la sonication est effectuée à l'aide d'une sonde 24 têtes de 3mm, à une amplitude de 20 %, par pulse de 2 secondes pendant une minute. 10  $\mu\text{L}$  (extrait total) sont prélevés dans chaque puits pour l'analyse. Après centrifugation (45 min 4000 rpm 4°C), on prélève derechef 10  $\mu\text{L}$  (extrait soluble), afin de déterminer quelles protéines sont restées en solution et quelles protéines ont précipitées.

### *Analyse*

L'analyse s'effectue par électrophorèse sur gel de polyacrylamide en conditions dénaturantes (SDS-PAGE). Pour chaque vecteur testé, plusieurs dépôts sont effectués :

- \_ Le témoin avant induction (avant l'induction) permet de voir s'il n'y a pas de fuite du vecteur et de comparer contre le témoin après induction.
- \_ Le témoin après induction (en fin de culture) permet de voir d'évaluer le niveau d'expression de la protéine produite.
- \_ L'extrait total est le témoin pour les protéines exprimées ce qui permet la comparaison contre l'extrait soluble.
- \_ L'extrait soluble permet d'évaluer la solubilité des protéines par comparaison contre l'extrait total.

## Résultats

Pour INI1, les séquences des différents domaines délimités par l'étude *in silico* ont été transférées à partir du même plasmide initial par recombinaison (stratégie Gateway®) dans différents vecteurs d'expression. Ces vecteurs comportent différentes fusions pour la purification d'affinité et/ou pour augmenter la solubilité de la protéine d'intérêt (GST, MBP et 6His). Actuellement, la banque des fragments d'INI1 comprend 75 vecteurs d'expression pour 15 constructions d'INI1 différentes (Figure 17). Les tests d'expression dans *E.coli* BL21 DE3 et le criblage des conditions de solubilité ont permis de dégager deux fragments solubles : INI1 (173-257) et INI1 (173-290) (Figure 16). En effet, ces deux fragments sont les seuls à être solubles en l'absence de tag et avec un tag hexahistidine (6His). Cependant, il existe une perte non négligeable de protéine entre la fraction totale (après sonication) et la fraction soluble (après centrifugation), ce qui indique que le tampon 150mM NaCl 50mM TrisHCl pH8.0 2mM βmercaptoéthanol n'est pas optimal et nécessite des ajustements afin d'augmenter la solubilité des fragments sélectionnés. Ces ajustements du tampon de solubilité seront effectués après l'optimisation de la production à l'échelle du litre.

	74/ 257	74/ 290	74/ 302	74/ 345	165/ 257	165/ 290	165/ 302	165/ 345	173/ 257	173/ 290	173/ 302	173/ 345	141/ 302	141/ 304	183/ 243	188/ 243
notag	bleu ciel	bleu ciel	bleu ciel	bleu ciel	bleu nuit	bleu ciel	bleu ciel	bleu ciel	orange	orange	bleu ciel	bleu ciel	bleu ciel	bleu ciel	bleu ciel	bleu ciel
GST-Tag	bleu ciel	bleu ciel	bleu ciel	bleu ciel	orange	bleu ciel	orange	bleu ciel	orange	bleu ciel	bleu ciel	bleu ciel				
His-Tag	bleu ciel	bleu ciel	bleu ciel	bleu ciel	bleu nuit	bleu nuit	bleu nuit	bleu ciel	orange	orange	bleu ciel	bleu ciel	bleu ciel	bleu ciel	bleu ciel	bleu ciel
His- MBP-Tag	orange	orange	orange	orange	orange	orange	orange	orange	orange	orange	orange	orange				
	74/ 257	74/ 290	74/ 302	74/ 345	165/ 257	165/ 290	165/ 302	165/ 345	173/ 257	173/ 290	173/ 302	173/ 345	141/ 302	141/ 304	183/ 243	188/ 243

Figure 16 : Tableau récapitulatif de l'expression et de la solubilité des différents fragments d'INI1 en fonction du tag. En bleu ciel = exprimé dans E coli BL21 DE3 ; bleu nuit = non exprimé dans E coli BL21 DE3; orange = exprimé et soluble dans 150mM NaCl 50mM TrisHCl pH 8.0 2mM βmercaptoéthanol.

VE #	tag en N-terminal	Nom du vecteur	limites du domaine de INI1	vecteur d'entrée d'origine	longueur du fragment+ tag (en nt)	longueur (en nt) entre UP1 et GWR	longueur du fragment+ tag (en aa)	Poids moléculaire théorique	pHi théorique	expression E.coli BL21(DE3)	Solubilité à 150mM NaCl 50mM TrisHCl pH8 2mMβMet
VE3539		p0GWA	74/257	E0962	546	827	182	21628,2	4,78	+	-
VE3540		p0GWK	74/257	E0962	546	827	182	21628,2	4,78	+	-
VE3541	GST-Tag	pGGWA	74/257	E0962	1551	2495	517	59220,5	5,33	+	-
VE3542	His-Tag	pHGWA	74/257	E0962	651	959	217	24656,6	5,59	+	-
VE3543	His-Tag	pHGWK	74/257	E0962	651	961	217	24656,6	5,59	+	-
VE3544	His-MBP	pHMGWA	74/257	E0962	546	1958	182	64668	5,25	+	+
VE3545		p0GWA	74/290	E0963	642	923	214	25355,4	4,74	+	-
VE3546		p0GWK	74/290	E0963	642	923	214	25355,4	4,74	+	-
VE3547	GST-Tag	pGGWA	74/290	E0963	642	1586	214	62714,1	5,25	+	-
VE3548	His-Tag	pHGWA	74/290	E0963	747	1055	249	28383,7	5,39	+	-
VE3549	His-Tag	pHGWK	74/290	E0963	747	1057	249	28383,7	5,39	+	-
VE3550	His-MBP	pHMGWA	74/290	E0963	642	2054	214	68395,1	5,18	+	+
VE3551		p0GWA	74/302	E0964	681	962	227	26772,1	4,78	+	-
VE3552		p0GWK	74/302	E0964	681	962	227	26772,1	4,78	+	-
VE3553	GST-Tag	pGGWA	74/302	E0964	681	1625	227	64130,8	5,26	+	-
VE3554	His-Tag	pHGWA	74/302	E0964	786	1094	262	29800,4	5,41	+	-
VE3555	His-Tag	pHGWK	74/302	E0964	786	1096	262	29800,4	5,41	+	-
VE3556	His-MBP	pHMGWA	74/302	E0964	681	2093	227	69811,8	5,19	+	+
VE3557		p0GWA	74/345	E0965	810	1091	270	31536,4	4,86	+	-
VE3558		p0GWK	74/345	E0965	810	1091	270	31536,4	4,86	+	-
VE3559	GST-Tag	pGGWA	74/345	E0965	810	1754	270	68895,1	5,29	+	-
VE3560	His-Tag	pHGWA	74/345	E0965	915	1223	305	34564,8	5,43	+	-
VE3561	His-Tag	pHGWK	74/345	E0965	915	1225	305	34564,8	5,43	+	-
VE3562	His-MBP	pHMGWA	74/345	E0965	810	2222	270	74576,2	5,21	+	+
VE3563		p0GWA	165/257	E0966	273	554	91	11370,7	3,97	-	-
VE3564		p0GWK	165/257	E0966	273	554	91	11370,7	3,97	-	-
VE3565	GST-Tag	pGGWA	165/257	E0966	273	1217	91	38729,5	4,93	+	+
VE3566	His-Tag	pHGWA	165/257	E0966	273	581	91	14130,8	4,69	-	-
VE3567	His-Tag	pHGWK	165/257	E0966	273	583	91	14130,8	4,69	-	-
VE3568	His-MBP	pHMGWA	165/257	E0966	273	1685	91	54410,5	4,94	+	+
VE3569		p0GWA	165/290	E0967	369	650	123	15097,9	4,11	+	-
VE3570		p0GWK	165/290	E0967	369	650	123	15097,9	4,11	+	-
VE3571	GST-Tag	pGGWA	165/290	E0967	369	1313	123	42456,6	4,88	+	-
VE3572	His-Tag	pHGWA	165/290	E0967	369	677	123	17857,9	4,66	+	-
VE3573	His-Tag	pHGWK	165/290	E0967	369	679	123	17857,9	4,66	+	-
VE3574	His-MBP	pHMGWA	165/290	E0967	369	1781	123	58137,7	4,9	+	+
VE3575		p0GWA	165/302	E0968	408	689	136	16514,6	4,19	-	-
VE3576		p0GWK	165/302	E0968	408	689	136	16514,6	4,19	+	-
VE3577	GST-Tag	pGGWA	165/302	E0968	408	1352	136	43873,3	4,91	+	+
VE3578	His-Tag	pHGWA	165/302	E0968	408	716	136	19274,6	4,72	-	-
VE3579	His-Tag	pHGWK	165/302	E0968	408	718	136	19274,6	4,72	-	-
VE3580	His-MBP	pHMGWA	165/302	E0968	408	1820	136	59554,4	4,92	+	+
VE3581		p0GWA	165/345	E0969	537	818	179	21278,9	4,35	+	-
VE3582		p0GWK	165/345	E0969	537	818	179	21278,9	4,35	+	-
VE3583	GST-Tag	pGGWA	165/345	E0969	537	1481	179	48637,7	4,95	+	-
VE3584	His-Tag	pHGWA	165/345	E0969	537	845	179	24039	4,82	+	-
VE3585	His-Tag	pHGWK	165/345	E0969	537	847	179	24039	4,82	+	-
VE3586	His-MBP	pHMGWA	165/345	E0969	537	1949	179	64318,7	4,95	+	+
VE3587		p0GWA	173/257	E0970	249	530	83	10427,7	4,01	+	+
VE3588		p0GWK	173/257	E0970	249	530	83	10427,7	4,01	+	+
VE3589	GST-Tag	pGGWA	173/257	E0970	249	1193	83	37786,5	5,03	+	+
VE3590	His-Tag	pHGWA	173/257	E0970	249	557	83	13187,8	4,88	+	+
VE3591	His-Tag	pHGWK	173/257	E0970	249	559	83	13187,8	4,88	+	+
VE3592	His-MBP	pHMGWA	173/257	E0970	249	1661	83	53467,5	5,01	+	+
VE3593		p0GWA	173/290	E0971	345	626	115	14154,9	4,15	+	+
VE3594		p0GWK	173/290	E0971	345	626	115	14154,9	4,15	+	+
VE3595	GST-Tag	pGGWA	173/290	E0971	345	1289	115	41513,6	4,97	+	-
VE3596	His-Tag	pHGWA	173/290	E0971	345	653	115	16914,9	4,79	+	+
VE3597	His-Tag	pHGWK	173/290	E0971	345	655	115	16914,9	4,79	+	+
VE3598	His-MBP	pHMGWA	173/290	E0971	345	1757	115	57194,7	4,96	+	+
VE3599		p0GWA	173/302	E0972	384	665	128	15571,6	4,25	+	-
VE3600		p0GWK	173/302	E0972	384	665	128	15571,6	4,25	+	-
VE3601	GST-Tag	pGGWA	173/302	E0972	384	1328	128	42930,3	4,99	+	-
VE3602	His-Tag	pHGWA	173/302	E0972	384	692	128	18331,6	4,85	+	-
VE3603	His-Tag	pHGWK	173/302	E0972	384	694	128	18331,6	4,85	+	-
VE3604	His-MBP	pHMGWA	173/302	E0972	384	1796	128	58611,4	4,98	+	+
VE3605		p0GWA	173/345	E0973	513	794	171	20335,9	4,41	+	-
VE3606		p0GWK	173/345	E0973	513	794	171	20335,9	4,41	+	-
VE3607	GST-Tag	pGGWA	173/345	E0973	513	1457	171	47694,7	5,03	+	-
VE3608	His-Tag	pHGWA	173/345	E0973	513	821	171	23096	4,94	+	-
VE3609	His-Tag	pHGWK	173/345	E0973	513	823	171	23096	4,94	+	-
VE3610	His-MBP	pHMGWA	173/345	E0973	513	1925	171	63375,7	5,02	+	+
VE4534	His-Tag	pHGWA	141/302	E1307	483	791	161	22171	5,17	+	-
VE4535	His-Tag	pHGWA	141/304	E1308	489	797	163	22341,2	5,17	+	-
VE4536	His-Tag	pHGWA	183/243	E1309	180	488	60	10618,1	5,54	+	-
VE4537	His-Tag	pHGWA	188/243	E1310	165	473	55	10080,4	5,77	+	-

Figure 17 Tableau récapitulatif de la banque de plasmides montrant les paramètres des vecteurs d'expression des fragments d'INI1. Résultats des tests d'expression dans E coli BL21 DE3 et des tests de solubilité. Le numéro de VE correspond au numéro attribué par la plateforme.

## Production et purification

### Matériels et méthodes

#### La production dans le système bactérien

La souche utilisée pour produire les protéines recombinantes est la même que pour les tests de mini-expression (*E. coli* BL21 (DE3) (Novagen) : F<sup>-</sup> *ompT hsdS<sub>B</sub>(r<sub>B</sub><sup>-</sup> m<sub>B</sub><sup>-</sup>) gal dcm (λ*Its857*, *ind1*, *Sam7 nin5 lacUV5-T7 gene1*)). Le génome bactérien contient le lysogène codant pour l'ARN polymérase du phage T7, sous le contrôle de l'opéron *lacUV5* inductible à l'IPTG. Dans le plasmide d'expression, le gène d'intérêt est cloné sous le contrôle d'un opéron lactose et d'un promoteur de transcription du bactériophage T7 si bien que la transcription ne peut s'effectuer que par l'ARN polymérase de T7 et en présence de lactose ou d'analogue (IPTG). L'IPTG (IsoPropyl β-D-1-ThioGalactopyranoside) est un analogue du lactose non métabolisable par *E. coli* utilisé comme inducteur gratuit de l'opéron lactose. La concentration d'IPTG reste constante ainsi le taux d'expression des gènes sous dépendance de l'opéron lactose ne varie pas. En absence d'inducteur, le répresseur produit par le gène *lacI* se fixe spécifiquement le site opérateur Lac ce qui empêche l'ARN polymérase de se fixer au site promoteur Lac, réprimant la transcription/traduction de l'ensemble des gènes. L'IPTG induit un changement conformationnel du répresseur LacI qui devient incapable de se fixer sur le site opérateur Lac. L'ARN polymérase T7 peut alors se fixer sur le promoteur T7 présent sur le vecteur et le gène cible est exprimé.*

#### Production à l'échelle du litre

Le changement d'échelle implique des ajustements. Plusieurs paramètres tels que la température d'induction, le temps de production, l'addition de 10 % p/v de saccharose permettent de compenser les variations dues au changement d'échelle et ainsi optimiser la production à l'échelle du litre.

#### Pré-culture

Après transformation de bactéries *E. coli* BL21 (DE3) par le plasmide d'intérêt, 2 mL de LB supplémenté de(s) antibiotique(s) approprié(s) par une colonie isolée est incubé pendant 6 heures à 37 °C sous agitation. Cette pré-culture est ensuite étalée sur plusieurs grandes boîtes de Pétri (LB agar et antibiotique(s)) puis ces boîtes sont placées à 37 °C durant la nuit. Cette

dernière étape a pour objectif d'obtenir un tapis bactérien pour l'ensemencement des cultures en grand volume et pour vérifier l'absence de plage de lyse.

### ***Culture***

Une culture d'un litre s'effectue en fiole de 5L pour une meilleure oxygénation des cellules, sous agitation (190 rpm) dans des incubateurs (INFORS™). Généralement, les cultures s'effectuent par 6 L ou 12 L réparties dans 6 ou 12 fioles. Extemporément à l'ensemencement, le milieu est supplémenté de(s) antibiotique(s) approprié(s), et/ou de 10 % p/v de saccharose (solution stock à 50 % p/v de saccharose). Les boîtes de Pétri sont raclées pour récupérer les bactéries. La solution bactérienne est ensuite homogénéisée et mesurée (DO600nm). Le volume d'ensemencement est calculé pour obtenir une DO600nm finale de 0,1 dans le milieu de culture. La phase de pousse de la culture s'effectue à 37°C sous agitation constante (190 rpm). La progression de la culture est suivie par mesure de la DO600nm. Quand la DO600nm atteint entre 0,4 et 0,5 la température est abaissée à 18 °C. Après une demi-heure à 18 °C, l'IPTG est ajouté pour une concentration finale de 0,5mM, la phase de production commence.

### ***Récolte***

Après une nuit à 18 °C, les bactéries sont récoltées par centrifugation (20 min à 4000 g à 4°C) et soigneusement lavées avec une solution de 10 g/L de NaCl stérile. Les culots bactériens secs sont ensuite pesés, congelés et conservés à -20°C. Les rendements masse/vol moyens sont de l'ordre de 4 à 5 g de cellules/L de culture.

### ***Purification***

#### ***Lyse des bactéries***

Selon les protéines étudiées, il est nécessaire d'adapter les conditions de lyse. Les culots bactériens sont repris dans le volume adéquat de tampon de lyse, et homogénéisé par mixage (UltraTurax® T-25). D'une manière générale il convient d'utiliser 10 mL de tampon de lyse par gramme de culot bactérien. La lyse s'effectue sous agitation à l'aide d'un sonicateur (Vibracell 72412 Fischer BioBlock™) équipé d'une sonde de 13 mm de diamètre, durant 1 min/(g de culot) de temps effectif par des impulsions d'ultra-sons toutes les 2 sec à 40% d'amplitude à 4°C. La sonication dégage de la chaleur, afin d'éviter la détérioration de la

protéine, l'échantillon est plongé dans un bain froid (mélange glace/eau/NaCl) et une sonde de température stoppe la sonication si l'échantillon atteint 10 °C. De plus, des inhibiteurs de protéases (DIFP et comprimés du cocktail d'inhibiteurs Complete, EDTA-free (Roche™)) sont ajoutés au début et au cours de la lyse. En fin de lyse l'aspect général du lysat doit être homogène et fluide.

### *Ultracentrifugation.*

Les débris bactériens sont culotés par ultracentrifugation pendant une heure et demie à 45 000 g à 4°C. Le surnageant contenant l'extrait soluble est récupéré, ré-implémenté en inhibiteurs de protéases et immédiatement utilisé.

### *Principes de chromatographies*

Les étapes de chromatographie sont réalisées sur un système FPLC (Fast Protein Liquid Chromatography) soit BioLogic Workstation® (BioRad™), soit ÄKTA Purifier® (GE Healthcare™).

#### *La chromatographie d'affinité*

La chromatographie d'affinité (Figure 18) est basée sur l'interaction réversible entre une protéine (ou un groupe de protéines) et un ligand spécifique couplé à une matrice de la résine. Cette technique permet de séparer sélectivement la protéine d'intérêt des protéines bactériennes. La résine de la chromatographie d'affinité est formée de billes d'agarose couplé au ligand qui doit être spécifique du tag d'affinité fusionné à la protéine d'intérêt. Grâce à ce tag, la protéine d'intérêt sera retenue puis élue spécifiquement par un ligand libre compétitif. Les deux types de tag utilisés pour les purifications sont l'hexahistidine (6H) et la Glutathion S-Transférase (GST).

#### *La Glutathion S-Transférase (GST)*

La GST (Glutathion S-Transférase de *Schistosoma japonicum*) est une protéine de fusion de 26 000 Da qui exprimée dans *E. coli* conserve l'activité enzymatique et la capacité à former des dimères. Les protéines de fusion à la GST sont facilement purifiées à partir de lysat bactérien par colonne d'affinité utilisant de la glutathion immobilisée sur une matrice de

sépharose. La glutathion, complémentaire au site de liaison de la Glutathion S-Transférase, est couplée aux billes d'agarose par l'intermédiaire d'un bras espaceur de dix atomes de carbone. La cinétique de liaison entre la glutathion sépharose et la GST est relativement lente. Lorsque la charge de l'échantillon s'effectue sur un système FPLC, le débit doit être de 0,2 mL/min pour maximiser la capacité de liaison. L'élution de la protéine d'intérêt s'effectue par injection du tampon lyse supplémenté par 50 mM glutathion réduit.

#### *La Chromatographie d'affinité par cations métalliques divalents immobilisés*

Ce type de chromatographie d'affinité permet de séparer les protéines portant un tag polyhistidine (ici l'hexahistidine). Des cations métalliques divalents ( $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$ ,  $\text{Zn}^{2+}$ ,  $\text{Cu}^{2+}$ ), immobilisés par des groupements imido-diacétique qui sont greffés *via* un espaceur de sept atomes sur une matrice de type sépharose. Ces cations interagissent avec les hexahistidines de la protéine recombinante. Dans les purifications effectuées les cations divalents utilisés sont  $\text{Ni}^{2+}$  d'une couleur cyprine caractéristique. L'étape de lavage permet de d'éliminer les protéines bactériennes endogènes. La protéine d'intérêt peut être éluee par différentes méthodes : modification du pH, agent chélateur ou molécule compétitrice (ici l'imidazole). L'élution par l'imidazole peut s'effectuer selon deux tactiques. La première consiste en un gradient croissant de la concentration d'imidazole. En général, cette méthode permet d'évaluer la concentration d'imidazole à laquelle la protéine recombinante se sépare des cations divalents immobilisée sur la matrice. La seconde par palier de concentration d'imidazole. Cette dernière est utilisée lorsque la concentration d'imidazole d'élution est connue. La concentration initiale d'imidazole permet d'éviter la liaison aspécifique de protéines bactériennes endogènes sur les cations. La concentration finale d'imidazole permet l'élution effective de la protéine d'intérêt.

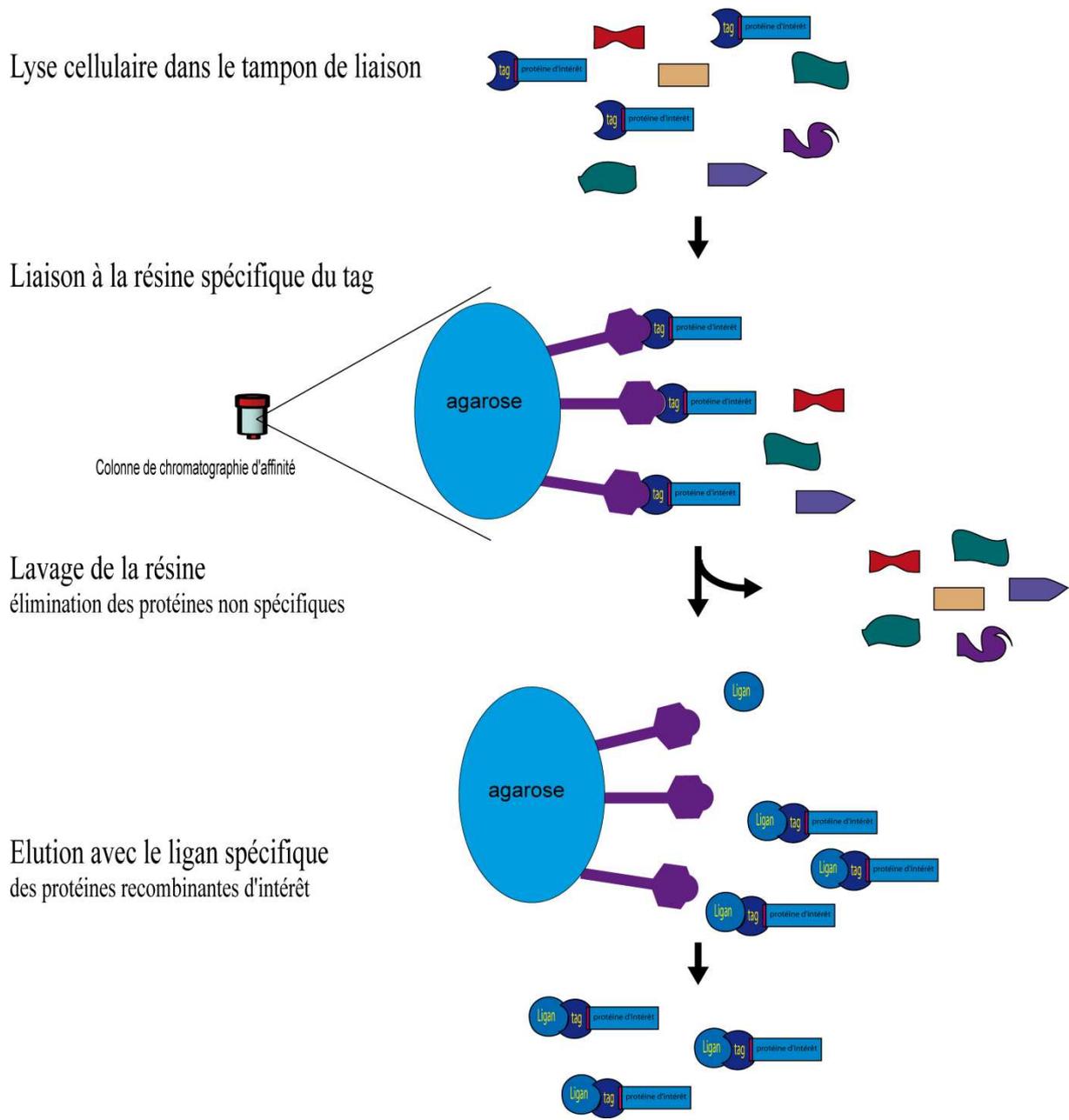


Figure 18 : Principe de la chromatographie d'affinité

### Chromatographie d'exclusion de taille (ou filtration sur gel)

Cette technique sépare les protéines en fonction de leur taille et de leur forme. La résine est constituée d'une matrice de dextrane (polymère de dextrose) et d'agarose qui forment des granules poreux. Les protéines sont éluées dans l'ordre décroissant de leur poids moléculaire dans la gamme de résolution de la colonne. Toutes les protéines dont le diamètre est supérieur à celui des pores sont exclues de la matrice et par conséquent éluées au niveau du volume

d'exclusion, appelé volume mort. La migration des protéines dont le diamètre est inférieur à celui des pores est fonction de leur taille et de leur forme. Les protéines seront ainsi éluées linéairement en fonction du logarithme de leur masse moléculaire. Enfin les protéines dont la taille est inférieure à la réticulation seront incluses dans la matrice et seront éluées en même temps au niveau du volume total.

### Dialyse

Cette technique est utilisée pour purifier les solutions, notamment séparer les petites molécules (sels détergents,...) des grosses molécules telles que les protéines (Figure 19). Une membrane poreuse semi-perméable permet aux petites molécules de diffuser vers une solution moins concentrée et retient les grosses molécules. La membrane est définie selon le seuil de poids moléculaire (MWCO) auquel 95 % des molécules ayant ce poids moléculaire sont retenues. Durant la purification, les dialyses ont deux finalités différentes. La première consiste à ôter les ligands nécessaires à l'éluion. La deuxième consiste en la formation des différents complexes. Le tampon par défaut est celui où la protéine la moins soluble est stable. Les autres partenaires sont également dans ce tampon. Après mélange des différents partenaires, la dialyse permet d'éliminer le détergent et en diminuer la force ionique (concentration en sel). Les interactions hydrophobes et électrostatiques entre les protéines sont ainsi favorisées. Cette dialyse graduelle permet un repliement induit des protéines par la formation de complexes avec les partenaires et sont ainsi maintenues solubles.

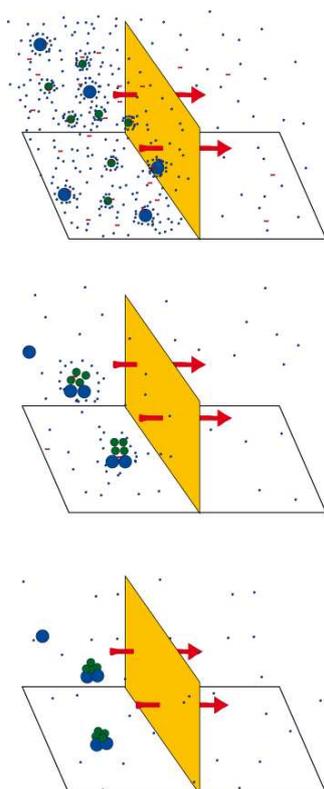


Figure 19 Principe de formation in vitro de complexe par dialyse. Les deux protéines (ronds bleus (LEDGF) et ronds verts (INtegrase)) sont solubles indépendamment dans un tampon à forte concentration en sel (points bleus) et en présence de détergent (rouge). La membrane de dialyse (en jaune) laisse passer les ions et les détergents (flèches rouges). Au fur et à mesure des dialyses la concentration en sel et détergent diminue favorisant ainsi la formation du complexe. Au final, le partenaire le plus soluble stabilise le second, lui permettant de rester en solution dans un tampon où normalement la protéine précipite

## Les différentes stratégies de formation de complexe

La formation d'un complexe *in vitro* est basée sur la stabilisation d'un ou plusieurs partenaires par d'autres protéines. Cette méthode évite de solubiliser artificiellement la protéine en formant des agrégats solubles à la manière des protéines de fusion tels la GST ou MBP (Nominé, Ristriani et al. 2001). De plus l'interaction est structurellement et/ou fonctionnellement significative. Plusieurs stratégies de formation de complexe ont été envisagées : par dialyse, par co-broyage, par co-expression. L'impossibilité de tester la totalité des combinaisons m'a poussé à chercher la plus forte probabilité d'obtenir des résultats rapidement en se focalisant sur les têtes de séries des différents tests effectués.

### *La formation de complexe par co-expression*

Dans le même système d'expression (ici la bactérie *E.coli*), à partir d'un vecteur exprimant les différents partenaires du complexe, ou de plusieurs plasmides exprimant séparément les protéines du complexe, les différentes protéines sont produites dans l'expectative que le complexe se forme au sein de la bactérie. De manière générale, la position des séquences des différentes protéines exprimées dans un plasmide de co-expression a une importance dans la formation du complexe. Pour la co-transformation, la quantité de chaque plasmide inséré dans la bactérie est une variable à contrôler. Pour la production de certaines protéines humaine, notamment LEDGF, un plasmide (pRARE) supplémentaire codant pour des ARN de transferts portant les codons rares est nécessaire.

Les différentes tentatives de co-expression dans la souche bactérienne *E.coli BL21(DE3)* du complexe IN-LEDGF à partir de plasmide bicaténaire se sont avérées peu prometteuses. *A fortiori*, la co-expression de trois partenaires n'était pas un choix tactique judicieux.

### *La formation de complexe par co-broyage*

La technique du co-broyage consiste à produire séparément les différents partenaires du complexe. Dans un tampon favorisant la formation du complexe, les cellules sont mélangées et lysées dans des proportions adéquates en fonction de la stœchiométrie du complexe et du taux d'expression lors de la production. C'est le principal frein à cette méthode. La variabilité des différentes productions des différentes protéines rend difficile la reproductibilité. En effet, cette variabilité au sein des productions nécessite une phase de mise au point importante pour chaque nouvelle production et ce pour chaque protéine.

### *La formation de complexe par dialyse*

La technique de la formation de complexe par dialyse s'est avérée être la plus robuste en terme de reproductibilité. La production, la lyse et la purification s'opèrent séparément, de sorte à obtenir en un premier temps les différents partenaires séparément. Cette étape de purification gomme la variabilité de la production et permet un meilleur contrôle du ratio stoechiométrique des différents partenaires du complexe. Lors du mélange des différents partenaires, le tampon dans lequel se trouvent l'ensemble des protéines est identique au tampon où la protéine la moins soluble est stable seule. Ce tampon contient une concentration en chlorure de sodium élevée pour favoriser les interactions électrostatiques et un détergent zwitterionique qui stabilise les régions hydrophobes des protéines évitant le phénomène de « salting out » (Figure 37). En dialysant le détergent et en diminuant la concentration en chlorure de sodium, les interactions hydrophobes et électrostatiques entre les protéines sont favorisées. Cette dialyse graduelle permet la formation *in vitro* de complexes stable.

### *Mode opératoire de la formation du complexe Intégrase/LEDGF : dialyse*

Les constructions utilisées pour la formation du complexe sont : Intégrase étiquetée GST avec un site de coupure à la Protéase 3C (P3C) = (GS-IN) et LEDGF étiquetée par l'hexahistidine avec un site de coupure à la thrombine = (HT-LEDGF). Séparément, pour les deux protéines, la culture et la lyse s'effectue selon le protocole standard décrit ci-dessus dans le tampon de lyse (NaCl 1000 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM, CHAPS 7 mM, glycérol 5 %).

Capture : Chromatographie d'affinité

Après avoir clarifié séparément par ultracentrifugation, les lysats sont chargés indépendamment sur des colonnes d'affinité.

*GS-IN :*

Une colonne GSTrap FF 5 mL (GE Healthcare™) est préalablement équilibrée avec 10 VC (volume colonne) de tampon de lyse. La bonne équilibration est vérifiée par une conductivité stable. Le lysat clarifié est chargé sur la colonne à 0,2 mL/min. Puis le lavage (15 à 20 VC) de la colonne s'effectue à 5 mL/min jusqu'à stabilisation de la DO<sub>280nm</sub> à la valeur initiale. La

GS-IN est ensuite éluée par injection sur la colonne d'un tampon de lyse supplémenté par 50 mM de glutathion réduit pH7.0. L'éluion est considérée terminée lors de la stabilisation de la DO280nm (environ 10 VC). La quantification et la qualification de la protéine s'effectue respectivement par spectrophotométrie (NanoDrop®) et par SDS-PAGE 12,5 %.

#### *HT-LEDGF*

La colonne utilisée pour la purification de HT-LEDGF est HiTrap Chelating HP 5 mL (GE Healthcare™). L'équilibration s'opère à l'identique, la charge de l'échantillon s'effectue à 5 mL/min. Le lavage est réalisé avec 3 % de tampon d'éluion (tampon de lyse + 250 mM Imidazole pH7.0). Ce lavage permet l'élimination des protéines bactériennes endogènes fixées non spécifiquement sur la résine de la colonne. L'éluion s'effectue sur un gradient de 3 % à 65 % de tampon d'éluion sur 40 VC. HT-LEDGF est éluée entre 12 % et 40 % de tampon d'éluion. Un gradient de 5 VC de 65% à 100 % permet une vérification du recouvrement de l'éluion. La quantification et la qualification de la protéine s'effectue respectivement par spectrophotométrie (NanoDrop®) et par SDS-PAGE 12,5 %.

#### Formation du complexe : dialyse

GS-IN et HT-LEDGF sont mélangées selon un ratio molaire de un pour deux. Cet excès de HT-LEDGF a pour but d'empêcher que de l'intégrase seule précipite lors de la dialyse. La dialyse s'effectue avec une membrane MWCO 10 kDa (SpetraPor Dialysis® Spectrum Lab™) à 4 °C en trois temps. Le volume de chaque dialyse est cent fois supérieur au volume à dialyser. Typiquement, GS-IN et HT-LEDGF sont dans 20 mL pour 2000 mL de tampon de dialyse. La première dialyse s'effectue pendant 2 à 3 heures contre le tampon NaCl 600 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM, CHAPS 1,3 mM. La deuxième et la troisième dialyse durent 2 à 3 heures contre le tampon NaCl 150 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM. La quantification et la qualification de la protéine s'effectue respectivement par spectrophotométrie (NanoDrop®) et par SDS-PAGE 12,5 %.

#### Purification du complexe : chromatographie d'affinité par gravité

Afin d'élimination de l'excédent de LEDGF, et de ne garder que le complexe GS-IN/HT-LEDGF, l'échantillon est incubé pendant un nuit à 4 °C sous agitation douce dans un tube

avec de la résine glutathion sepharose 4B (GE Healthcare™) préalablement équilibré avec du tampon de dialyse NaCl 150 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM. Le volume de résine est fonction de la quantité de protéines, soit environ 1 mL de résine pour 5 mg de protéines. La résine est transvasée dans une colonne de paillasse, puis est lavé avec 2 fois 15 VC, l'excédent de LEDGF est éliminé dans le lavage. Le complexe est élué en 5 à 6 fois 1 VC avec du tampon de dialyse supplémenté par 50 mM de glutathion réduit pH7.0. La quantification et la qualification de la protéine s'effectue respectivement par spectrophotométrie (NanoDrop®) et par SDS-PAGE 12,5 %.

#### Élimination du glutathion réduit : dialyse

L'échantillon est dialysé deux fois à 4 °C pendant 2 à 3 heures contre cent fois le volume de l'échantillon dans le tampon NaCl 150 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM afin d'éliminer le glutathion réduit.

#### Coupure et élimination du tag : protéolyse spécifique

La protéase 3C (P3C) est ajoutée à hauteur de 20 µg (GST-P3C produite au laboratoire) par mg de complexe. La P3C coupe spécifiquement la séquence insérée entre le tag GST et l'intégrase. Pendant deux jours, l'échantillon est incubé à 4°C sous agitation douce dans un tube avec de la résine glutathion sepharose 4B (GE Healthcare™) (environ 250 µL de billes pour 1 mg de complexe) préalablement équilibré avec du tampon de dialyse NaCl 150 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM. La résine est transvasée dans une colonne de paillasse, puis laver deux fois avec 2 VC. Le complexe se retrouve dans le volume de lavage, tandis que la GST et la GST-P3C sont retenues spécifiquement sur la résine. La quantification et la qualification du complexe s'effectue respectivement par spectrophotométrie (NanoDrop®) et par SDS-PAGE 12,5 %.

#### Polissage : filtration sur gel.

L'échantillon est concentré sur Amicon Ultra 15 50kDa (Millipore™) pour obtenir un volume de 5 mL. La filtration sur gel s'effectue par FPLC grâce à une colonne de type Superdex 200 High Load 16/60 PrepGrad dans le tampon NaCl 150 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM. La quantification et la qualification du complexe s'effectue respectivement par spectrophotométrie (NanoDrop®) et par SDS-PAGE 12,5 %.

### *Mode opératoire de la formation du complexe Intégrase/LEDGF/IBD(INI1)*

Les constructions utilisées pour la formation du complexe ternaire sont : Intégrase étiquetée GST avec un site de coupure à la Protéase 3C (P3C) = (GS-IN) et LEDGF étiquetée par l'hexahistidine avec un site de coupure à la thrombine = (HT-LEDGF) et le domaine de liaison à l'intégrase (IBD) d'INI1 étiquetée par l'hexahistidine avec un site de coupure à la TEV = (HV-IBD(INI1)). Le complexe Intégrase LEDGF est préalablement formé comme décrit précédemment. Le domaine d'INI1 est purifié séparément. La culture et la lyse s'effectue selon le protocole décrit ci-dessus dans le tampon de lyse (NaCl 2000 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM, CHAPS 20 mM).

#### Capture : Chromatographie d'affinité

Après avoir clarifié par ultracentrifugation, le lysat est chargé sur une colonne d'affinité HiTrap Chelating HP 5 mL (GE Healthcare™). Après équilibration, la charge de l'échantillon s'effectue à 3 mL/min. Le lavage par le tampon de lyse se déroule sur 20 VC, puis l'élimination des protéines fixées non spécifiquement s'effectue à 20 % de tampon d'élution (tampon de lyse + 500 mM imidazole) s'effectue également sur 20 VC. Enfin un gradient de 20 % à 100 % de tampon d'élution permet d'éluer HV-IBD(INI1) entre 40 % et 60 %. La quantification et la qualification de la protéine s'effectue respectivement par spectrophotométrie (NanoDrop®) et par SDS-PAGE 15 %.

#### Polissage : filtration sur gel.

La filtration sur gel s'effectue par FPLC grâce à une colonne de type Superdex 75 High Load 16/60 PrepGrad dans le tampon NaCl 2000 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM CHAPS 20 mM. La quantification et la qualification de la protéine s'effectue respectivement par spectrophotométrie (NanoDrop®) et par SDS-PAGE 15 %.

#### Formation du complexe ternaire: dialyse en flux continu

Le tampon (150 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM) du complexe Intégrase/LEDGF est changée par dilution au demi par un tampon (NaCl 4000 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM CHAPS 40 mM) pour obtenir un tampon final NaCl 2000 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2

mM 20 mM CHAPS. L'ajout de HV-IBD(INI1) s'effectue dans un ratio 4 HV-IBD(INI1) pour 1 complexe soit une molécule d'HV-IBD(INI1) pour une molécule d'Intégrase. La dialyse en flux continu (Figure 20) dans un volume cent fois supérieur au volume à dialyser, avec une membrane MWCO 3 à 6 kDa (SpectraPor Dialysis® Spectrum Lab™) à 4 °C pendant deux jours. Le boudin de dialyse est placé dans le tampon de départ NaCl 2000 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM CHAPS 20 mM. Avec l'aide d'une pompe péristaltique, le tampon de départ est aspiré dans l'éprouvette où se situe le boudin de dialyse, et dans le même temps le tampon d'arrivée (NaCl 1000 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM CHAPS 7 mM) y est injecté. Un barreau magnétique assure l'homogénéisation de la solution. Une fois que cent volumes à dialyser de tampon (NaCl 1000 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM CHAPS 7 mM) a été injecté, le tampon d'arrivée est changé pour NaCl 500 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM. Quand cent volumes du second tampon d'arrivée a été injecté, le boudin de dialyse est placé pour une ultime dialyse dans le tampon NaCl 500 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM.

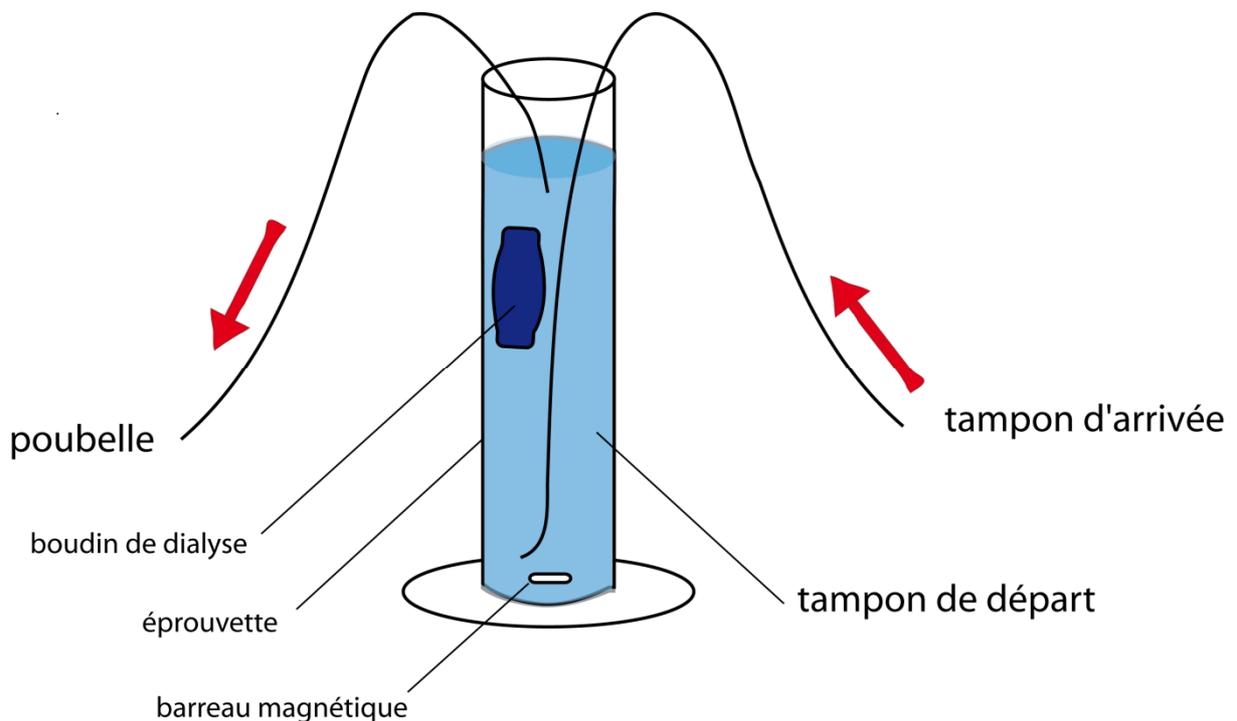


Figure 20: dialyse en flux continu

Polissage : filtration sur gel.

L'échantillon est concentré sur Amicon Ultra 15 10kDa (Millipore™) pour obtenir un volume de 5 mL. La filtration sur gel s'effectue par FPLC grâce à une colonne de type Superdex 200 High Load 16/60 PrepGrad dans le tampon NaCl 500 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM. La quantification et la qualification du complexe s'effectue respectivement par spectrophotométrie (NanoDrop®) et par SDS-PAGE 15 %.

### **Fin purification**

Les complexes Intégrase/LEDGF et Intégrase/LEDGF/IBD(INI1) sont maintenant purifiés et prêts pour une étude structurale et fonctionnelle.

## **Résultats**

### **Introduction**

Deux types de complexes ont été envisagés. 1) Le complexe entre le domaine de liaison à l'intégrase d'INI1 et le domaine du cœur catalytique de l'intégrase VIH-1 mutant F185K. La structure du cœur catalytique (50-212) de l'intégrase F185K ayant déjà été résolue par cristallographie aux rayons X (Dyda, Hickman et al. 1994), ce complexe pourrait être la base d'une étude cristallographique. 2) Un complexe ternaire comprenant l'intégrase pleine longueur type sauvage, LEDGF pleine longueur type sauvage et le domaine de liaison à l'intégrase d'INI1. L'étude structurale par microscopie électronique du complexe Intégrase LEDGF (Michel, Crucifix et al. 2009) servira de base à l'étude d'autres partenaires cellulaires (INI1) et de leurs interactions entre-eux et avec l'intégrase.

### ***Essais de formation du complexe INI1 (173-257) / intégrase (50-F185K-212)***

Dans un premier temps, la production, la lyse et la purification s'opèrent indépendamment de sorte à obtenir les différents partenaires séparément. Puis dans un deuxième temps, des essais de formation de complexe IN(50-F185K-212) / INI1(173-257) sont réalisés par dialyse. La vérification de la formation du complexe est effectuée par chromatographie d'exclusion de

taille. En effet, une co-élution des deux partenaires à un poids moléculaire apparent cohérent stoechiométriquement constitue une première preuve de la formation d'un complexe.

#### *Purification du domaine de liaison à l'Intégrase d'INI1 (173-257)*

Le fragment INI1(173-257) a été fusionné au tag 6His-MBP pour sa surexpression et sa solubilité significativement supérieures aux autres fragments. La production s'effectue à partir de 3 L de culture (*E.coli* BL21 DE3, milieu LB, sélection ampicilline) soit 13,2 g de cellules. La phase de croissance bactérienne s'effectue à 37°C sous agitation constante (190 rpm), puis la phase de production s'effectue à 18 °C sur une période de 12 heures.

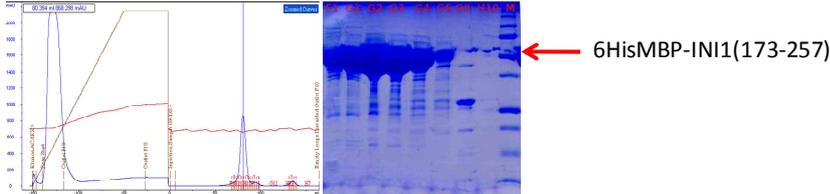
La purification du domaine de liaison à l'intégrase INI1(173-257) est réalisée en cinq étapes (Figure 21). Les culots bactériens sont repris dans 130 mL de tampon de lyse tampon L1 (NaCl 0,5 M HEPES 50 mM pH7.5 MgCl<sub>2</sub> 5 mM β-mercaptoéthanol 2 mM ) puis lysés par sonification. Après centrifugation, l'extrait soluble est injecté sur une colonne de chromatographie d'affinité (HiTrap Chelating HP 5 mL (GE Healthcare™)) dont les cations divalents sont Ni<sup>2+</sup>. Un gradient d'élution de 0 à 1 M d'imidazole permet d'éluer HisMBP-INI1(173-257) vers 400 mM d'imidazole. Une seconde étape de purification par filtration sur gel (SUPERDEX 200 HighLoad 16/60 PrepGrad) dans le tampon L1 permet d'obtenir 60 mg de protéine d'intérêt et d'éliminer l'imidazole. La troisième étape est une protéolyse spécifique du tag HisMBP. Le site de clivage TEV se situe entre le tag HisMBP et INI1(173-257). La protéolyse s'effectue à 4 °C pendant 48 heures. Afin de séparer le HisMBP du fragment d'intérêt INI1(173-257), l'échantillon est injecté sur une colonne de chromatographie d'affinité (HiTrap Chelating HP 5 mL (GE Healthcare™)) dont les cations divalents sont Ni<sup>2+</sup>. Avec ces deux étapes, le rendement chute à 10 %, soit 6 mg d'INI1(173-257). Au final, 1,5 mg d'INI1(173-257) seront obtenus après la dernière étape de filtration sur gel (SUPERDEX 200 HighLoad 16/60 PrepGrad) dans le tampon L1.

**Purification**  
6HisMBP-INI1(173-257)

**Production:** LB 18°C  
**Lyse:** sonification  
Tampon L1

**Chromatographie d'affinité**  
HiTrap Chelating HP 5mL Ni<sup>2+</sup>  
Tampon L1 / Tampon E1

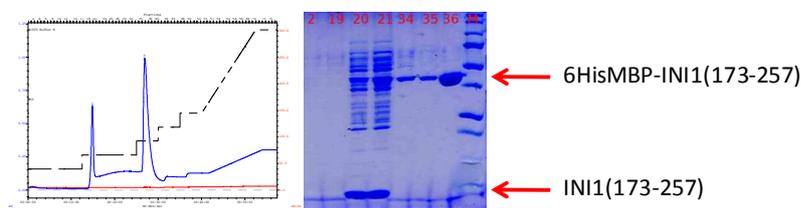
**Filtration sur gel**  
SUPERDEX 200 HighLoad 16/60 PrepGrad  
Tampon L1



**Protéolyse spécifique ménagée**  
His-TEV  
Tampon L1



**Chromatographie d'affinité**  
HiTrap Chelating HP 5mL Ni<sup>2+</sup>  
Tampon L1 / Tampon E1



**Filtration sur gel**  
SUPERDEX 200 HighLoad 16/60 PrepGrad  
Tampon L1

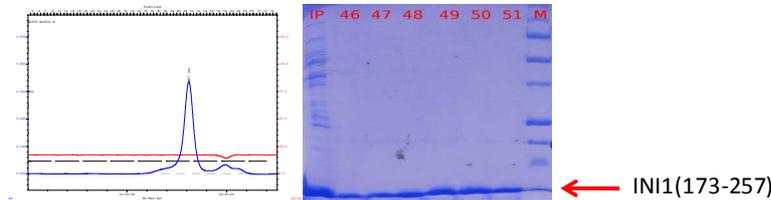


Figure 21 : Récapitulatif de la production et de la purification du 6His-MBP – INI1(173-257). Pour 13,2g de cellules (3L), 60 mg de protéine d'intérêt sont obtenus après la deuxième étape soit 4,5 mg/g de cellules, et 3 mg après séparation du tag. L'ultime étape de filtration sur gel permet d'obtenir 1,5 mg soit environ 0,1 mg/g de cellules. La composition des tampons est: pour le tampon L1 : NaCl 0,5 M HEPES 50 mM pH7.5 MgCl<sub>2</sub> 5 mM β-mercaptoéthanol 2 mM ; pour le tampon E1 : NaCl 0,5 M HEPES 50 mM pH7.5 MgCl<sub>2</sub> 5 mM β-mercaptoéthanol 2 mM imidazole 1000 mM.

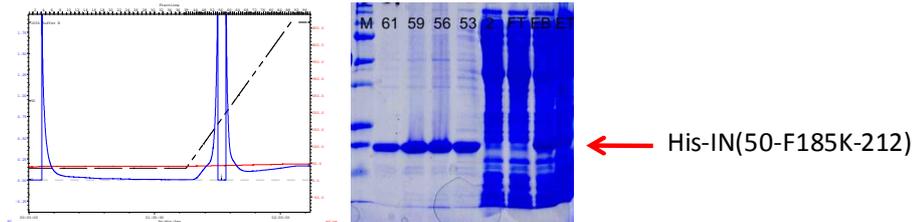
### *Purification du cœur catalytique (50-212) de l'intégrase F185K*

La purification du cœur catalytique (50-212) de l'intégrase F185K (IN(50-F185K-212)) est documentée (Goldgur, Dyda et al. 1998). Après la production en fiole, 7,9 g de cellules sont récoltées pour 3 L de culture (*E.coli* BL21 DE3, milieu LB, sélection ampicilline). Les conditions de croissance et de production sont identiques à celles d'INI1 (173-257). La purification d'IN(50-F185K-212) s'effectue en deux étapes (Figure 22). Les culots bactériens sont repris dans 80 mL tampon de lyse (tampon L2 : NaCl 0,5M HEPES 20mM pH7,5  $\beta$ MEt 2mM imidazole 5mM) puis lysés par sonification. Après centrifugation, l'extrait soluble est injecté sur une colonne de chromatographie d'affinité (HiTrap Chelating HP 5 mL (GE Healthcare™)). Un gradient d'élution de 5 mM à 800 mM d'imidazole permet d'éluer 22,5 mg de protéine d'intérêt vers 300 mM d'imidazole. Une seconde étape de purification par filtration sur gel (SUPERDEX 200 HighLoad 16/60 PrepGrad) dans le tampon G2 (NaCl 0,5M TrisHCl 20mM pH7,5 EDTA 1mM DTT 5mM) permet d'évaluer un poids moléculaire apparent de 40 KDa correspondant à un dimère de l'IN(50-F185K-212). En fin de purification la quantité de protéine d'intérêt est évaluée à 7,6 mg.

## Purification His-IN(50-F185K-212)

**Production:** LB 18°C  
**Lyse:** sonification  
Tampon L2

**Chromatographie d'affinité**  
HiTrap Chelating HP 5mL Ni<sup>2+</sup>  
Tampon L2 / Tampon E2



**Filtration sur gel**  
SUPERDEX 200 HighLoad 16/60 PrepGrad  
Tampon G2

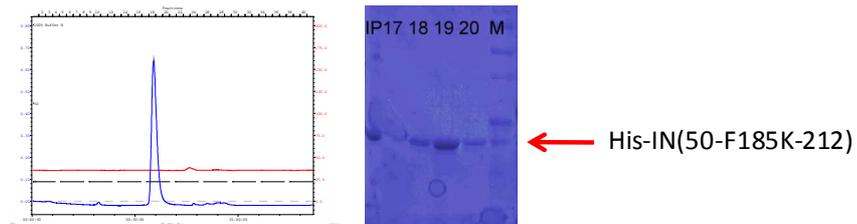


Figure 22: Récapitulatif de la production et de la purification du 6His-tag - cœur catalytique (50-212) de l'intégrase F185K. Pour 7,9 g de cellules (3L), 22,5 mg de protéine d'intérêt pour la première étape (affinité) soit 2,8 mg/g de cellules, et 7,6 mg pour la seconde étape (filtration sur gel) soit environ 1,0 mg/g de cellules. La composition des tampons sont: pour le tampon L2 : NaCl 0,5 M HEPES 50 mM pH7.5 β-mercaptoéthanol 2 mM imidazole 5 mM ; pour le tampon E2 : NaCl 0,5 M HEPES 50 mM pH7.5 β-mercaptoéthanol 2 mM imidazole 800 mM ; pour le tampon G2 : NaCl 0,5 M TrisHCl 20 mM pH7.5 EDTA 1 mM DTT 5 mM.

## Essais de formation du complexe IN(50-F185K-212) / INI1(173-257)

Plusieurs essais de formation du complexe entre IN(50-F185K-212) et INI1(173-257) ont été réalisés. Cependant, aucun essai n'a pu aboutir à la formation du complexe entre ces deux protéines (Figure 23). Différents ratios ont été testés de 2 IN(50-F185K-212) pour 1 INI1(173-257) à 1 IN(50-F185K-212) pour 4 INI1(173-257) à différentes force ionique (NaCl 250 mM ; 150 mM et 50 mM) dans la même base de tampon HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM.

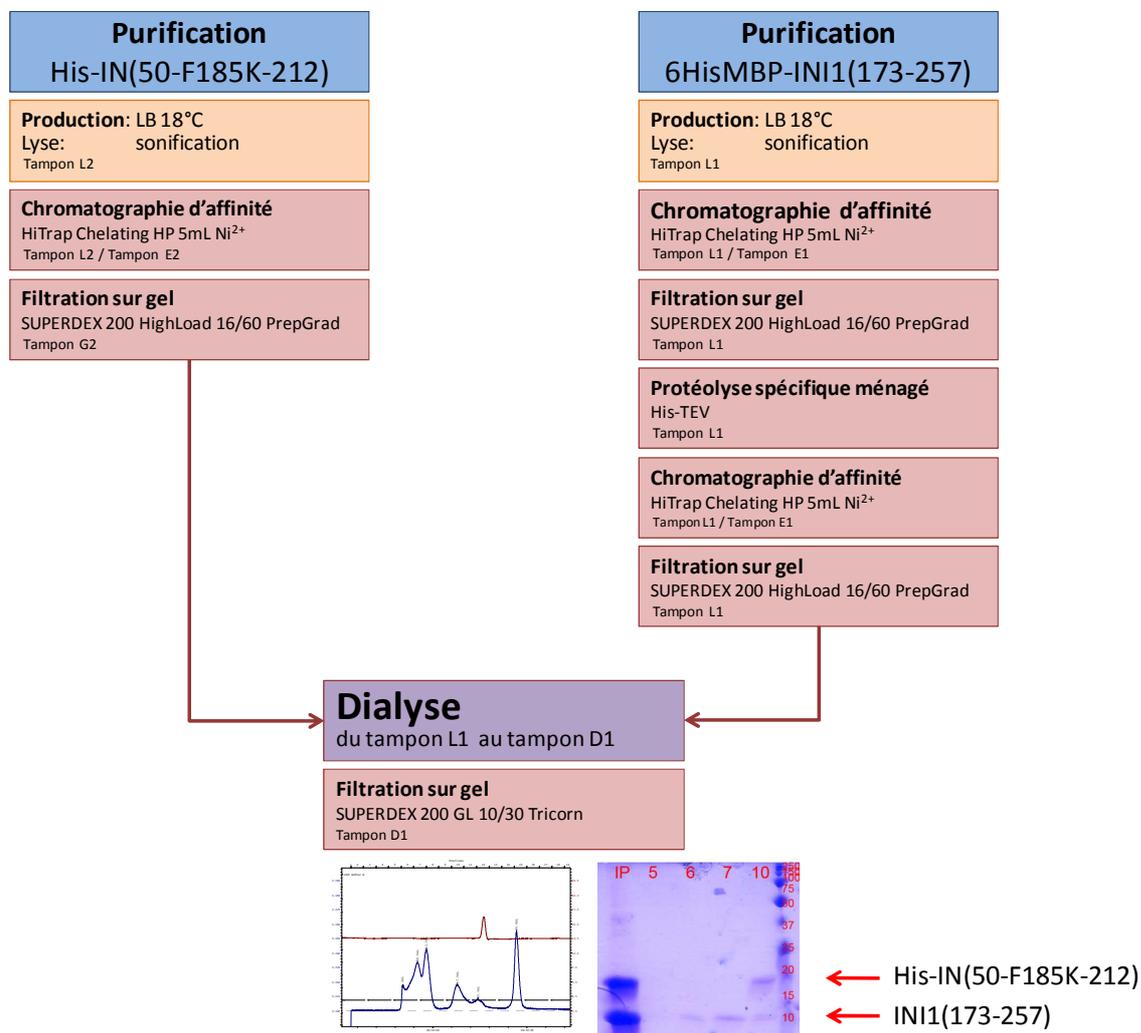


Figure 23: Récapitulatif de l'essai de formation du complexe entre 6His-tag - cœur catalytique (50-212) de l'intégrase F185K et INI1(173-257) (ratio 1:2) à NaCl 250 mM HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM (tampon D1). La composition du tampon L1 est NaCl 0,5 M HEPES 50 mM pH7.5 MgCl<sub>2</sub> 5 mM β-mercaptoéthanol 2 mM, du tampon E1 est NaCl 0,5 M HEPES 50 mM pH7.5 MgCl<sub>2</sub> 5 mM β-mercaptoéthanol 2 mM imidazole 1000 mM ; du tampon L2 est NaCl 0,5 M HEPES 50 mM pH7.5 β-mercaptoéthanol 2 mM imidazole 5 mM ; du tampon E2 est NaCl 0,5 M HEPES 50 mM pH7.5 β-mercaptoéthanol 2 mM imidazole 800 mM ; du tampon G2 est NaCl 0,5 M TrisHCl 20 mM pH7.5 EDTA 1 mM DTT 5 mM ; du tampon D1 est NaCl 0,25 M HEPES 50 mM pH7.0 MgCl<sub>2</sub> 5 mM β-mercaptoéthanol 2 mM.

Une autre alternative pour la formation de complexe a été testée : la coupure de l'étiquette 6HisMBP pendant la dialyse. La coupure du tag en même temps que la dialyse de formation du complexe a également été testée à différents ratio IN(50-F185K-212/INI1(173-257) (2:1 ; 1:1 ; 1:2 ; 1:4) pour différentes forces ioniques (NaCl 250mM ; 150mM ; 50mM). Dans ce cas, le mélange s'effectue après la première filtration sur gel de 6HisMBP-INI1(173-257). Les deux protéines sont mélangées dans le même tampon L1, puis la protéase His-TEV est ajoutée pour une protéolyse spécifique du tag 6HisMBP et l'échantillon est mis à dialyser. Après la dialyse une filtration sur gel est effectuée. Le tag 6HisMBP est coélué avec IN(50-F185K-212) mais séparément d'INI1(173-257) (Figure 24).

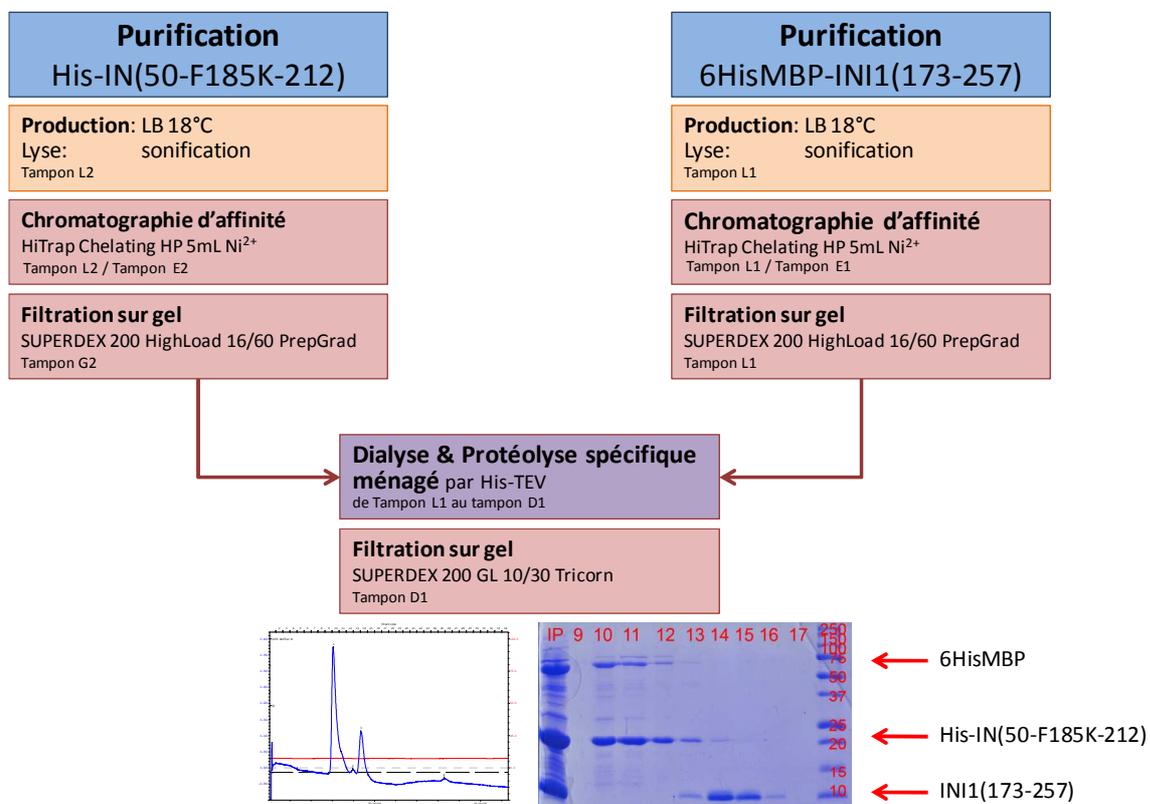


Figure 24: Récapitulatif de l'essai de formation du complexe entre 6His-tag - cœur catalytique (50-212) de l'intégrase F185K et INI1(173-257) (ratio 1:1) à NaCl 250 mM HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM (tampon D1) avec l'étiquette de solubilisation 6HisMBP.

## Purification du complexe Intégrase / LEDGF / INI1(173-290)

### Optimisation de la production

La formation du complexe INI1(173-257)/IN(50-F185K-212) étant impossible dans les conditions étudiées, une autre construction a été utilisée INI1(173-290). Pour éliminer l'étape de clivage de l'étiquette, l'hexahistidine (6His) a été choisie. L'utilisation d'un autre tag a nécessité une optimisation de production à l'échelle du litre de la construction 6His-INI1(173-290). L'optimisation de la production à l'échelle du litre s'est effectuée par variation de la température d'induction, variation du temps de production et l'addition ou non de saccharose (Figure 25). La production à 18 °C avec addition de 10 % de saccharose a été retenue.

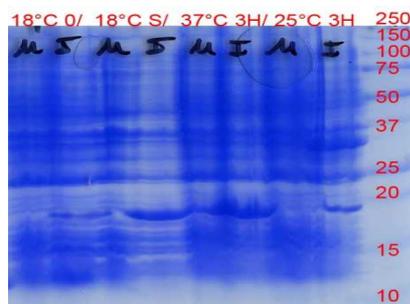


Figure 25 SDS-PAGE 15 %. Test de production à l'échelle du litre. U = avant induction ; I = après induction. La production s'effectue soit : à 18 °C sans saccharose ; à 18 °C avec 10 % de saccharose ; à 37 °C pendant 3 heures ; et à 25 °C pendant 3 heures.

Lors de la purification de INI1(173-290) des problèmes de reproductibilité ont été rencontrés dû à la solubilité mal évaluée du fragment. Dans les tests d'expression, la fraction soluble d'INI1(173-290) était surestimée. Il est apparu que au cours des différentes purifications du fragment 6His-INI1(173-290) une proportion importante de la protéine s'agrégeait. Des tests de solubilité et de purification ont été réalisés. Prenant pour base le tampon HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM, différentes concentrations en sel et détergent ont été ajoutés (Figure 26). Le CHAPS est un détergent zwitterionique non dénaturant plus facilement dialysé que le zwittergent 3-12. Le zwittergent 3-12 (Z[3-12]) est utilisé ici comme témoin positif de solubilité.

VE3596	NaCl				
pH7,5	150mM	500mM	1M	2M	2,5M
O	-	-	-		
CHAPS 7mM	-	-	-	-	-
CHAPS 20mM	-		-	+	++
Z[3-12] 4mM		+++			

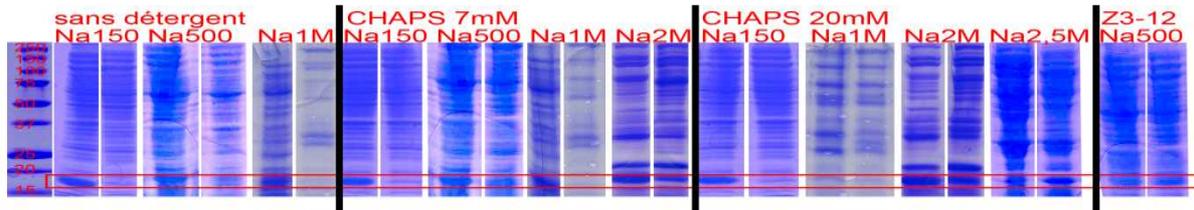


Figure 26: Analyse des tests de solubilité du fragment 6His-*INI1*(173-290) par SDS-PAGE (15 %). Pour les conditions de solubilité, la première piste correspond à l'extrait total (après sonification : témoin pour la quantité de protéine exprimée), la seconde piste correspond à l'extrait soluble (quantité de protéine soluble après centrifugation). Le tableau récapitulatif présente les résultats obtenus.

### Purification du fragment 6His-*INI1*(173-290)

Après la production en fiole, 12,4 g de cellules sont récoltées pour 3 L de culture (*E.coli* BL21 DE3, milieu LB, sélection ampicilline). Les conditions de croissance et de production sont identiques à celles d'*INI1* (173-257). La composition du tampon retenu pour la purification (Figure 27) du fragment 6His-*INI1*(173-290) est NaCl 2 M HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, βMEt 2 mM CHAPS 20 mM (tampon L4). Les culots bactériens sont repris dans 130 mL de tampon de lyse tampon L4 puis lysés par sonication. Après centrifugation, l'extrait soluble est injecté sur une colonne de chromatographie d'affinité (HiTrap Chelating HP 5 mL (GE Healthcare™)). Un gradient d'éluion de 0 à 1 M d'imidazole permet d'éluier le fragment 6His-*INI1*(173-290) vers 180 mM d'imidazole. Une seconde étape de purification par filtration sur gel (SUPERDEX 200 HighLoad 16/60 PrepGrad) dans le même tampon permet d'obtenir 6,2 mg de protéine d'intérêt et d'éliminer des contaminants protéiques et l'imidazole. Le fragment 6His-*INI1*(173-290) est maintenant prêt pour l'étape de formation du complexe ternaire par dialyse en flux continu.

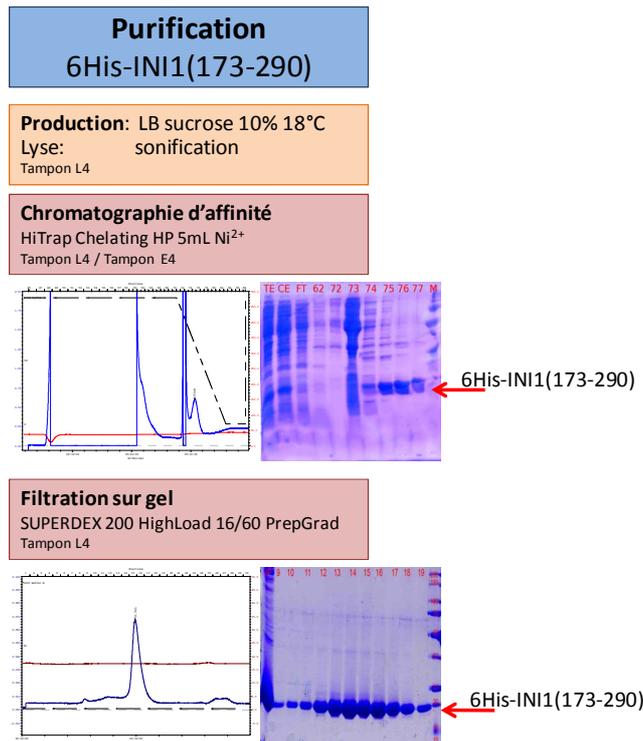


Figure 27: Récapitulatif de la production et de la purification du 6His-INI1 (173-290). Pour 12,3g de cellules (3L), 12,4 mg de protéine d'intérêt sont obtenus en fin de première étape (affinité) soit 1 mg/g de cellules, et 6,2 mg sont obtenus en fin de seconde étape (filtration sur gel) soit environ 0,5 mg/g de cellules. Le tampon L4 a la composition suivant : NaCl 2 M HEPES 50 mM pH7.5 MgCl<sub>2</sub> 5 mM βMEt 2mM CHAPS 20 mM ; et le tampon E4 : NaCl 2 M HEPES 50 mM pH7.5 MgCl<sub>2</sub> 5 mM βMEt 2mM CHAPS 20 mM imidazole 1000 mM

### *Purification du complexe Intégrase / LEDGF*

Purification de GST-Intégrase pleine longueur type sauvage

Après la production en fiole, 22,4 g de cellules sont récoltées pour 6 L de culture (*E.coli* BL21 DE3, milieu LB, sélection ampicilline). Les bactéries sont cultivées à 37 °C puis à 18 °C après induction par 0,5 mM d'IPTG. La production est effectuée sur 12 heures. Les culots bactériens sont repris dans 250 mL de tampon de lyse L3 (NaCl 1 M HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, βMEt 2 mM CHAPS 7 mM) puis lysés dans un système de traitement de fluides à fort cisaillement (microfluidizer®). Après centrifugation, l'extrait soluble est injecté sur une colonne de chromatographie d'affinité (GSTrap HP 5 mL (GE Healthcare™)). L'élution s'effectue en ajoutant 50 mM de glutathion réduit. Cette étape de purification permet d'obtenir 118 mg de GST-IN soit 5,2 mg /g de cellules (Figure 28).

## Purification de 6His-LEDGF pleine longueur type sauvage

Après la production en fiole, 45,6 g de cellules sont récoltées pour 12 L de culture (*E.coli* BL21 DE3, milieu LB, sélection ampicilline, chloramphénicol). Les conditions de croissance et de production sont identiques à celles de GST-Intégrase. Les culots bactériens sont repris dans 450 mL de tampon L3 puis lysés dans un système de traitement de fluides à fort cisaillement (microfluidizer®). Après centrifugation, l'extrait soluble est injecté sur une colonne de chromatographie d'affinité HiTrap Chelating HP 5 mL (GE Healthcare™). Un gradient d'élution de 0 à 800 mM d'imidazole permet d'éluer 6His-LEDGF vers 480 mM d'imidazole. Cette étape de purification permet d'obtenir 275 mg de 6His-LEDGF soit 6,0 mg / g de cellules (Figure 29).

## Formation du complexe Intégrase / LEDGF

La formation du complexe s'effectue en excès (2x) de la protéine la plus soluble 6His-LEDGF. Les deux protéines sont dans le même tampon (les concentrations d'imidazole et de glutathion réduit ne sont pas pris en compte, car n'ayant pas d'incidence). Le mélange s'effectue avec 118 mg de GST-intégrase et 240mg de 6His-LEDGF, soit au total 358 mg de protéines. L'étape suivante a pour objectif la formation effective du complexe. La force ionique (concentration en NaCl) est progressivement diminuée et le détergent (CHAPS) est éliminé par dialyses successives. L'élimination progressive des agents solubilisants (NaCl et CHAPS) force les partenaires protéiques à s'associer en complexe stable à faible force ionique et sans détergent. Après les trois dialyses (tampon final NaCl 0,15 M HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, βMEt 2 mM), la quantité de protéines est de 294 mg soit un rendement de 88 %. L'excès de 6His-LEDGF non complexé à GST-intégrase est éliminé en incubant le complexe sur des billes de glutathion. Les billes de glutathion sépharose 4B (GE Healthcare™) retiennent spécifiquement la GST-intégrase qui est maintenant complexée à 6His-LEDGF. Après élution par ajout de 50 mM de glutathion réduit, la quantité du complexe GST-Intégrase / 6His-LEDGF est estimée à 66 mg. La protéolyse spécifique ménagée permet de séparer le tag GST de l'intégrase et le tag 6His de LEDGF. Le tag GST est ensuite éliminé par chromatographie par gravité (batch) de glutathion sépharose 4B (GE Healthcare™). Le complexe n'étant plus immobilisé sur la résine, le lavage des billes permet d'obtenir 13 mg de complexe intégrase / LEDGF. Le complexe intégrase / LEDGF après concentration sur Amicon Ultra 15 50kDa (Millipore™) est injecté sur une colonne de filtration sur gel. Après

cette ultime étape, la quantité de complexe intégrase / LEDGF obtenu est de 6 mg. Ce complexe intégrase / LEDGF est une des composants du complexe ternaire Intégrase / LEDGF / INI1(173-290) (Figure 28).

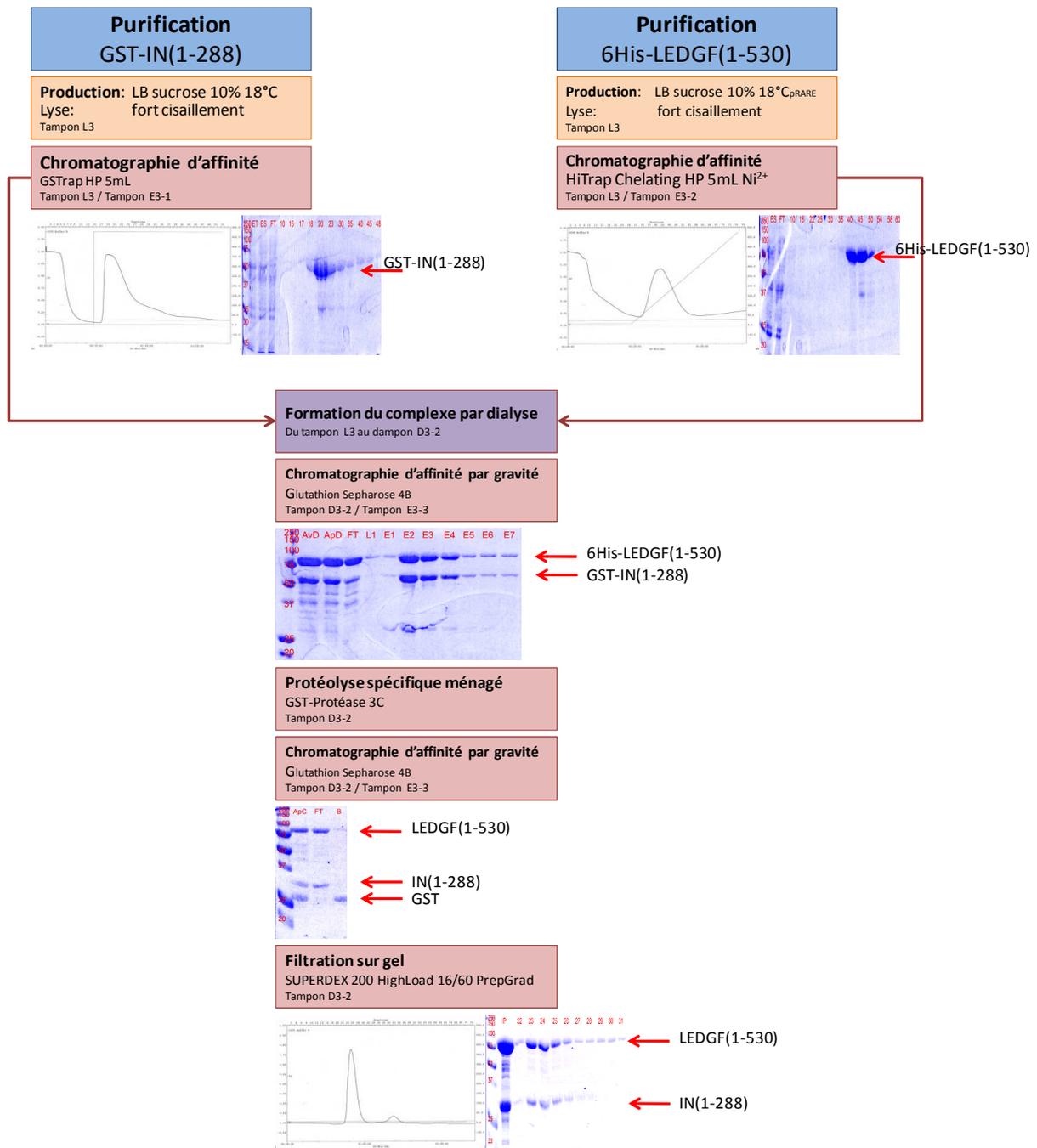


Figure 28: Récapitulatif de la purification et de la formation du complexe Intégrase / LEDGF. La composition des tampons est : tampon L3 NaCl 1 M HEPES 50 mM pH7.5 MgCl<sub>2</sub> 5 mM β-mercaptoéthanol 2mM CHAPS 7mM ; tampon E3-1 NaCl 1 M HEPES 50 mM pH7.5 MgCl<sub>2</sub> 5 mM β-mercaptoéthanol 2mM CHAPS 7mM glutathion réduit 50 mM ; tampon E3-2 NaCl 1 M HEPES 50 mM pH7.5 MgCl<sub>2</sub> 5 mM β-mercaptoéthanol 2mM CHAPS 7mM Imidazol 800 mM ; tampon D3-1 NaCl 0,6 M HEPES 50 mM pH7.0 MgCl<sub>2</sub> 5 mM β-mercaptoéthanol 2 mM CHAPS 1,3mM ; tampon D3-2 NaCl 0,15 M HEPES 50 mM pH7.0 MgCl<sub>2</sub> 5 mM β-mercaptoéthanol 2 mM ; tampon E3-3 NaCl 0,15 M HEPES 50 mM pH7.0 MgCl<sub>2</sub> 5 mM β-mercaptoéthanol 2 mM glutathion réduit 50 mM.

### *Formation du complexe ternaire*

Préalablement à la formation du complexe ternaire, les différents partenaires sont mis dans les conditions où INI1(173-290) est le plus soluble, soit NaCl 2 M HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, βMEt 2 mM CHAPS 20 mM. Ainsi le complexe intégrase LEDGF est dilué 2x dans le tampon NaCl 4 M HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, βMEt 2 mM CHAPS 40 mM. Le mélange s'effectue entre 6 mg de complexe intégrase LEDGF et 1,6 mg d'INI1(173-290) soit au total 7,6 mg de protéines. Le ratio des différentes protéines est de 4 intégrases, 2 LEDGF et 4 INI1(173-290). Par dialyse en flux continu, la force ionique (concentration en NaCl) est progressivement diminuée et le détergent (CHAPS) est éliminé pour former le complexe. Une ultime dialyse parachève la formation du complexe dans le tampon final NaCl 0,5 M HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, βMEt 2 mM, la quantité de complexe formé est de 3,4 mg. La filtration sur gel finale permet de co-éluer les trois partenaires du complexe ensemble dans le tampon NaCl 0,5 M HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, βMEt 2 mM. La masse de complexe intégrase / LEDGF / INI1(173-290) obtenu est de 1,5 mg pour 22,4 g de cellules exprimant la GST-intégrase, 45,6 g de cellules exprimant la 6His-LEDGF, et 3,2 g de cellules exprimant INI1(173-290) (Figure 29).

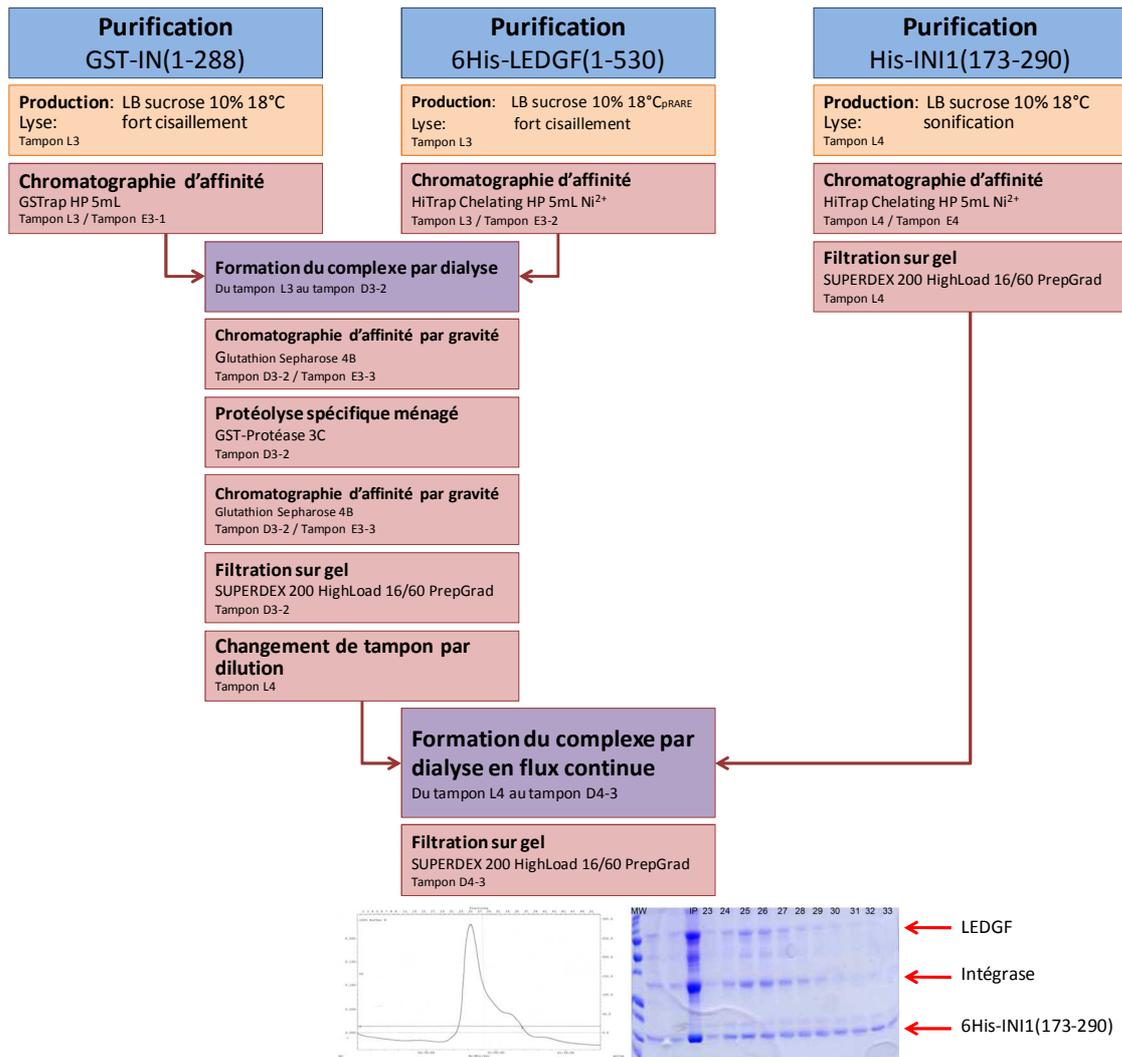


Figure 29: Récapitulatif de la purification et de la formation du complexe ternaire Intégrase / LEDGF / 6His-INI1(173-290). La composition des tampons est : tampon D4-1 NaCl 1 M HEPES 50 mM pH7.5 MgCl2 5 mM βMEt 2mM CHAPS 7 mM ; tampon D4-3 NaCl 0,5 M HEPES 50 mM pH7.0 MgCl2 5 mM β-mercaptoéthanol 2 mM

## Études fonctionnelles

### Matériel et méthodes

Les trois activités (liaison à l'ADN, *3' processing* et transfert de brin) de l'intégrase peuvent être mesurées selon divers procédés (Chow 1997; Deprez, Tauc et al. 2000; Deprez, Tauc et al. 2001; Wang, Klock et al. 2005; Merkel, Andrade et al. 2008; Merkel, Andrade et al. 2009). L'anisotropie de fluorescence est la méthode choisie pour la liaison à l'ADN et le *3' processing* (Merkel, Andrade et al. 2009). Les différents tests fonctionnels ont été réalisés à l'aide d'oligonucléotides marqués fluorescent (fluorescéine (FAM)). Les oligonucléotides marqués ont été synthétisés par IBA™ GmbH et les oligonucléotides standards (non marqués) ont quant à eux été produits par SIGMA Life Science™. Les oligonucléotides ont été hybridés selon les besoins. Deux séquences différentes d'ADN ont été utilisées : l'ADN mimant l'extrémité U5 de l'ADN viral et l'ADN mimant une séquence cible génomique.

### Anisotropie de fluorescence

#### Principe

Une molécule fluorescence qui est excitée par une lumière linéairement polarisée émettra elle-même une fluorescence linéairement polarisée. Une population de molécules fluorescentes dans leur état fondamental a une répartition naturellement isotrope. Dans cette population, une lumière polarisée excitera une sous-population ayant des moments dipolaires répartis de manière préférentielle (anisotrope) en faveur de la direction de polarisation de la lumière d'excitation. Ainsi les molécules photo-sélectionnées sont spécifiquement excitées quand leur orientation moléculaire coïncide avec la direction de polarisation de la lumière d'excitation. Comme la distribution de molécules à l'état excité est anisotrope, il en résulte une lumière de fluorescence elle-même polarisée en fonction de l'orientation moyenne de la sous-population de molécules excitées au moment de l'émission.

La valeur de la mesure de l'anisotropie de fluorescence est le résultat de la perte partielle ou totale de l'anisotropie de distribution résultant de la photo-sélection. Cette valeur est fonction pour une large part de la diffusion brownienne de rotation qui entraîne un déplacement angulaire pendant le temps de vie du fluorophore. Ainsi la lumière polarisée émise est progressivement dépolarisée en fonction de la diffusion rotationnelle. Plus la taille de la molécule est importante, plus la diffusion rotationnelle est lente, donc plus l'anisotropie de

fluorescence est grande et plus la molécule est petite, plus la diffusion rotationnelle est rapide, donc plus la valeur d'anisotropie est faible. L'anisotropie de fluorescence est définie selon la formule :

$$r = \frac{(I_{\text{par}} - I_{\text{per}})}{(I_{\text{par}} + 2I_{\text{per}})}$$

Où  $r$  est l'anisotropie de fluorescence. L'intensité de fluorescence  $I_{\text{par}}$  est mesurée lorsque les polarisateurs d'excitation et d'émission sont parallèles, elle reflète la proportion des molécules ayant subi une faible diffusion rotationnelle. A l'inverse,  $I_{\text{per}}$  est l'intensité de fluorescence lorsque les polarisateurs d'excitation et d'émission sont perpendiculaires, et reflète la proportion des molécules ayant subi une forte diffusion rotationnelle. La différence est normalisée par l'intensité de fluorescence totale définie par  $I_{\text{par}} + 2I_{\text{per}}$ .

L'anisotropie de fluorescence peut également être exprimée en fonction des trois paramètres  $\tau_c$ ,  $\tau_F$  et  $r_0$  : le temps de corrélation rotationnelle  $\tau_c$  d'une molécule fluorescente est fonction de sa masse  $M$ , de son volume  $v$ , de son volume spécifique d'hydratation  $\delta$ , de la viscosité  $n$  et de la température  $T$ . Ce paramètre est inversement proportionnel au coefficient de diffusion brownien de rotation

$$\tau_c = \frac{M(v + \delta)n}{RT}$$

L'anisotropie est fonction de  $\tau_c$  et du temps de vie du fluorophore :  $\tau_F$  *i.e.* le temps de latence entre l'absorption des photons excitateurs et l'émission des photons fluorescents. Le temps de vie du fluorophore définit donc la fenêtre temporelle durant laquelle la diffusion rotationnelle de la molécule sera considérée.

$$\frac{r_0}{r} = 1 + \frac{\tau_F}{\tau_c}$$

La valeur  $r_0$  représente l'anisotropie limite pour un fluorophore donné dans un milieu rigide (molécule immobile). Les valeurs d'anisotropie peuvent être prédites et sont presque reliées linéairement à la masse de la molécule jusqu'à ce que la valeur de  $\tau_c$  approche le tiers de celle de  $\tau_F$ .

Les tests fonctionnels ont été mis au point dans l'optique d'un criblage haut-débit d'inhibiteurs. Ils ont été réalisés dans un premier temps en cuvette de 10mm, puis adaptés à un lecteur de plaque PheraStar© (BMG LabTech™).

### *Liaison à l'ADN*

#### Principe

Le principe du test de liaison à l'ADN (Figure 30) est basé sur l'augmentation d'anisotropie quand l'ADN marqué est lié au complexe protéique. A concentration d'ADN fixe, la concentration en complexe protéique est augmentée progressivement. Plusieurs facteurs doivent être pris en considération. Le changement de masse dû à la liaison avec la protéine, la masse absolue des composés, la concentration de fluorophore requise pour une détection robuste, et la constante de dissociation de l'interaction. La réaction de liaison entre l'ADN marqué [A] et le complexe protéique [C] ayant n sites de liaison à l'ADN, s'écrit :



A l'équilibre, la constante de dissociation globale  $K_d$  de la réaction s'écrit :

$$K_d = \frac{[C][A]^n}{[CA_n]}$$

Le taux de saturation des sites de liaison à l'ADN :

$$\vartheta = \frac{[A_{lié}]}{[C_{tot}]} = \frac{f[A_{tot}]}{[C_{tot}]}$$

où f est la fraction d'ADN lié de l'ensemble de l'ADN présent  $A_{tot}$  et  $[C_{tot}]$  est la concentration totale de complexe. L'anisotropie observée est la somme des anisotropies individuelles de chaque composant pondérées par la brillance.

$$r_{obs} = \sum_i B_i r_i$$

Les composants sont l'ADN marqué libre  $[A_{lib}]$  et lié  $[A_{lié}]$  correspondent à  $r_0$  la mesure d'anisotropie en l'absence de complexe, et  $r_t$  la mesure d'anisotropie à saturation (plateau) lorsque tout le complexe ADN/complexe protéique est formé. La fraction f représentant la partie liée de  $[A_{tot}]$  peut s'écrire :

---

Il est ainsi possible de modifier l'équation de Scatchard afin d'ajuster l'anisotropie,  $r$  :

---

Cette équation permet ainsi de prévoir les paramètres de liaison de l'ADN avec le complexe protéique.

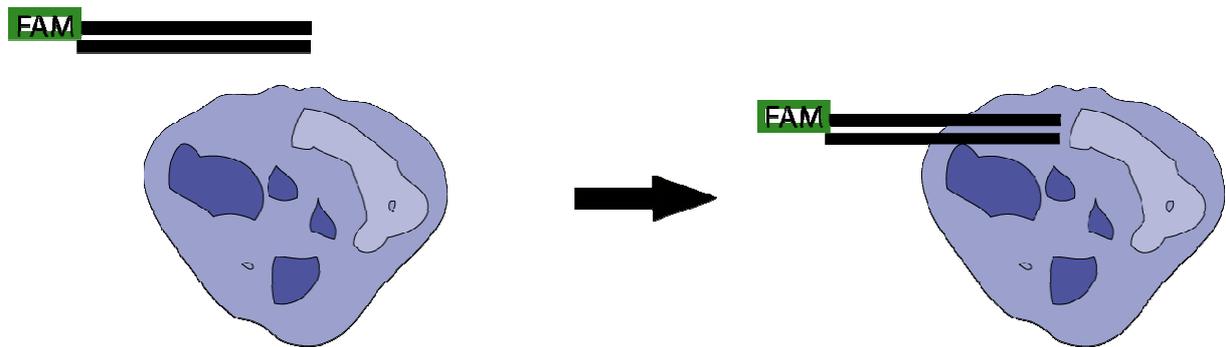


Figure 30: Principe du test de liaison à l'ADN. En violet le complexe protéique, en noir la séquence de l'ADN marqué à la fluorescéine.

### Protocole

En pratique, la titration a été réalisée à 25°C sur 12 puits dans une plaque Greiner™ Black HiBase 384©. Le volume réactionnel final est de 20 µL dans un tampon NaCl 150 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM. La concentration finale d'ADN est de 2 nM dans chacun des 12 puits de la mesure. Typiquement, la concentration finale de complexe protéique varie de 0 nM à 300 nM. Dans chaque puits, 10 µL d'ADN marqué (solution stock à 4 nM) sont déposés pour une concentration finale de 2 nM, puis 10 µL d'une solution de concentration variable de complexe sont ajoutés. Pour obtenir les différentes concentrations de complexe nécessaires, une dilution en cascade d'un facteur 1,5 est effectuée dans des tubes Eppendorf 1,5mL. La première dilution s'opère pour obtenir une concentration de 600 nM dans un volume de 75 µL. 50 µL sont pris puis diluer dans 25 µL de tampon seul. Cette étape est répétée 10 fois. Onze solutions de concentration décroissante en complexe sont ainsi préparées (Figure 31).

puits	concentration stock	concentration dans le puits
1	600,0 nM	300,0 nM
2	400,0 nM	200,0 nM
3	266,7 nM	133,3 nM
4	177,8 nM	88,9 nM
5	118,5 nM	59,3 nM
6	79,0 nM	39,5 nM
7	52,7 nM	26,3 nM
8	35,1 nM	17,6 nM
9	23,4 nM	11,7 nM
10	15,6 nM	7,8 nM
11	10,4 nM	5,2 nM
12	0,0 nM	0,0 nM

Figure 31: La concentration en nM de complexe suivant une dilution en cascade de facteur 1,5.

La plaque est placée dans le lecteur à 25 °C et la lecture est réalisée immédiatement. Trois mesures sont effectuées à 5 minutes d'intervalle.

### Analyse

L'analyse s'effectue à l'aide du logiciel Origin® version 6.1. Pour comparer les différentes séquences d'ADN entre elles, la normalisation s'effectue selon le principe :

$$r = \frac{r - r_0}{r_t - r_0}$$

Où  $r_0$  est la mesure d'anisotropie en l'absence de ligand, et  $r_t$  est la mesure d'anisotropie à saturation (plateau) lorsque tout le complexe ADN/complexe protéique est formé. L'analyse grâce à l'équation de Scatchard modifiée permet de déterminer la constante de dissociation globale et le nombre de sites de liaison à l'ADN du complexe.

### 3'Processing

#### Principe

Un ADN double brin spécifique de 40 paires de bases est marqué en 3'. Le principe du test de l'activité enzymatique de 3'processing (Figure 32) est basé sur la diminution d'anisotropie quand les deux derniers nucléotides coté 3' marqué est excisé spécifiquement par le complexe protéique qui s'est préalablement lié à l'ADN spécifiquement. La séquence utilisée pour le

test d'activité est identique à celle de l'ADN viral utilisé précédemment. Les concentrations d'ADN et de complexe protéique sont fixes.

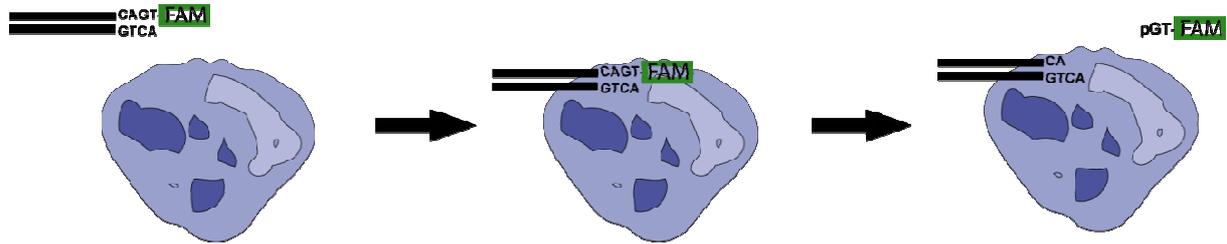


Figure 32: Principe du test pour l'activité de 3'processing. L'ADN marqué se fixe sur le complexe protéique. L'activité enzymatique de l'intégrase entraîne l'excision spécifique du dinucléotide pGT(FAM).

### Protocole

En pratique, la mesure s'effectue en plaque Greiner™ Black 96© à 37 °C. Le volume final de la réaction s'effectue dans 100 µL dans le tampon NaCl ~10 mM, BisTris 25 mM pH6.5, MgCl<sub>2</sub> 10 mM, DTT 5 mM. La concentration finale d'ADN marqué est de 50 nM, et la concentration finale de complexe protéique est de 200 nM. Dans un premier temps, l'ADN marqué est dilué dans le tampon, une mesure d'anisotropie est effectuée, puis le complexe protéique est ajouté. Le puits est recouvert de 50 µL d'huile de paraffine afin d'éviter l'évaporation. La mesure d'anisotropie se déroule ensuite sur 6 heures (une mesure toute les 85 secondes).

### Analyse

L'analyse s'effectue à l'aide du logiciel Origin® version 6.1. La normalisation s'effectue selon la formule suivante :

$$\frac{\text{Anisotropie}}{\text{Anisotropie}} = \frac{\text{Anisotropie}}{\text{Anisotropie}}$$

Où est la valeur maximale d'anisotropie mesurée. Cette valeur correspond à la liaison totale de l'ADN viral marqué au complexe protéique. La valeur est une mesure expérimentale d'un dinucléotide pGT(FAM) marqué dans les mêmes conditions. Cette valeur correspond à la valeur limite d'anisotropie de 100 % de pGT(FAM) libre.

## Résultats

### Intégrase et liaison à l'ADN

La liaison à l'ADN de l'intégrase a été étudiée par anisotropie de fluorescence, la formation du complexe protéine/ADN résultant en une augmentation de l'anisotropie associée du fait de l'accroissement en masse. Les courbes obtenues après analyse avec le logiciel Origin© sont présentées en Figure 34.

L'intégrase lie des séquences d'ADN distinctes : les extrémités de l'ADN viral (U5 et U3) sont liées spécifiquement, alors que l'ADN génomique est lié de manière peu spécifique. Pour être au plus proche de la réalité cellulaire, les tests *in vitro* ont été effectués en duplicata et sur deux séquences différentes : l'une de 40 paires de bases mimant l'extrémité U5 de l'ADN viral, l'autre de 49 paires de bases mimant l'ADN génomique cible. L'activité de liaison à l'ADN des deux complexes IN/LEDGF et IN/LEDGF/INI1(173-290) pour chaque ADN (viral et cible) a été mesurée dans les mêmes conditions : NaCl 150 mM, HEPES 50 mM pH7.0, MgCl<sub>2</sub> 5 mM, β-mercaptoéthanol 2 mM à 25 °C.

En résumé, deux ADN ont été utilisés :

- ADN 40mer avec la séquence de l'extrémité U5 viral (ADN viral)

```
6FAM-5' -GACTACGGTTCAAGTCAGCGTGTGGAAAATCTCTAGCAGT-3'
      3' -CTGATGCCAAGTTCAGTCGCACACCTTTTAGAGATCGTCA-5'
```

- ADN 49mer mimant l'ADN génomique cible (ADN cible)

```
5' -AGTTAAGTGCTGAATTATGATAGTAATCAATATCTACTCCTAACCTCTT-3' -6FAM
3' -TCAATTCACGACTTAATACTATCATTAGTTATAGATGAGGATTGGAGAA-5'
```

Au sein du complexe IN/LEDGF, l'intégrase est tétramérique (Michel, Crucifix et al. 2009). D'après le modèle développé au laboratoire, trois sites de liaison à l'ADN sont nécessaires au processus d'intégration. Les deux sites de liaison à l'ADN viral (un pour chaque extrémité U5 et U3) sont probablement spécifiques et équivalents. Le site de liaison à l'ADN cible n'est pas spécifique. Les paramètres de modélisation de liaison de l'ADN avec le complexe ont été choisis en fonction de ce modèle. Les variables de l'équation de Scatchard modifiée sont ajustées pour obtenir un coefficient de détermination (R<sup>2</sup>) le plus proche de un. Les valeurs de départ des paramètres sont :

Pour l'ADN viral marqué :

$$r_0 = 0 \quad r_t = 1 \quad [A_{\text{tot}}] = 2 \text{ nM} \quad n = 0,5$$

Pour l'ADN cible marqué :

$$r_0 = 0 \quad r_t = 1 \quad [A_{\text{tot}}] = 2 \text{ nM} \quad n = 1$$

La constante de dissociation  $K_d$  et la valeur finale d'anisotropie  $r_t$  sont laissés variables, les autres variables ( $n$  et  $[A_{\text{tot}}]$ ), connues, sont fixées. Selon les données traitées, un nombre variable d'itération de l'algorithme de Levenberg-Marquardt est nécessaire pour calculer une solution numérique de minimisation. Le nombre de degrés de liberté étant identiques, le calcul du coefficient de détermination  $R^2$  est un bon indicateur de la qualité de la minimisation.

Les expériences de liaison à l'ADN ont été réalisées en absence et en présence d'ADN froid dans des proportions 100 fois supérieures à l'ADN marqué (soit 200 nM). La séquence des ADN froids est identique à celle des ADN marqués. En théorie, l'ADN viral froid agit comme un inhibiteur compétitif de l'ADN viral marqué. Pour le complexe IN/LEDGF, la constante de dissociation  $K_d$  de la liaison à l'ADN viral ne varie pas entre l'absence et la présence de 200 nM d'ADN cible froid. Cependant, la constante d'affinité est divisée par cinq en présence de 200 nM d'ADN viral froid. Les deux sites de fixation de l'ADN viral semblent être spécifiques puisque la constante de dissociation ne varie qu'en présence du même ADN viral froid. (Figure 33).

La constante de dissociation de la liaison à l'ADN cible n'est pas significativement différente dans les trois cas sans ADN froid, avec ADN viral froid, avec ADN cible froid (Figure 33/Figure 34). L'ADN froid de quelque nature qu'il soit (viral ou cible) ne semble rentrer en compétition ni avec l'ADN viral ni avec l'ADN cible. L'effet sur la liaison à l'ADN cible ou viral serait donc nul. L'ADN viral se fixe préférentiellement sur les sites spécifiques de liaison à l'ADN viral, il est normal que la liaison avec l'ADN cible n'en soit pas affecté. En revanche, la constante de dissociation observé pour le site de liaison à l'ADN cible ne varie pas car l'affinité pour le site est plus faible. L'ajout d'ADN cible froid d'un excès plus important permettrait d'obtenir un effet décroissant significatif.

Le complexe ternaire IN/LEDGF/INI1(173-290) a une affinité pour l'ADN viral et cible environ 4 fois moindre que le complexe IN/LEDGF (Figure 34c et f). Le domaine de liaison à

l'intégrase d'INI1(173-290) réduit l'affinité pour l'ADN du complexe IN/LEDGF. Cette réduction est probablement due à la baisse de l'accessibilité des sites de liaison à l'ADN de l'intégrase. Cependant, la constante globale de dissociation pour l'ADN viral double en présence d'ADN froid (viral ou cible). Concernant le site de l'ADN cible, le comportement de l'affinité du complexe IN/LEDGF/INI1(173-290) suit celui du complexe IN/LEDGF. Pour le complexe IN/LEDGF/INI1(173-290), il est à noter que les courbes de titration en présence d'ADN froid (viral ou cible) décrivent une légère sigmoïde ce qui pourrait trahir une coopérativité entre les sites de liaison à l'ADN. Dans ce cas, le modèle utilisé ne serait plus satisfaisant car il implique que les sites de liaison soient identiques.

En résumé, en présence d'INI1(173-290) l'affinité pour l'ADN cible et viral est diminuée. Les courbes a) et b) de la Figure 34 montre la spécificité des sites de liaison à l'ADN viral. La diminution de l'affinité des sites de l'ADN viral en présence d'ADN viral froid en est la conséquence.

	IN/LEDGF									IN/LEDGF/INI1(173-290)								
	sans			ADN viral froid			ADN cible froid			sans			ADN viral froid			ADN cible froid		
	n	Kd	R <sup>2</sup>	n	Kd	R <sup>2</sup>	n	Kd	R <sup>2</sup>	n	Kd	R <sup>2</sup>	n	Kd	R <sup>2</sup>	n	Kd	R <sup>2</sup>
ADN viral marqué	0.5	4.4 nM <sup>2</sup>	0.99	0.5	21.7 nM <sup>2</sup>	0.99	0.5	5.3 nM <sup>2</sup>	0.99	0.5	20.1 nM <sup>2</sup>	1.00	0.5	49.3 nM <sup>2</sup>	0.984	0.5	46.3 nM <sup>2</sup>	0.952
ADN cible marqué	1.0	5.4 nM	0.97	1.0	8.6 nM	0.99	1.0	6.5 nM	0.97	1.0	23.5 nM	0.95	1.0	54.1 nM	0.94	1.0	33.8 nM	0.928

Figure 33 : Tableau des valeurs modélisées. n = nombre de sites de liaison au complexe/ ADN ; Kd = constante de dissociation ; R<sup>2</sup> = coefficient de détermination, sont reportées par l'ADN marqué (viral ou cible) en fonction du complexe (IN/LEDGF ou IN/LEDGF/INI1(173-290)) avec ou sans ADN froid (viral ou cible). Pour l'ADN viral, le Kd est exprimé en nM<sup>2</sup> et pour l'ADN cible, le Kd est exprimé en nM.

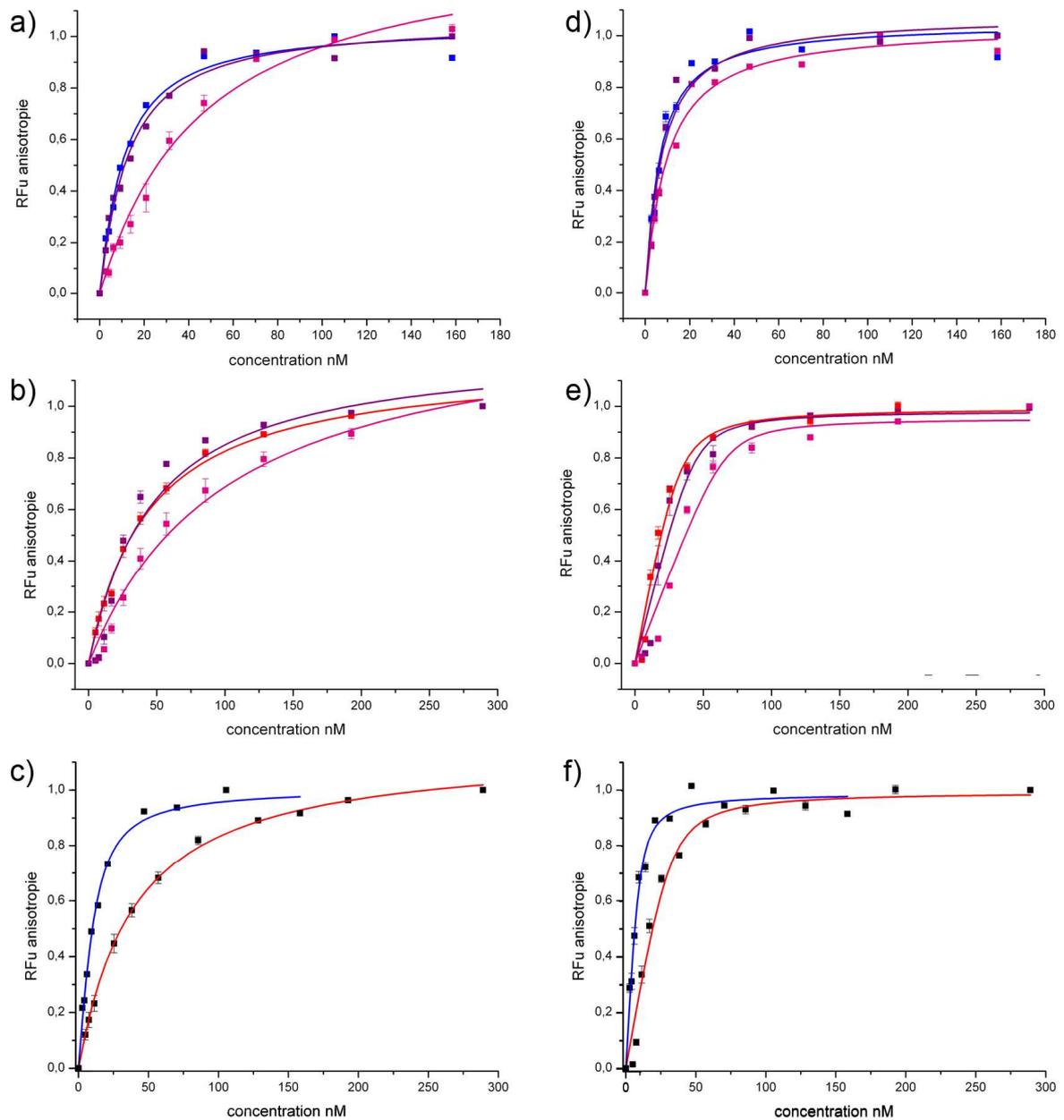


Figure 34: Expériences et calculs des sites de liaisons à l'ADN des complexes IN/LEDGF (bleu) et IN/LEDGF/INI1(173-290) (rouge). a) courbe de titration de l'ADN viral contre le complexe IN/LEDGF (sans (bleu) ou avec ADN froid (viral (rose) ou cible (pourpre)). b) courbe de titration de l'ADN viral contre le complexe IN/LEDGF/INI1(173-290) (sans ADN froid (rouge) ou avec ADN froid (viral (rose) ou cible (pourpre)). c) comparaison des courbes de titration de l'ADN viral contre IN/LEDGF (bleu) et IN/LEDGF/INI1(173-290) (rouge). d) courbe de titration de l'ADN cible contre le complexe IN/LEDGF (sans (bleu) ou avec ADN froid (viral (rose) ou cible (pourpre)). e) courbe de titration de l'ADN cible contre le complexe IN/LEDGF/INI1(173-290) (sans (rouge) ou avec ADN froid (viral (rose) ou cible (pourpre)), f) comparaison des courbes de titration de l'ADN cible contre IN/LEDGF (bleu) et IN/LEDGF/INI1(173-290) (rouge).

## Intégrase et 3' processing

L'activité catalytique 3' processing de l'intégrase a été étudiée par anisotropie de fluorescence en suivant la cinétique. La diminution de l'anisotropie associée à l'activité enzymatique du complexe protéique a été mesurée (*i.e.* la libération dans le milieu réactionnel du dinucléotide pGT(FAM) excisé spécifiquement). La fraction d'anisotropie due à l'activité enzymatique doit être séparée de celle due à la liaison à l'ADN. Il est nécessaire que la totalité de l'ADN soit lié au complexe. Bien que la totalité du dinucléotide pGT(FAM) soit larguée dans le milieu (Guiot 2006), il est nécessaire de prendre en compte la possibilité que le complexe peut se dégrader à une concentration saline inférieure à 10 mM, à 37 °C et pendant un laps de temps de 6 heures. En effet, le complexe dégradé ne lie plus l'ADN qui resté libre diminue la mesure de l'anisotropie globale. L'utilisation des complexes mutants IN(D64E)/LEDGF et IN(D64E)/LEDGF/INI1(173-290) permet de soustraire à la fraction d'anisotropie due à cette perte de liaison et la fraction d'anisotropie due à l'activité 3' processing. Le résidu 64 de l'intégrase du VIH-1 appartient à la triade catalytique D64 D116 E152 nécessaire à l'activité enzymatique de l'intégrase. La mutation D64E entraîne l'inactivation du site catalytique empêchant la réaction de 3' processing et de transfert de brin. Cette mutation D64E est cependant neutre pour la liaison à l'ADN (Engelman and Craigie 1992; Drelich, Haenggi et al. 1993; Jenkins, Esposito et al. 1997). Par cette méthode, il est possible de suivre la réaction en continu. La fraction d'anisotropie représentant cette perte de la liaison à l'ADN au complexe (Figure 35 a) diminue de 40 % de manière linéaire en 6 heures. Cette diminution est identique quelque soit le complexe envisagé (IN(D64E)/LEDGF ou IN(D64E)/LEDGF/INI1(173-290)). Afin de vérifier la spécificité pour l'ADN viral, des tests d'activité enzymatique ont été effectués en remplaçant l'ADN viral par un ADN quelconque marqué en 3'. Une diminution d'anisotropie comparable à celle des complexes mutants a été observée. Cependant la pente indique que le complexe IN(D64E)/LEDGF/INI1(173-290) est plus stable que IN(D64E)/LEDGF durant les trois première heures.

Plusieurs essais ont été effectués pour trouver les conditions où l'activité enzymatique est mesurable (Figure 35 b). A une force ionique de 150 mM de NaCl, l'activité enzymatique est négligeable comparée à la diminution d'anisotropie mesurée avec les complexes mutants inactifs. La réaction de 3' processing semble nécessiter une force ionique faible, inférieure à 10 mM de NaCl.

Après soustraction de la fraction d'anisotropie due à la perte de liaison (Figure 35 c), la comparaison de l'activité enzymatique des complexes IN(D64E)/LEDGF et IN(D64E)/LEDGF/INI1(173-290), permet de déduire qu'INI1(173-290) inhibe totalement l'activité enzymatique de *3' processing*. La courbe d'activité du complexe IN/LEDGF atteint un plateau après 4 heures à 37 °C alors que la fraction de GT<sup>FAM</sup> libéré n'est que de 50 %. Les 60 % du complexe resté en solution effectuent la réaction de *3' processing* sur 50 % du substrat disponible. Un complexe IN/LEDGF effectue donc la réaction sur un seul ADN. Le site catalytique de l'intégrase ayant l'activité enzymatique n'est pas régénéré, l'intégrase n'est donc pas processive.

En résumé, l'action d'INI1(173-290) au sein du complexe IN/LEDGF diminue la liaison à l'ADN viral et à l'ADN cible. Le domaine de liaison à l'intégrase d'INI1 (173-290) inhibe totalement l'activité catalytique *3' processing* de l'intégrase.

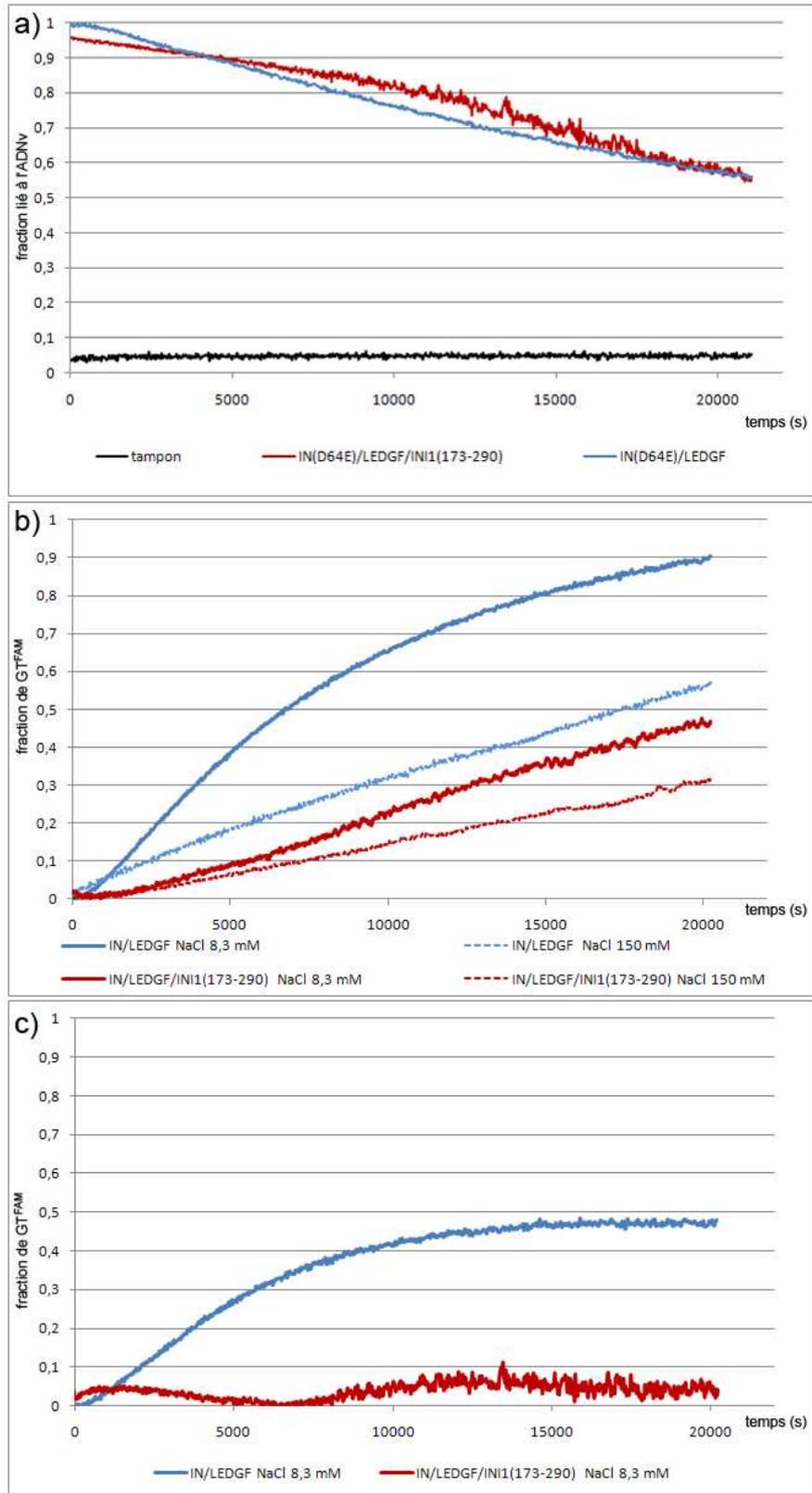


Figure 35 : Activité enzymatique 3' processing des complexes IN(D64E)/LEDGF/INI1(173-290) et IN(D64E)/LEDGF. a) Les contrôles négatifs de l'activité 3' processing des complexes mutants D64E (en rouge : IN(D64E)/LEDGF/INI1(173-290) ; en bleu : IN(D64E)/LEDGF) (à 37 °C pendant 6 heures) b) Fraction d'anisotropie de GTFAM non pondérées c) Fraction d'anisotropie de GTFAM pondérées par la diminution d'anisotropie respective des courbes des complexes mutants non actifs.

## Études structurales

### Matériels et méthodes

#### Introduction

Le principe est de combiner deux niveaux de résolution qu'offre la cristallographie par diffraction des rayons X pour une résolution atomique des fragments de protéines et la microscopie électronique pour une résolution de la dizaine d'angströms pour le complexe entier (Baker and Johnson 1996). Par cette approche, des modèles peuvent être développés mettant en exergue la relation structure-fonction.

#### La stratégie de cristallisation

Un cristal est un agencement périodique tridimensionnel de molécules liées par des interactions intermoléculaires non covalentes (interactions de Van der Waals, liaison hydrogènes et interactions ioniques). L'obtention de cristaux uniques est nécessaire pour déterminer la structure tridimensionnelle d'une protéine par diffraction des rayons X. La cristallogénèse consiste à favoriser la formation de cristaux en diminuant la solubilité de la protéine encourageant les interactions intermoléculaires spécifiques. Le caractère spécifique de ces interactions est fondamental car des interactions non spécifiques mènent à la précipitation du soluté. Au début les molécules de protéines sont solubles car stabilisées par des interactions protéine/solvant. La déstabilisation des interactions protéines/solvant au profit d'interactions protéine/protéine. La technique de cristallisation utilisée est la diffusion de vapeur soit en gouttes assises ou suspendues (Figure 36). La goutte est constituée d'un mélange de protéine/solution dont il est possible de varier le ratio. Le volume du réservoir est considéré comme infiniment grand par rapport au volume de la goutte. La solution réservoir dans la goutte est diluée ou concentrée par rapport à la solution dans le réservoir et il y a diffusion de vapeur de la goutte vers le réservoir jusqu'à obtention d'un équilibre. C'est au cours de ce processus que la protéine cristallise.

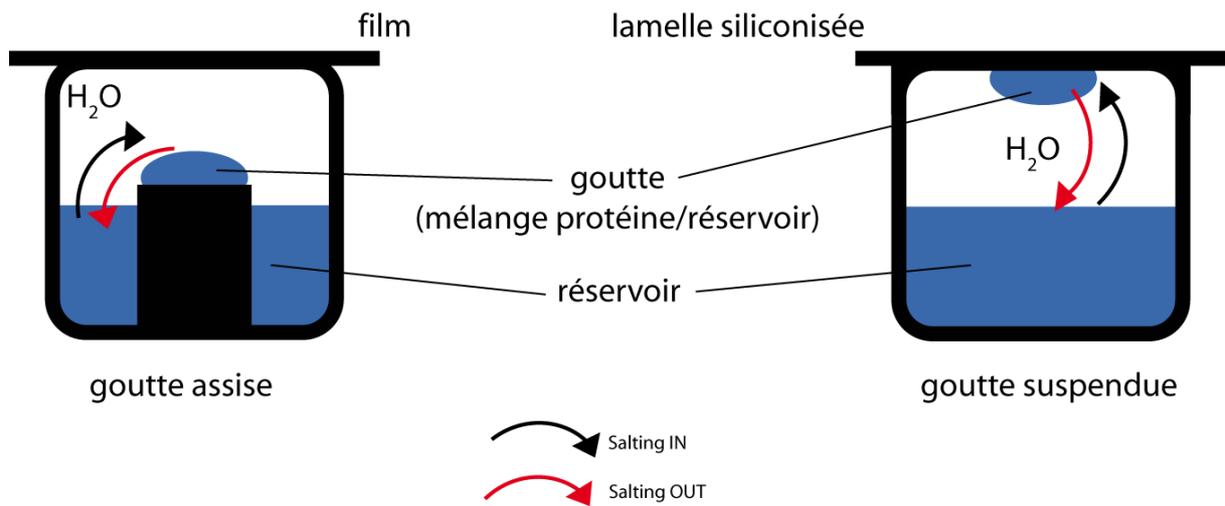


Figure 36: Représentation schématique de la technique de la diffusion de vapeur en goutte suspendue et en goutte assise.

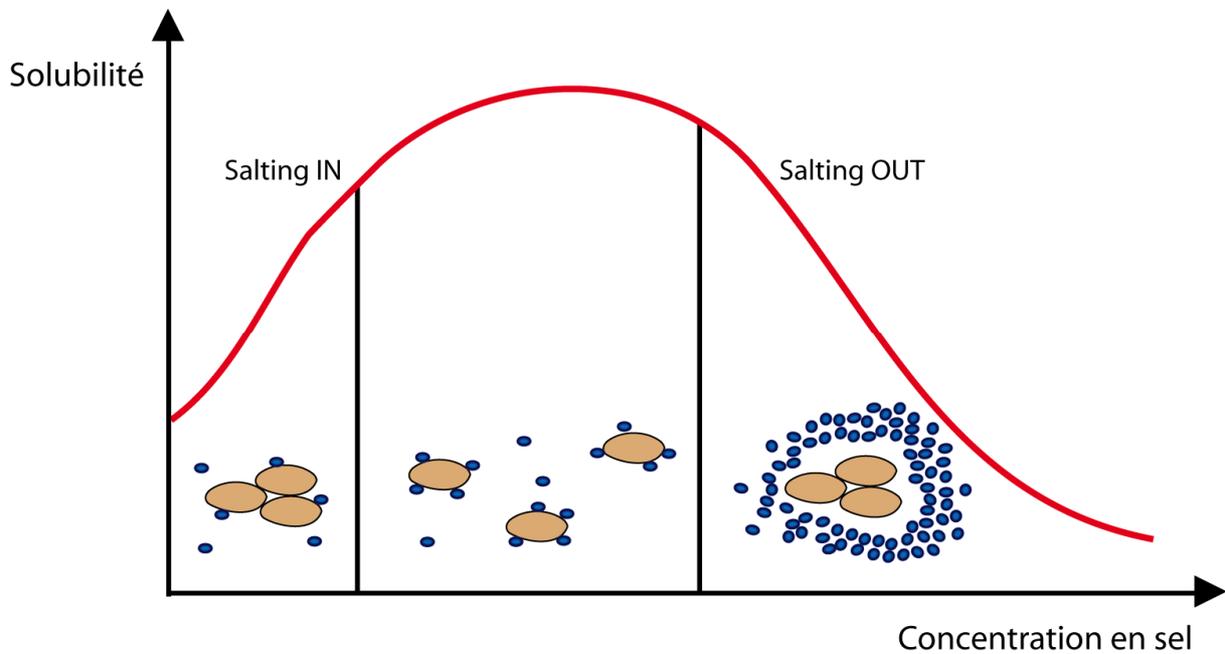


Figure 37: Représentation schématique du phénomène de « salting in/out » ; solubilité en fonction de la concentration en sel.

## La stratégie de microscopie électronique

### Introduction

La microscopie électronique moléculaire vise à obtenir un modèle en trois dimensions (3D) de complexes supramoléculaires en observant des particules isolées immobilisés sur une grille. En combinant les différentes méthodes de préparation d'échantillon, d'enregistrement des images et d'analyse des données, il est possible d'obtenir une enveloppe 3D du complexe supramoléculaire à une résolution de l'ordre d'une dizaine d'Angström.

### Les méthodes de préparations

Diverses méthodes sont utilisées pour amplifier le contraste, par addition d'atomes de haut numéro atomique tels que l'acétate uranyl ou l'acide phosphotungstique, soit sur le matériel biologique (coloration positive), soit autour de l'objet (coloration négative). Ainsi un atome d'uranium ( $Z=92$ ) diffuse 60 fois plus les électrons que le carbone ( $Z=6$ ). La coloration négative, la cryo-coloration négative en sandwich et la congélation dans une couche de glace amorphe (Figure 38) sont les techniques de préparation d'échantillon utilisées pour réaliser le modèle 3D.

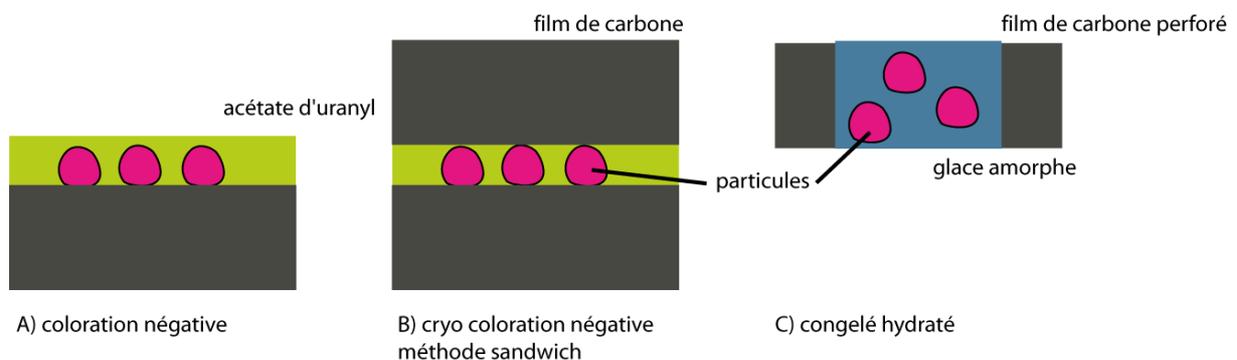


Figure 38: Les différentes techniques de préparation d'échantillon pour la microscopie électronique moléculaire.

### La méthode de coloration négative

La méthode de coloration négative consiste à inclure les particules à étudier dans une couche de sel d'atomes lourds (la plupart du temps de l'acétate d'uranyl). Le sel entoure l'échantillon et pénètre toutes les régions accessibles préservant de la déshydratation dans le vide du microscope. Les atomes lourds utilisés dans la coloration négative doivent être solubles dans l'eau, amorphes après séchage, stables sous le faisceau et ne doivent pas interférer avec le matériel biologique. Les particules sont enrobées dans un film constitué de matériel dense aux

électrons. Les électrons incidents sont proportionnellement diffusés élastiquement en fonction du numéro atomique  $Z$  car ils interagissent avec la charge positive du noyau. Plus le numéro atomique est élevé plus la diffusion élastique sera importante selon une proportionnalité de  $Z^{3/2}$ . Ce phénomène de diffusion élastique est responsable de la formation de l'image. Ainsi un atome ayant un numéro atomique élevé apparaîtra plus sombre qu'un atome dont le  $Z$  est inférieur. Le contraste entre deux points est la différence de densité électronique entre ces deux points. Si la densité électronique est élevée, une proportion importante des électrons incidents seront diffusés. L'échantillon biologique étant constitué d'atomes dont le numéro atomique (carbone, d'azote, d'hydrogène et d'oxygène et de quelques éléments représentés de façon mineure) est faible apparaît en clair sur l'image car la diffusion élastique est faible.

Pratiquement, l'échantillon biologique est adsorbé sur une grille recouverte d'un film de carbone rendu hydrophile pendant 1 minute. Ensuite, une solution contenant de l'acétate d'uranyle de 0,5% à 2% est ajoutée au complexe supramoléculaire. L'excès de liquide est ensuite éliminé à l'aide d'un papier filtre et la grille est séchée. Une meilleure répartition des particules est obtenue quand la grille est effluée avant adsorption. Pour cela, la grille est mise dans une enceinte où l'on peut créer un vide peu poussé avant de réaliser une décharge électrique qui ionisera les molécules d'air résiduelles, la surface des grilles sera ainsi rendue hydrophile. L'acétate d'uranyle permet une grande rapidité d'infiltration et présente un grain très fin. Même avec son pH très acide, les objets sont rarement détruits. L'acétate d'uranyle présente l'inconvénient de précipiter avec les ions  $\text{OH}^-$ , ainsi qu'en présence de lumière ou de groupements phosphate.

L'observation des échantillons est aisée avec cette technique, la préparation rapide et l'analyse d'images permet de mettre en évidence des détails structuraux à une résolution de 20 Å et permet d'identifier des vues caractéristiques de l'objet étudié, obtenir un premier modèle 3D et valider la possibilité de débiter une étude en cryo-microscopie (Brenner and Horne 1959). Cependant cette méthode ne donne cependant qu'une information de surface de l'échantillon car ce qui est observé est la fraction de l'objet qui exclut le colorant. Aucune information structurale interne ne pourra être obtenue. La préservation de l'objet n'est pas optimale car l'utilisation de sels d'atomes lourds entraîne une augmentation de la concentration saline qui peut déstabiliser les complexes protéiques. Mais l'échantillon est toutefois mieux préservé que lors du séchage à l'air car le moule de sel réduit les forces de tensions superficielles qui s'appliquent sur l'échantillon au cours du séchage et le sel masque les forces ioniques qui apparaissent lorsque les molécules d'eau sont retirées.

## La cryo-coloration négative et la méthode sandwich

La méthode de cryo-coloration négative en sandwich (CCNS) (Golas, Sander et al. 2003) minimise les variations de coloration, les particules sont prises en sandwich entre deux films de carbone fins après coloration uniforme à l'acétate d'uranyle puis congelées dans l'éthane liquide. La densité protéique la plus éloignée du film de carbone présente le même contraste que celle au contact du film. L'analyse d'images portera essentiellement sur les variations entre les molécules plutôt que sur les variations de coloration qui ne sont pas forcément liées à la structure de l'objet. La congélation des échantillons et leur observation à basse température réduit les effets des dommages d'irradiation. A  $-180^{\circ}\text{C}$ , la perte de résolution est 2 à 3 fois moins importante qu'à température ambiante. Cependant cette technique n'améliore pas le phénomène d'aplatissement de la molécule qui induit des erreurs dans la mesure de sa taille.

Pratiquement, une plaque de mica recouverte d'un film de carbone est plongée (selon un angle d'environ  $45^{\circ}$ ) dans la solution protéique. Le carbone se détache de la plaque de mica et flotte à la surface. Les protéines s'adsorbent sur le film de carbone puis l'ensemble est retiré de la solution protéique. La plaque de mica avec le film de carbone recouvert de protéines est plongée dans une solution d'acétate d'uranyle. Le film de carbone se détache et flotte à la surface, permettant ainsi la coloration des particules protéiques. Une grille de microscopie électronique est déposée sur la surface du film de carbone. Le film de carbone s'adsorbe à la grille, il peut donc être détaché de la plaque de mica. Un autre film de carbone avec un support en mica est flotté dans une solution d'acétate d'uranyle. La grille recouverte du film de carbone auquel sont adsorbées les protéines colorées est immergée dans la solution d'acétate d'uranyle et plaquée contre la face inférieure du film de carbone avant d'être détaché de la plaque de mica. Finalement, l'échantillon protéique se trouve dans un « moule » de colorant coincé entre deux films de carbone. L'ensemble est plongé dans l'éthane liquide puis observé à froid au microscope.

## La CRYO : hydratés congelés

Cette méthode (Dubochet, Adrian et al. 1988) s'approche le plus des conditions physiologiques. Cette technique consiste à maintenir les particules dans un état hydraté piégé dans de l'eau vitrifiée (glace amorphe) obtenue par un brusque plongeon de l'échantillon dans un agent cryogénique (éthane liquide). La vitesse de congélation est de l'ordre de  $2000^{\circ}\text{C}$  par seconde évitant la formation de cristaux de glace hexagonale opaque aux électrons. Contrairement à la coloration négative, l'absence d'atomes lourds préserve la force ionique

des particules et permet de s'affranchir des variations de contraste dues à une répartition inégale des atomes lourds sur la surface des particules. Ainsi le contraste ne provient que de la différence entre la particule et le milieu.

Les particules ne collapsent pas au sein d'une matrice hydratée, et n'adoptent pas une orientation préférentielle ainsi, il n'est pas nécessaire d'avoir recouru aux techniques de « tilts » (inclinaisons à différents angles).

Pratiquement, les particules sont observées sur des grilles recouvertes de carbone perforé sur lesquels est tendu un film mince contenant l'échantillon. L'épaisseur de ce film est de l'ordre de 20-100 nm et le film est obtenu en appliquant un papier absorbant sur une goutte de l'échantillon déposé sur la grille (par « blotting »), avant d'être plongé dans l'éthane liquide pour transformer le film d'eau liquide en glace vitreuse. Le support technique utilisé pour congeler l'échantillon est appelé la guillotine (Dubochet, Booy et al. 1981). Ce dispositif est composé d'une cuve contenant de l'éthane entouré d'une autre cuve remplie d'azote liquide (son point d'ébullition est de  $-189.85^{\circ}\text{C}$  et son point de fusion est de  $-195.85^{\circ}\text{C}$ ) qui sert à maintenir l'éthane dans un état liquide (son point d'ébullition est de  $-188.75^{\circ}\text{C}$  et son point de fusion est de  $-183.45^{\circ}\text{C}$ ). Quand le film fin de suspension protéique est formé, la guillotine est rapidement activée, la grille plonge brusquement dans l'éthane liquide ce qui entraîne une vitrification instantanée de la fine suspension protéique. A partir du moment où l'échantillon est congelé, le maintien de la chaîne du froid de la grille est primordial et préserver la grille d'une contamination de glace hexagonale par la condensation de l'humidité ambiante. Dans le microscope, le dépôt de molécules d'eau sur la grille refroidie est empêché par un piège froid placé autour de l'échantillon. Depuis l'invention de la guillotine, d'autres appareils ont été développés pour rendre plus reproductible et mieux contrôler les conditions de congélation des grilles. Le laboratoire est équipé d'un appareil VITROBOT qui permet de contrôler l'humidité de l'environnement dans lequel se fait le dépôt de la solution protéique sur la grille jusqu'à la congélation, ceci afin d'éviter l'évaporation de l'échantillon avant sa congélation. Cet appareil permet aussi de contrôler avec précision le temps d'application du papier absorbant sur la grille et ainsi de mieux contrôler l'épaisseur de la suspension protéique qui constitue un facteur déterminant pour le contraste lors de l'observation.

### *Observation de l'échantillon*

Le microscope utilisé est un TECNAI F20 (Philips™) équipé d'une platine cryogénique où se place l'échantillon, et ainsi permet la prise d'image à  $-174^{\circ}\text{C}$ . Une fois l'échantillon

préparé, il faut le transférer dans le vide du microscope, afin que seuls les atomes avec lesquels doivent interagir les électrons soient ceux de l'échantillon. La fragilité des échantillons sous le flux d'électrons oblige une faible exposition à l'irradiation pour limiter les dégradations induites par le transfert d'énergie des électrons au spécimen. L'acquisition des images s'effectue sur une émulsion photographique à un agrandissement de 50 000x dans la gamme de défocalisation de 0,5  $\mu\text{m}$  à 1,5  $\mu\text{m}$ . La numérisation du film argentique s'effectue à 5000 dpi (dot per inch) ce qui correspond à 1 Å/pixel.

### *L'analyse numérique des images*

#### Extraction et création d'une référence

Parmi tous les signaux présents dans une image, seule une fraction provient de l'objet. Le rapport signal/bruit des images de microscopie électronique est faible et ne permet pas l'interprétation directe des détails structuraux fins de l'objet observé. Les méthodes numériques d'analyse d'images ont pour but de séparer le signal contenant l'information structurale du bruit de nature aléatoire. Une image peut-être considérée comme étant constituée d'une partie signal et d'une partie bruit. Pour augmenter le rapport signal/bruit, les techniques d'analyse d'images visent à calculer la moyenne d'un nombre important de ces images bruitées. Cette opération permet d'éliminer le bruit réparti de façon aléatoire et d'améliorer le rapport signal/bruit d'un facteur N puisque le signal reste constant d'une image à l'autre (N étant le nombre d'images intervenant dans la moyenne). Ce calcul de la moyenne n'est justifié que si les images présentent toutes la même projection. La méthode de l'analyse par corrélation et moyenne d'image (van Heel and Frank 1981) est appliquée dans le cas général où aucune information d'orientation et de position des molécules n'est connue. Les images sont partitionnées en sous-groupe homogènes correspondant chacun à une projection de l'objet dans une orientation particulière.

L'image originale contient un grand nombre de différentes particules isolées. Une première sélection grossière et visuelle permet d'extraire des sous-images centrées sur les objets. Par sélection des zones d'intérêt, la quantité d'information à traiter sur l'image est ainsi réduite et la translation à appliquer pour amener les images en superposition est partiellement compensée. Le but de cette opération est de rendre les images comparables entre elles en éliminant les variations indépendantes de la structure de la macromolécule. Manuellement, les images qui se ressemblent le plus sont moyennées donnant ainsi une forme générale très

grossière de la macromolécule. Cette forme qui servira de première référence est nécessairement effectuée manuellement pour éviter tout biais dans le modèle. Le choix de la référence est très important puisqu'il détermine l'orientation des images recalées. Elle devra être représentative de la population d'images à comparer pour les recalculer de façon à permettre leur répartition en sous-groupes représentant la même orientation.

#### L'alignement en rotation et translation

L'orientation privilégiée selon laquelle les images individuelles vont être alignées est donnée par l'image de référence. Toutes les images de la population seront alignées sur cette référence en rotation d'abord, l'angle de rotation étant quelconque, en translation ensuite, la translation étant supposée faible. La méthode d'alignement consiste à calculer pour toutes les rotations et translation possibles, une mesure de similarité entre l'image et la référence. Cette mesure de similarité est donnée par le calcul du coefficient d'intercorrélation pour toutes les combinaisons de rotations et de translations. La recherche du maximum de la courbe d'intercorrélation permet de définir les paramètres d'alignement. Dans le cas de l'alignement en rotation, le calcul du coefficient d'intercorrélation se fait en coordonnées polaires pour se replacer dans le contexte d'un recalage en translation.

#### Classification des images

Si l'objet possède plusieurs orientations, le calcul direct de l'image moyenne n'améliore pas le rapport signal/bruit. Le critère de ressemblance (coefficient d'intercorrélation) garantit la similarité d'images comparées deux à deux mais pas celle de l'ensemble des données. L'intervention de méthodes statistiques automatiques va partitionner la population d'images en classe de projections selon des critères de ressemblance maximale. Cette classification se fait en deux étapes : l'analyse factorielle des correspondances et une classification hiérarchique ascendante.

#### *Analyse factorielle des correspondances*

Cet outil a été adapté à l'analyse d'images et permet de décrire une population d'images alignées dans un espace de dimensions restreintes. Dans un espace à N dimensions, une image constituée de N pixels est représentée par un point. La population d'images alignées est décrite par un nuage de points dans cet espace. La diminution du nombre de dimensions

s'effectue par le calcul des vecteurs propres décrivant la variance inter-images maximale des images de la population. Chaque image peut alors être décrite selon une combinaison linéaire de quelques vecteurs de base. Le principal avantage de cette étape est de permettre de diminuer les temps de calcul pour l'étape de classification. Elle permet également de n'utiliser que les tendances lourdes du jeu de données.

La référence pouvant influencer sur le résultat de la classification (van Heel, Schatz et al. 1992), les images moyennes obtenues à l'issue de ce cycle serviront aux cycles ultérieurs de classification, ainsi le biais du choix des références est interne au jeu de données.

#### *Classification hiérarchique ascendante*

Dans un premier temps, chaque image alignée est considérée comme une classe à part entière, avec ses propres coordonnées dans l'espace déterminées lors de l'analyse factorielle des correspondances. Selon un critère de proximité, les classes fusionnent alors deux à deux jusqu'à ne former plus qu'une seule et unique classe. L'historique de fusion des classes peut être représenté par un dendrogramme de structure dichotomique. Le critère de fusion correspond à la plus courte distance euclidienne entre les centres des deux classes, pondérée par le nombre de membres. Le partitionnement est affiné par une réaffectation des images aux classes dont elles sont le plus proche, pour de minimiser la variance à l'intérieur de chaque classe. Le nombre de classes est déterminé manuellement en coupant le dendrogramme à un niveau choisi. Le choix du partitionnement s'effectue selon les critères d'homogénéité: à l'intérieur d'une classe, la variation est la plus faible possible (les images sont le plus semblables possibles); et d'homologie: les classes éloignées correspondent à une vue différente de l'objet. L'inspection manuelle des images moyennes des classes de ces différentes partitions permet de regrouper les classes selon des critères de formes et de contours semblables pour déterminer les différentes familles. Cette opération permet de se concentrer sur les familles les plus stables. Pour chaque famille déterminée, une image moyenne est choisie et servira de nouvelle référence d'alignement.

#### *Alignement multi-références et itération*

Un alignement est de qualité si et seulement si la classification est efficace. Ainsi en bouclant le processus alignement - classification, la sélection des images moyennes de classes est améliorée. Le premier cycle d'alignement - classification permet de répartir des images

moyennes de classes dans différentes familles selon des caractéristiques morphologiques différentes. Afin de parfaire l'analyse, la boucle alignement - classification est réitérée. Les itérations successives stabilisent les familles d'images : une famille peu représentative va perdre de l'importance alors qu'une nouvelle famille peut apparaître, bien que masquée lors des cycles précédents. L'analyse par corrélation et moyenne d'images permet de calculer des images moyennes de l'objet à partir d'une grande population hétérogène d'images. Ces images moyennes correspondent à différentes : orientations, conformations ou espèces moléculaires. Plus la population d'images est complexe (nombre élevé de projections différentes), plus il faut d'images pour pouvoir distribuer ces images en classes homogènes. Par conséquent, une analyse d'un nombre important d'images est nécessaire pour mettre en évidence une sous-population faiblement représentée à l'intérieur d'un jeu de données. La stabilisation est dépendante de la diversité et de la qualité des références établies. Les images représentant les différentes familles sont sélectionnées, puis alignées entre elles en translation et en rotation selon un cycle itératif par comparaison des coefficients de corrélation. Les différentes projections (familles) de l'objet sont ainsi visualisées. Au final, la stabilisation des cycles d'alignement - classification est déterminée manuellement. Les dernières classes déterminées sont en partie différentes ou non de celles du cycle précédent.

### Reconstruction 3D

Pour reconstruire un objet en 3D à partir de projection 2D (images), la combinaison des différentes projections d'un même objet nécessite de déterminer les orientations relatives de ces projections entre elles. Afin de déterminer les relations angulaires entre les différentes projections, la technique de la reconstruction angulaire a été appliquée.

#### *La technique de reconstitution angulaire*

La technique de la reconstruction angulaire (Van Heel 1987) se base sur le principe que les particules sont distribuées aléatoirement sur le support de la grille de microscopie. Aucune inclinaison de grille n'est alors nécessaire dans la mesure où toutes les vues de la molécule sont présentes dans les images de départ. Cependant les orientations relatives des différentes projections sont inconnues. La technique de reconstitution angulaire permet de déterminer les angles des degrés de liberté en rotation à partir du théorème de la projection des lignes communes (« Common Line Projection (CLP) theory ») qui stipule que deux projections

différentes d'un même objet 3D partagent au moins une ligne de projection. A partir de ces lignes communes, la comparaison de projections deux à deux permet de déterminer les angles relatifs entre elles et donc de construire un modèle 3D.

Pour déterminer l'angle relatif de projections 2D, toutes les projections 1D sont préalablement calculées pour chaque image 2D. Un masque est appliqué à chaque image de manière à ce qu'essentiellement ne soit considérée que l'information provenant de la particule. Un sinogramme est représenté sous forme d'une image ayant pour lignes toutes les projections 1D de l'image 2D. Par le calcul des 360 projections 1D (incrément angulaire de  $1^\circ$ ), une projection 2D de 128x128 pixels permet de générer un sinogramme de 128 pixels de largeur sur 360 pixels de haut. Les sinogrammes sont normalisés afin de ne pas considérer des variations d'intensité qui ne sont pas dues à des différences d'orientation de l'objet. La comparaison deux à deux et ligne par ligne : permet d'obtenir une fonction de corrélation des sinogrammes (Sinogram Correlation Function, SCF). Le calcul de corrélation d'une ligne contre une autre est représenté numériquement selon un niveau de gris dans une image de la fonction de corrélation. Cette information est concentrée dans un pixel. La détermination de l'orientation se fait par observation des pics dans cette fonction. Une forte corrélation se traduit par un pic intense dont les coordonnées donnent l'angle relatif entre les deux projections. L'étude de la hauteur et la déviation standard des pics permet également d'éliminer les mauvaises projections.

#### *Rétroprojection pondérée*

Les angles ainsi déterminés pour un jeu de projections moyennes permettent ainsi un calcul de la densité par la méthode de la rétroprojection pondérée (Radermacher 1988). Cette technique est utilisée pour reconstruire effectivement le volume de l'objet. La rotation de l'échantillon doit se faire sur une gamme angulaire la plus large possible pour éviter de trop grandes distorsions entre la structure d'origine et sa reconstruction 3D, et il est aussi nécessaire d'avoir un repère central sur les images 2D pour permettre de centrer les images les uns par rapport aux autres pour effectuer la rétroprojection. Le principe consiste à replacer la moyenne des pixels de chaque projection dans le voxel correspondant, *i.e.* reconstruire dans un cube d'un pixel de face (voxel), chaque face est la moyenne des pixels pondérée en fonction de l'orientation des différentes projections.

### *Itération du modèle*

A partir du modèle tridimensionnelle, des projections 2D sont calculées pour créer un jeu de références représente toutes les projections 2D théoriques. Ce jeu de référence permet d'aligner derechef les images, et d'effectuer un nouveau cycle d'affinement. Les projections moyennes obtenues n'ont pas toutes forcément une origine 3D commune (centrées sur le centre de masse de la molécule). Grâce au calcul du premier modèle 3D, l'alignement sera amélioré en générant des projections rares passées inaperçues lors de la sélection des références et parfaire ainsi la reconstruction 3D par itération.

### *Fitting : Positionnement*

#### *Superposition*

Dans un premier temps, l'enveloppe 3D (IN/LEDGF/INI1(173-290)) obtenue est superposée à l'enveloppe 3D (IN/LEDGF) précédemment obtenue au laboratoire (Michel, Crucifix et al. 2009) en utilisant le programme CHIMERA (Pettersen, Goddard et al. 2004). Une première superposition grossière est faite manuellement suivit d'une superposition des deux cartes dans CHIMERA par une optimisation de la corrélation entre les deux cartes de densité. Dans un deuxième temps, les structures atomiques positionnées dans l'enveloppe (IN/LEDGF) (un tétramère d'intégrase, et le domaine de liaison à l'intégrase de LEDGF) sont transférées telles quelles dans l'enveloppe (IN/LEDGF/INI1(173-290)). Dans un troisième temps la position du modèle est affinée par corps rigide dans CHIMERA.

#### *Positionnement flexible en mode normal*

La dernière position du modèle est utilisée pour un positionnement flexible en utilisant les modes normaux avec le programme NORMA (Suhre, Navaza et al. 2006). Il a été montré que des changements importants de conformation peuvent être décrites par un nombre restreint de modes normaux basses fréquences (Yang and Bahar 2005). L'utilisation du positionnement flexible en modes normaux en utilisant des cartes de densité électronique basse résolution provenant de la microscopie électronique a formalisé par (Tama, Miyashita et al. 2004). Pour repositionner les structures atomiques dans la nouvelle enveloppe, une approche en plusieurs étapes a été utilisée. Un premier affinement a été réalisé en utilisant un seul mode basse fréquence (mode 7) et un premier modèle intermédiaire a été généré. Cette étape est

complétée par un affinement par REFMAC (CCP4-6.1 1994). Cette première étape est suivit par un affinement utilisant 3 modes (7,8,9), suivit par une régularisation de la structure par REFMAC.

## Résultats

### Cristallogénèse

Des tests de cristallisation ont été réalisés, cependant aucune condition testée n’a abouti à la formation de cristaux (**Erreur ! Source du renvoi introuvable.**). Les pré-tests de cristallisation (annexe 1) ont déterminé la concentration utilisée pour les tests de cristallisation. Le criblage a été réalisé en plaque Greiner™ LOWPROFILE™ par le robot de cristallisation (Cartesian©) en mélangeant 200 nL de conditions de cristallisation et 200 nL de protéine ou de complexe protéique.

Les différentes conditions de cristallisation

	concentration	JCSG +	Natrix/ MembFac	The Classics	PEG/Ion/pH	Sparse Matrix1	ProComplex	SaltRX	conditions de cristallisation IN(50-F185K- 212)
INI1(173-257)	4,2 mg/mL								
6His-MBP + INI1(173-257)	35 mg/mL								
INI1(173-257) + IN(50-F185K-212)	3,4 mg/mL								
6His-MBP + INI1(173-257) + IN(50-F185K-212)	25 mg/mL								

Figure 39 : Les différents panels de conditions (Hampton Research™) testés pour chaque protéine ou complexe protéique. Aucune condition testée n’a donné lieu à la formation de cristaux de protéines.

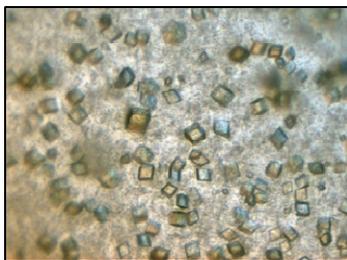
Les conditions de criblages couvrent les variations des paramètres habituellement utilisés pour la recherche des conditions de cristallisation. La meilleure façon d’identifier les conditions de cristallisation optimales consiste à explorer systématiquement les conditions (ou leurs variantes) les plus courantes pour obtenir des cristaux.

En cristallisation des protéines, la fonction générale des agents de précipitation est de diminuer la solubilité de la protéine. En théorie, les facteurs déclenchant entrent en compétition avec la protéine solutés pour les molécules d’eau, ce qui conduit à la

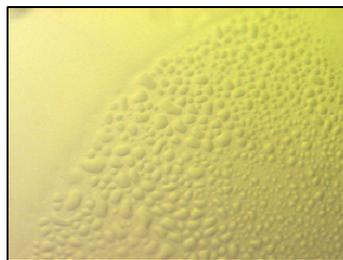
sursaturation des protéines. Les sels, les polymères, et les solvants organiques sont les facteurs déclenchant les plus populaires.

La stratégie « Sparse Matrix » (criblages « Sparse Matrix 1 », et « The Classics ») établit une carte typique d'un large spectre de conditions de cristallisation à l'aide de nombreux produits différents. Les conditions sont établies à partir des bases de données d'expériences de cristallisation ayant données des cristaux (criblage « JCSG+ »). Ces conditions constituent un excellent point de départ pour commencer la recherche de conditions de cristallisation pour le complexe.

De plus, la condition de cristallisation du domaine cœur catalytique de l'intégrase du VIH-1 mutant (F185K) a été testée. Il est intéressant de noter que seul l'intégrase (50-F185K-212) cristallise (Figure 40) Dans les mêmes conditions en présence d'INI1(173-257) avec ou sans l'étiquette HisMBP il n'y a pas cristallisation. La formation de sphérules en présence de HisMBP est probablement due à la formation d'agrégats solubles. Toutes les conditions approchant se sont révélées négatives pour la formation de cristaux.



**IN(50-F185K-212)  
5,2 mg/mL**



**6His-MBP +  
INI1(173-257) +  
IN(50-F185K-212)  
25 mg/mL**



**INI1(173-257) +  
IN(50-F185K-212)  
3,4 mg/mL**

Figure 40 : Condition de cristallisation de l'intégrase VIH-1 (50-F185K-212) (Dyda, Hickman et al. 1994). Cette condition a été testé sur l'IN(50-F185K-212) comme témoin positif, ainsi que sur 6His-MBP + INI1(173-257) + IN(50-F185K-212) et sur INI1(173-257) + IN(50-F185K-212).

## Microscopie électronique moléculaire

En collaboration avec le laboratoire de P. Schultz, un modèle en trois dimensions a été obtenu par microscopie électronique moléculaire. Plusieurs étapes ont été nécessaires à la réalisation des cartes de densité dans le but d'obtenir une enveloppe 3D du complexe IN/LEDGF/INI1(173-290). La comparaison avec le complexe IN/LEDGF permet de recalibrer la position d'INI1(173-290) au sein du complexe ternaire.

### *Coloration négative*

Afin de permettre de reconstruire un volume 3D du complexe IN/LEDGF/INI1(173-290), des micrographies ont été enregistrées sur des échantillons colorés négativement pour percevoir la qualité de l'échantillon. Ces micrographies sont sélectionnées en fonction de la qualité de la coloration, de l'astigmatisme de la micrographie et de l'intégrité structurale des particules qu'elles contiennent. Après extraction des images, l'analyse permet de calculer un volume pour donner un premier aperçu du complexe. En superposant le complexe IN/LEDGF et IN/LEDGF/INI1(173-290), il est possible de visualiser grossièrement les changements qu'opèrent l'ajout du domaine de liaison à l'intégrase d'INI1 (Figure 41). Le calcul d'une carte de différence entre IN/LEDGF et IN/LEDGF/INI1(173-290) permet de positionner INI1(173-290) au sein du complexe IN/LEDGF/INI1(173-290).

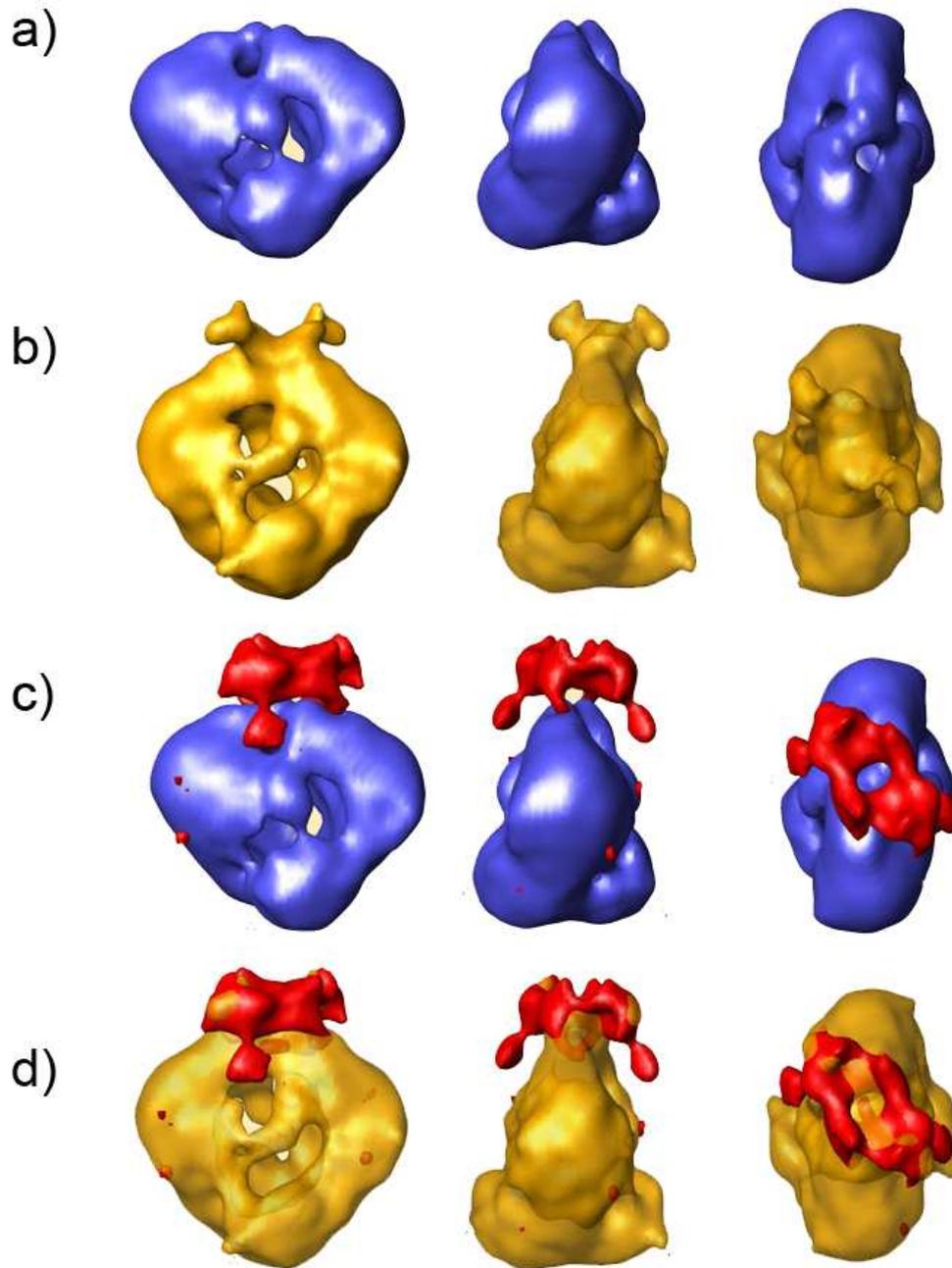


Figure 41: Ces cartes de microscopie électronique en coloration négative : en a) (bleu) trois vues du complexe IN/LEDGF sont représentées ; en b) (en or) est représenté le complexe IN/LEDGF/INI1(173-290). En c) Les deux complexes ont été superposés puis la différence entre IN/LEDGF et IN/LEDGF/INI1(173-290) a été représenté en rouge. INI1(173-290) apparaît en rouge. En d) la superposition de IN/LEDGF/INI1(173-290) et de la différence entre IN/LEDGF et IN/LEDGF/INI1(173-290) permet de visualiser le positionnement de INI1(173-290) au sein du complexe IN/LEDGF/INI1(173-290).

### *Cryo-Coloration négative*

Afin d'établir une référence des projections du complexe IN/LEDGF/INI1(173-290), des micrographies ont été enregistrées sur des échantillons cryo-colorés négativement. La reconstruction 3D à partir de ces images sera utilisée comme matrice pour l'alignement des images extraites des micrographies d'échantillons hydratés congelés. Après extraction des images, l'analyse permet de calculer un volume du complexe IN/LEDGF/INI1(173-290). Il est possible de superposer les cartes des complexes IN/LEDGF et IN/LEDGF/INI1(173-290) (Figure 42) Le coefficient de corrélation de la superposition des cartes est de 0,9.

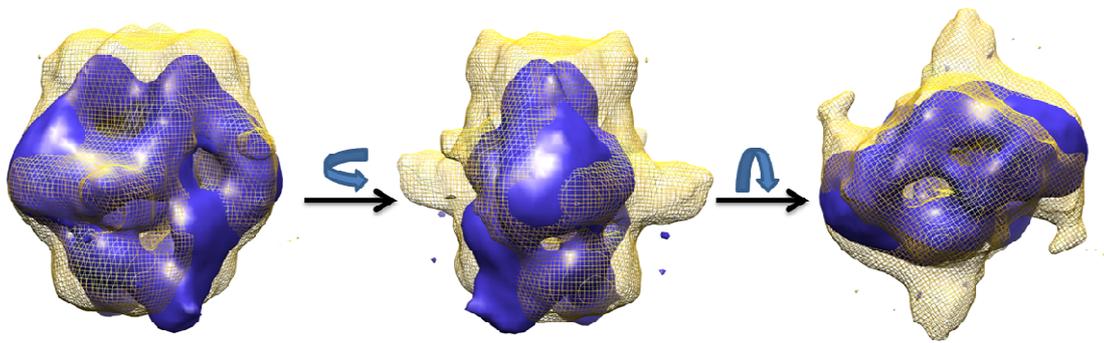


Figure 42: Cartes de microscopie électronique en cryo-coloration négative. la superposition des cartes du complexe IN/LEDGF (en bleu) et du complexe IN/LEDGF/INI1(173-290) (en or).

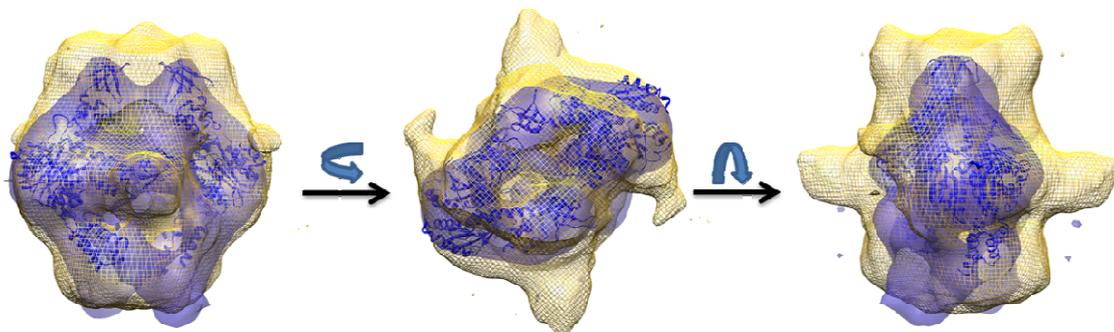


Figure 43: Le modèle en cartoon (bleu) positionné des domaines de l'intégrase et de l'IBD de LEDGF (Michel, Crucifix et al. 2009).

Le positionnement de la structure est ensuite affiné dans CHIMERA par corps rigide (Figure 43 a) et est ensuite affinée dans NORMA (Figure 43 b)

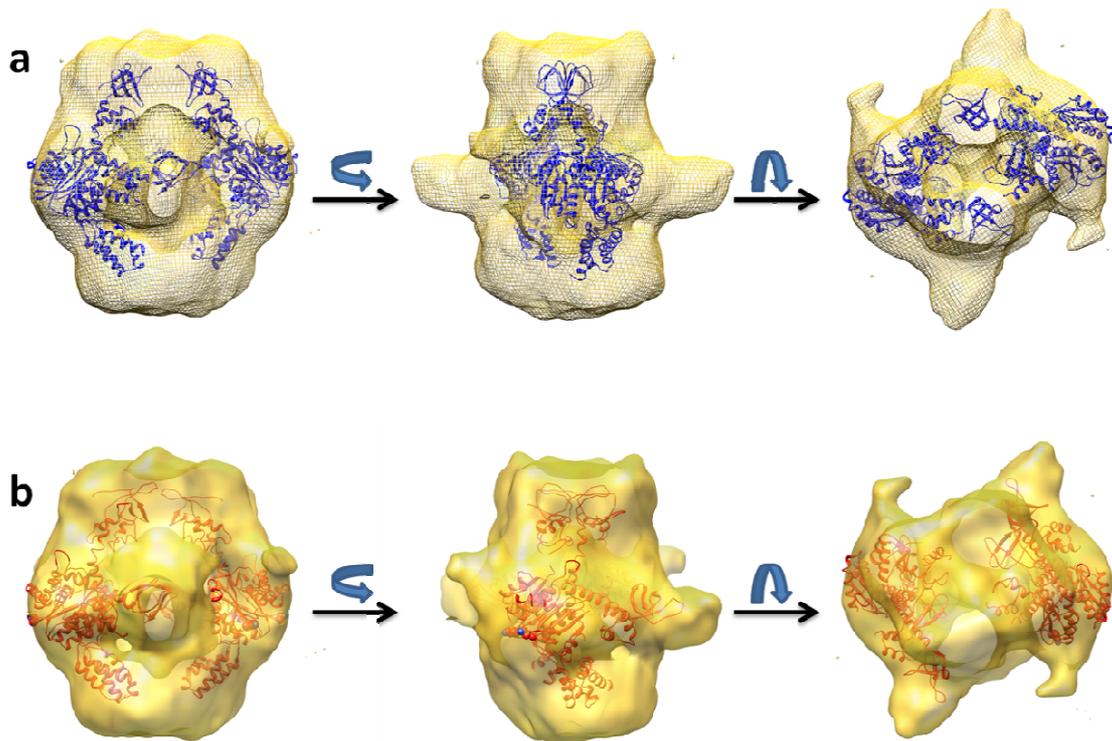


Figure 44: a) position initiale du complexe IN/LEDGF après positionnement dans la carte par corps rigides (Chimera) en bleu et b) position finale du complexe en jaune après positionnement flexible en modes normaux par NORMA.

Une carte différence entre la carte provenant de la microscopie électronique et le modèle final positionné a été générée par COOT (Emsley and Cowtan 2004) pour la partie supérieure de la carte de densité électronique. La densité résiduelle représente la position d'INI1 dans le complexe (Figure 45).

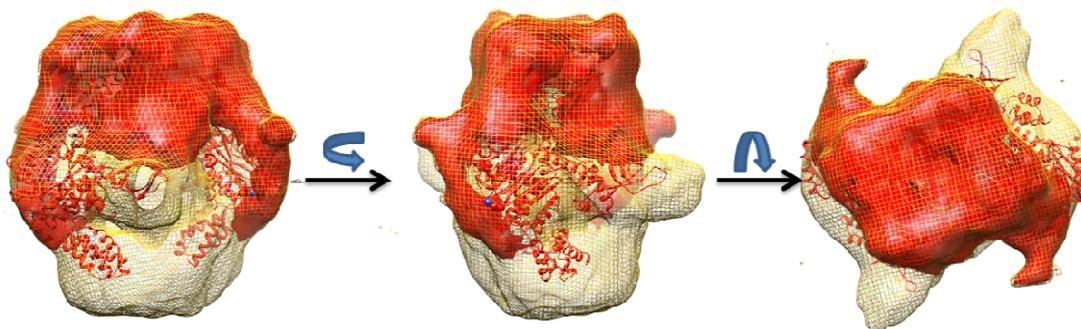


Figure 45: carte différence de la partie supérieure du modèle (la partie LEDGF a été retirée) montrant le positionnement de INI1 (en rouge).

Le volume occupé par le complexe IN/LEDGF/INI1 est en accord avec les expériences de spectrométrie de masse qui donnent une stœchiométrie de 4 intégrases, 2 LEDGF, et 2 INI1 (Figure 46).

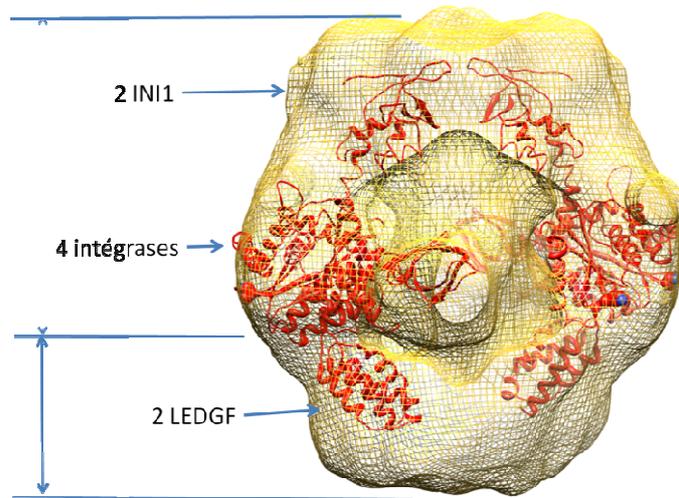


Figure 46: carte de densité (jaune) avec la structure de IN/LEDGF positionné (rouge) avec indiqué la stœchiométrie et la position de l'intégrase, de INI1 et de LEDGF

Les interactions de INI1 avec l'intégrase sont en accord avec la zone cartographié par double hybride et le résidu K71 crucial pour l'interaction (Maroun, Delelis et al. 2006) (Figure 47).

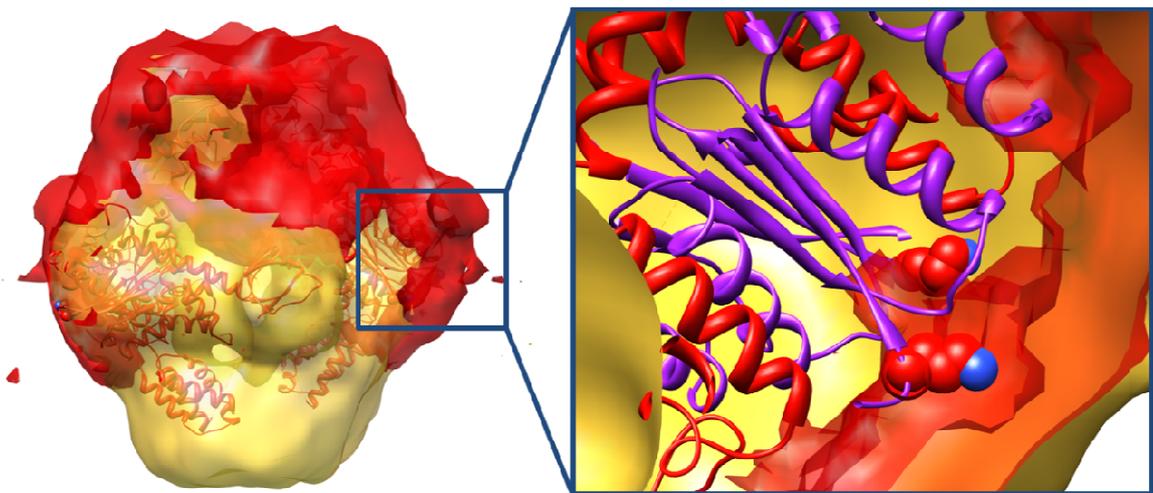


Figure 47: Zoom sur le corps catalytique de l'intégrase. Les résidus 53 -153 qui définissent le domaine d'interaction de l'intégrase avec INI1 sont en violet. Le résidu K71 crucial pour l'interaction est représenté en sphère.

On remarque que INI1 se positionne au sommet de l'intégrase en la maintenant dans une conformation qui diminuerait l'interaction avec l'ADN cible et empêcherait la réaction de 3' *processing* ce qui est accord avec les résultats fonctionnels.

## Discussions

### Discussion de l'étude *in silico*

#### Résumé des résultats de l'étude *in silico*

La stratégie *in silico* a permis de dégager des limites des domaines d'INI1. Les différentes prédictions obtenues à l'aide des différents programmes utilisés ont été combinées pour extraire de l'information pertinente. L'analyse des données ainsi produites a permis de définir les limites de domaines qui ont servi de base pour les clonages et tests d'expressions. Toutes les limites contiennent le domaine de liaison à l'intégrase d'INI1. Selon les limites utilisées, les différentes constructions imaginées englobent en plus soit le domaine potentiel de liaison à l'ADN (106-186), soit la répétition imparfaite (259-319), soit les deux.

#### Discussion

L'absence de structure analogue à INI1 englobant la totalité du domaine de liaison à l'intégrase a été un frein à l'établissement rapide de domaine structurellement intègre et soluble. En effet au sein du MACS aucune structure apparentée à INI1 n'a été mise en évidence. La compilation des différentes prédictions (structures secondaires, région intrinsèquement désordonnée, définition automatique des domaines) ont permis d'établir des limites. Parmi les constructions réalisées, un très petit nombre s'est avéré soluble. Cependant ces constructions ont permis la formation d'un complexe ternaire stable. A posteriori, la prédiction P173 la plus forte du côté N-ter était effectivement la meilleure de celles testées. En effet, la convergence des prédictions de structure secondaire était un indice très fort. Malheureusement, c'est l'unique convergence de prédictions de structure secondaire dans la séquence.

#### Perspectives

Certaines limites de domaines 85 et 106 n'ont pas été exploitées. La limite 85 exclurait totalement la partie manquante de l'isoforme B. La construction commençant par 106 permettrait d'englober au plus près le domaine potentiel de liaison à l'ADN. La construction tirée (183-243) d'études précédentes (Morozov, Yung et al. 1998) n'est pas soluble dans les conditions testées. Différentes limites (141, 145 et 294) peuvent générer des fragments intéressants pour une étude structurale. Cependant, la solubilité reste un problème récurrent.

En effet, sortis du contexte cellulaire, ces fragments seront probablement insolubles au vue des prédictions *in silico*. Un changement de système de production (baculovirus, mammifère) pourrait permettre d'obtenir des fragments ayant subi des modifications post-traductionnelles.

Une modélisation par homologie de la structure 3D du domaine 173-290 est en cours. Cependant, les homologies entre les séquences est inférieure à 20 % et les structures prédites comme homologue ne convergent pas. De plus, aucune structure ne recouvre entièrement le domaine.

## Discussion tests d'expression et production

### Résumé des résultats des tests d'expression

La construction de la banque de clones des fragments d'INI1 a été réalisée par la stratégie Gateway©. Les tests d'expression ont permis de mettre en évidence les constructions solubles des différents vecteurs, notamment les constructions 173-257 et 173-290.

### Discussion

Les tests d'expression ont mis en évidence les constructions les plus solubles. Cependant, plusieurs faux positifs ont été détectés à posteriori. La difficulté réside dans le changement d'échelle entre les 2 mL de milieu des tests d'expression et de solubilité, et 1 L de milieu utilisé en production de masse. Ce changement d'échelle induit une variabilité qui nécessite des ajustements. Les conditions de production étant différente, la variabilité peut entre autre diminuer la solubilité de la construction. Dans ce cas, de nouvelles conditions de solubilité doivent être trouvées.

### Perspectives

Les différentes tentatives de co-expression dans la souche bactérienne *E.coli BL21(DE3)* du complexe IN-LEDGF à partir de plasmide bicaténaire se sont avérées peu prometteuses. *A fortiori*, la co-expression de trois partenaires n'était pas un choix tactique judicieux. Cependant, la formation du complexe entre les fragments INI1 et IN(50-F185K-212) avec et sans Nter et Cter pourra être testée en co-expression.

## Discussion purification

### Résumé des résultats de purification

La purification du complexe à partir des partenaires exprimées séparément s'est révélé l'approche la plus robuste en termes de reproductibilité. Le rendement pour le complexe IN/LEDGF est de 6 mg de complexe purifié pour 6 L de culture d'IN (22,4 g de cellules) et 12 L de culture de LEDGF (45,6 g de cellules), Le rendement pour le complexe IN/LEDGF/INI1(173-290) est de 1,5 mg pour 6 L de culture d'IN (22,4 g de cellules), 12 L de culture de LEDGF (45,6 g de cellules) et 3 L de culture d'INI1(173-290) (3,2 g de cellules).

### Discussion

La purification du complexe ternaire dure en général trois semaines. Il n'a jamais été possible de réduire le temps de purification, par exemple en réduisant le temps de dialyse et/ou en augmentant la vitesse de protéolyse ménagé, sans affecter l'intégrité et la qualité du complexe. La purification de ce complexe nécessite un véritable savoir-faire empirique développé au sein du laboratoire.

La purification par co-broyage est une méthode beaucoup plus rapide pour la formation du complexe IN/LEDGF et de même IN/LEDGF/INI1(173-290). La variabilité des différentes productions des différentes protéines rend difficile la reproductibilité. En effet, cette variabilité au sein des productions nécessite une phase de mise au point importante pour chaque nouvelle production et ce pour chaque protéine. Une des solutions pourrait être de produire une grande quantité de chaque protéine (en fermenteur de 100 L), et d'optimiser le ratio des masses de cellules de chaque production de protéines. Le temps pour optimiser la formation du complexe risque de faire perdre un temps considérable d'autant qu'il faudra optimiser chaque production.

Pour les fragments d'INI1, la principale difficulté rencontrée au cours de l'étape de purification était l'instabilité du fragment INI1(173-257). En effet, la protéine s'agrégeait au fur et à mesure de l'avancement de la purification. L'utilisation de l'étiquette de solubilité MBP a masqué la solubilité réelle d'INI1(173-257) en formant des agrégats solubles (Nominé, Ristriani et al. 2001). La construction 6His-MPB-INI1(173-257) est apparemment soluble, cependant en réalité le fragment INI1(173-257) est maintenu en solution par la

multimérisation de 6His-MBP qui forme une cage de solvation et au centre le fragment INI1(173-257) agrégé.

Indéniablement, le zwittergent 3-12 est un détergent qui solubilise très bien les fragments d'INI1 notamment INI1(173-290). Cependant, le zwittergent 3-12 inhibe les interactions entre protéines (Nollstadt, Karkhanis et al. 1989), de plus une fraction résiduelle reste même après dialyse. En revanche, le CHAPS est un détergent électriquement neutre qui empêche l'agrégation mais non dénaturant (Hjelmeland 1980). De plus le CHAPS est plus facilement dialysable.

## Discussion Fonction

### Liaison à l'ADN

#### Résumé

La comparaison entre les complexes IN/LEDGF et IN/LEDGF/INI1(173-290) font apparaître l'action du domaine de liaison à l'intégrase d'INI1. Cette étude a montrée qu'en présence d'INI1(173-290) l'affinité pour l'ADN cible est diminuée de 4 fois et de 2 fois pour l'ADN viral pour chacun des deux sites. L'étude a montré une spécificité pour les sites de liaison à l'ADN viral de l'intégrase.

#### Discussion

Pour les mesures sur le complexe IN/LEDGF/INI1(173-290) en présence d'ADN froid, les courbes décrivent une légère sigmoïde ce qui pourrait trahir une perte de sensibilité de l'appareillage ou une coopérativité entre les sites de liaison à l'ADN. Dans ce cas, le modèle utilisé ne serait plus satisfaisant car il implique que les sites de liaison soient identiques. L'étude de la variation du ratio ADN/complexe pourrait établir s'il y a ou non coopérativité.

Une étude de la liaison à l'ADN (cible et viral) des complexes IN/LEDGF et IN/LEDGF/INI1(173-290) en fonction de la concentration en NaCl permettra déceler les interactions non ionique entre l'ADN et les protéines (Mély, de Rocquigny et al. 1995). Les effets des ions  $Mg^{2+}$ ,  $Mn^{2+}$  et  $Zn^{2+}$  sur les complexes devrons être aussi contrôlés.

## *Perspectives*

Une étude basée sur la technique de spectroscopie de corrélation de fluorescence est en cours pour confirmer le nombre de sites de liaison et leur spécificité. Cette technique est une alternative à l'anisotropie résolue en temps. En effet, le temps de vie de la fluorescéine (4 ns) est trop court au regard de la taille imposante des complexes étudiés (250 KDa à 300 KDa).

## **3'processing**

### *Résumé des résultats*

Au sein du complexe IN/LEDGF, le fragment INI1 (173-290) empêche complètement le 3' *processing*. Le domaine de liaison à l'intégrase d'INI1 (173-290) non seulement réduit l'affinité de l'intégrase pour l'ADN, mais inhibe également la maturation endonucléolytique de l'intégrase.

### *Discussion*

La perte de 40 % du complexe par dégradation est probablement due à combinaison de la faible force ionique, de la température (37 °C), et de la durée de la mesure (6 heures). Il est intéressant de noter que le fragment INI1(173-290) n'influence pas significativement la dégradation du complexe IN/LEDGF.

La comparaison avec d'autres études (Guiot 2006; Smolov, Gottikh et al. 2006; Merkel, Andrade et al. 2009; Michel, Crucifix et al. 2009) montre que l'activité du complexe IN/LEDGF est supérieure à celle de l'intégrase seule. La tétramérisation de l'intégrase induite par la formation du complexe avec LEDGF semble permettre une conformation facilitant le 3' *processing*. Quand l'intégrase est seule, les études montrant qu'un dimère d'intégrase est plus efficace qu'un tétramère d'intégrase seule (Guiot 2006). Le changement de conformation qu'opère LEDGF est probablement à l'origine de l'augmentation de l'activité. Maintenant que le test d'activité est mise au point en microplaque, il serait intéressant d'observer la variation du ratio ADN/complexe, ainsi que l'effet des différents cations divalents  $Mg^{2+}$ ,  $Mn^{2+}$  et  $Zn^{2+}$  sur l'activité du complexe IN/LEDGF/INI1(173-290).

Lorsque l'intégrase du VIH-1 est sous forme tétramérique, il existe une activité de clivage interne de l'ADN sur des séquences palindromiques (Delelis, Parissi et al. 2007). Une étude

comparative (avec et sans INI1(173-290) pourrait mettre en lumière le rôle d'INI1 dans le choix du site d'intégration. En effet, il a été suggéré depuis longtemps qu'INI1 pourrait intervenir dans le choix du site d'intégration

### Les autres activités de l'intégrase

Pour que l'étude soit la plus complète possible, plusieurs autres activités enzymatiques de l'intégrase doivent être testées sur le complexe avec et sans INI1(173-290).

### Transfert de brins

L'activité de transfert de brins du complexe pourrait être testée par transfert d'énergie entre molécules fluorescentes résolu en temps (Wang, Klock et al. 2005) (Figure 48). Cette méthode présente l'avantage de pouvoir mesurer l'activité enzymatique en continu. De plus, du fait des temps de vie très long des composés fluorescents à base d'euprium (de l'ordre plusieurs centaines de  $\mu$ s) les oligonucléotides utilisés pourraient également servir pour une étude de l'état d'oligomérisation des complexes IN/LEDGF par des mesures d'anisotropie résolue en temps.

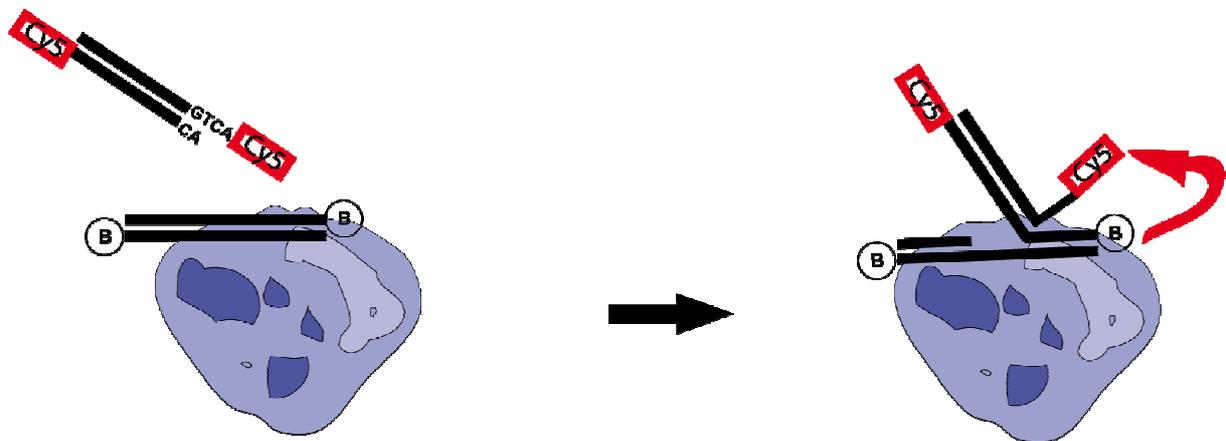


Figure 48: Principe du test pour le transfert de brins basé sur le TR-FRET. Le complexe catalyse le transfert de l'ADN viral « préprocessé » marqué au Cy5 (accepteur). L'ADN cible est marqué à l'Europium-chélaté (B) (donneur).

Une étude a montrée que LEDGF augmente l'activité de transfert de brin (Cherepanov, Maertens et al. 2003). La protéine pleine longueur (Kalpana, Marmon et al. 1994) et le fragment (183-294) d'INI1 (Das, Cano et al. 2009) stimule également l'activité de transfert de brin de l'intégrase. Il sera intéressant d'observer le rôle d'INI1 et du fragment (173-290) au

sein du complexe IN/LEDGF dans le cadre des deux réactions de trans-estérification (*3' processing* et transfert de brins). En effet, les tests d'activité de *3' processing* ont montrés une inhibition totale de l'activité en présence d'INI1 (173-290). Une explication serait qu'en l'absence d'ADN cible, INI1 inhiberait la réaction de *3' processing* afin d'éviter que l'ADN viral ne s'intègre lui-même.

### *Intégration concertée*

Enfin, pour être au plus proche de ce qui se passe dans la cellule, une étude de l'intégration concertée pourrait être envisagée (Grandgenett, Bera et al. 2009). La présence de partenaire protéiques viraux (NC) ou cellulaires de l'intégrase augmente l'efficacité (Carteau, Gorelick et al. 1999). Cependant, l'intégration concertée est difficile à mettre en évidence *in vitro* (Sinha and Grandgenett 2005). Une étude a montrée que LEDGF augmente modestement l'intégration concertée (Pandey, Sinha et al. 2007). Dans la cellule, un effet dose dépendant à LEDGF est mesuré.

### *Désintégration*

Le test de désintégration peut facilement être mis en œuvre selon une méthode similaire (Chow 1997; Merkel, Andrade et al. 2009) par anisotropie de fluorescence avec des oligonucléotides en Y ou en haltère (*dumbbell*) (Figure 49). Généralement, ce test est utilisé pour tester le mécanisme compétitif de certains inhibiteurs car le domaine central isolé est requis pour la désintégration. Il serait intéressant de savoir le mode d'action d'INI1 (173-290) sur cette réaction qui s'apparente à une activité endonucléasique non spécifique.

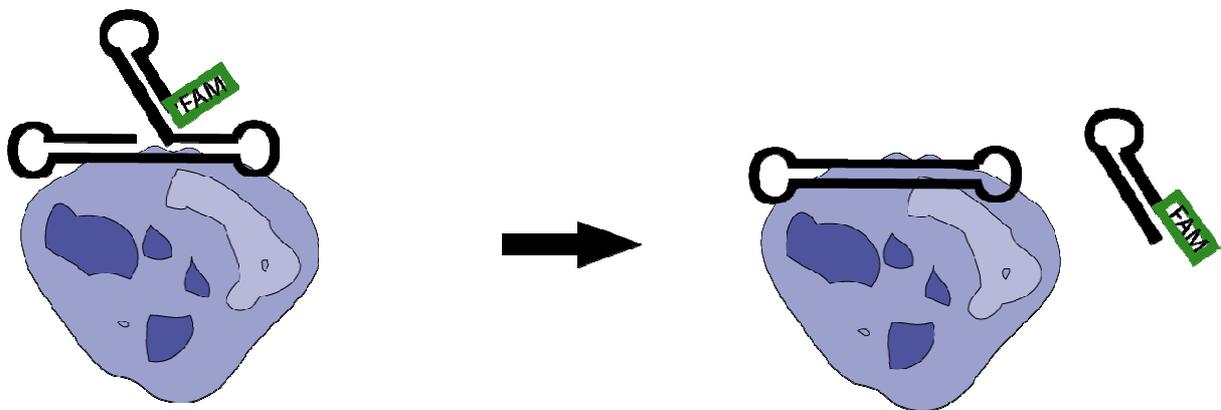


Figure 49: Principe du test de désintégration. En violet est représenté le complexe avec l'IN. L'ADN est marqué à la fluorescéine. La réaction est suivie par mesure de l'anisotropie de fluorescence. S'il y a désintégration, l'anisotropie devrait diminuer.

## Discussion structure

La structure présentée dans ce travail a été résolue par microscopie électronique par la technique de cryo-coloration négative. Elle montre l'organisation supramoléculaire du complexe IN<sub>4</sub>/LEDGF<sub>2</sub>/INI1<sub>2</sub>. INI1 se place dans le complexe à l'opposé de LEDGF dans la zone de d'interaction de l'ADN cible et bloque également le site actif pour la réaction de 3' *processing*. Les interactions de INI1 avec l'intégrase sont en accord avec la zone cartographié par double hybride et le résidu K71 crucial pour l'interaction (Maroun, Delelis et al. 2006).

Les résultats publiés sont contradictoires concernant le rôle que joue INI1 pendant l'infection du VIH-1. Une étude (Boese, Sommer et al. 2004) a suggéré qu'INI1 n'avait aucun effet sur l'intégration du virus dans des cellules déplétives en INI1. Tandis que d'autres rapports ont montrés qu'INI1 participe à une réponse antivirale cellulaire en interférant avec les étapes précoces de la réplication virale (Turelli, Doucas et al. 2001; Maroun, Delelis et al. 2006). A l'inverse, une autre étude (Ariumi, Serhan et al. 2006) a montrée qu'INI1 renforce indirectement la transcription provirale. INI1 a également été jugée essentielle pour la production de particules virales infectieuses (Yung, Sorin et al. 2001; Sorin, Yung et al. 2006).

Nos résultats structuraux et fonctionnels suggèrent que INI1, en bloquant le site actif de l'intégrase, en inhibant la réaction de 3' *processing* et en diminuant l'affinité de l'intégrase pour l'ADN cible, pourrait empêcher l'auto-intégration de l'ADN viral sur lui-même dans le noyau cellulaire avant de cibler l'ADN cellulaire. Ce qui serait cohérent avec les résultats publiés montrant un rôle apparemment contradictoire de INI1. En effet une surproduction de INI1 dans la cellule infecté empêcherait l'intégration alors qu'une cellule déplétive en INI1 montrerait des taux de réplication virale faible du à une diminution du taux d'intégration de l'ADN viral dans l'ADN cellulaire, conséquence de l'auto-intégration de l'ADN viral sur lui-même.

En perspective les structures de ce complexe avec et sans ADN viral avec des échantillons hydratés congelés sont en cours de résolution. La structure avec l'ADN viral nous donnera la position et les zones d'interaction de l'ADN viral avec le complexe IN/LEDGF/SNF5 et nous permettra de comprendre structurellement l'inhibition de la réaction de 3' *processing*.

## Annexes

### Composition des milieux

Les milieux sont stérilisés par autoclavage deux heures à 110 °C en atmosphère liquide et sous pression d'une atmosphère.

#### Composition de milieu LB

<i>composition</i>	<i>concentration</i>	<i>unité</i>
bacto-tryptone	1	% m/v
extrait de levure	0,5	% m/v
NaCl	0,5	% m/v

#### Composition de milieu LB agar

<i>composition</i>	<i>concentration</i>	<i>unité</i>
bacto-tryptone	1	% m/v
extrait de levure	0,5	% m/v
NaCl	0,5	% m/v
bacto-agar	1,5	% m/v

#### Composition de milieu SOB (pour bactéries chimiocompétentes).

<i>composition</i>	<i>concentration</i>	<i>unité</i>
bacto-tryptone	2	% m/v
extrait de levure	0,5	% m/v
NaCl	10	mM
KCl	2,5	mM
MgCl <sub>2</sub>	10	mM
MgSO <sub>4</sub>	10	mM
NaOH	ajuster le pH 7,0	

Le milieu est autoclavé avant l'ajout des sels. Ces derniers sont ajoutés stérilement.

### Composition de milieu TB (pour bactéries chimiocompétentes).

<i>composition</i>	<i>concentration</i>	<i>unité</i>
PIPES pH7,0	10	mM
CaCl <sub>2</sub>	15	mM
MnCl <sub>2</sub>	55	mM
KCl	250	mM
KOH	ajuster le pH 7,0	

Le pH est ajusté à 7,0 avant d'ajouter le MnCl<sub>2</sub>. le milieu est filtré sur membrane de porosité 0,22 µm.

### Composition de milieu ZYP-5052 (pour tests mini-expression).

<i>composition</i>	<i>concentration</i>	<i>unité</i>
milieu ZY	92	% v/v
MgSO <sub>4</sub>	1	mM
50x5052	2	% v/v
20xNPS	5	% v/v

Ce mélange s'effectue stérilement extemporanément.

#### *Détail :*

Milieu ZY

<i>composition</i>	<i>concentration</i>	<i>unité</i>
bactotryptone	1	% m/v
yeast extract	0,5	% m/v

Le milieu est autoclavé

50x5052

<i>composition</i>	<i>concentration</i>	<i>unité</i>
glycerol	25	%
glucose	2,5	%
alpha-lactose	10	%

La solution est filtré sur membrane de porosité 0,22 µm.

20xNPS

<i>composition</i>	<i>concentration</i>	<i>unité</i>
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	0,5	M
KH <sub>2</sub> PO <sub>4</sub>	1	M
Na <sub>2</sub> HPO <sub>4</sub>	1	M
	ajuster le pH 6.75	

La solution est filtré sur membrane de porosité 0,22 µm.

## Electrophorèse sur gel de polyacrylamide en conditions dénaturantes (SDS-PAGE)

### Composition du gel de séparation

<i>% de réticulation</i>	10%	12,50%	15%	
<i>composition</i>	<i>volume</i>	<i>volume</i>	<i>volume</i>	<i>unité</i>
H <sub>2</sub> O distillée	6,1	5,4	4,1	mL
TRIS-HCL 1,5mM pH8.8 SDS 0,4 % m/v	3,15	3,15	3,15	mL
acrylamide-bisacrylamide 29/1 à 40 %	3,1	3,9	4,7	mL
persulfate d'ammumium 10 %	125	125	125	µL
TEMED	12	12	12	µL

### Composition du gel de concentration

<i>composition</i>	<i>volume</i>	<i>unité</i>
H <sub>2</sub> O distillée	3,8	mL
TRIS-HCL 0,5mM pH6.8 SDS 0,4 % m/v	1,6	mL
acrylamide-bisacrylamide 29/1 à 40 %	0,8	mL
persulfate d'ammumium 10 %	125	µL
TEMED	12	µL

### Composition du tampon de dénaturation

<i>composition</i>	<i>concentration</i>	<i>unité</i>
TRIS-HCL pH6.8	100	mM
SDS	4	% v/v
glycérol	25	% v/v
bleu de bromophénol	0,2	% m/v
β-mercaptoéthanol	140	mM

### Composition du tampon de migration 5x concentré

<i>composition</i>	<i>quantité pour 1L</i>	<i>unité</i>
TRIS-base	15	g
glycine	72	g
SDS	0,5	% v/v
H2O distillée qsp	1	L

### Composition de la solution de coloration

<i>composition</i>	<i>quantité pour 200 mL</i>	<i>unité</i>
éthanol	30	% v/v
acide acétique	20	% v/v
bleu de Coomassie G250	2	g
H2O distillée qsp	200	mL

### Composition de la solution de décoloration

<i>composition</i>	<i>concentration</i>	<i>unité</i>
éthanol	30	% v/v
acide acétique	20	% v/v
H2O distillée	qsp	

### Composition de la solution de conservation

<i>composition</i>	<i>concentration</i>	<i>unité</i>
acide acétique	2	% v/v
H2O distillée	qsp	

### Protocole

Le gel est moulé entre deux plaques de verres prévues à cet effet. L'épaisseur du gel est d'environ 1 mm. Avec une pipette, on dépose la solution encore liquide du gel de séparation, puis on la recouvre d'éthanol absolu pour aplanir l'interface. Une fois le gel de séparation polymérisé, l'éthanol est enlevé, et on dépose de la même façon la solution du gel de concentration.

La dénaturation des 10  $\mu$ L d'échantillons se fait en ajoutant 5  $\mu$ L de tampon de dénaturation. La dénaturation des protéines est complétée par chauffage 5 minutes à 100 °C. La migration s'effectue à une tension constante de 250 V et une intensité limitant de 80 mA.

Le gel est démoulé de son support, mis à fixer et colorer par immersion dans une solution de coloration au bleu de Coomassie pendant 10 min à température ambiante et sous agitation. Le gel est ensuite décoloré dans plusieurs bains de solution de décoloration pour révéler les protéines précédemment fixées et colorées. Après rinçage à l'eau distillée, le gel est immergé dans la solution de conservation.

### Hybridation des oligonucléotides

Prendre chaque oligonucléotide dans du tampon (NaCl 50 mM, BisTris pH 6.5 10 mM) afin d'obtenir une concentration finale en oligonucléotide de 200  $\mu$ M. Puis mélanger les oligonucléotides complémentaires dans un ratio 1/1, typiquement 15  $\mu$ L / 15  $\mu$ L. Dans un appareil PCR, l'étape de dénaturation dure 10 minutes à 90 °C, puis l'hybridation consiste en une diminution de la température de 0,5 °C toutes les 5 minutes sur 14 heures 20 minutes pour atteindre 4 °C. La vérification de l'hybridation s'effectue par TBE-PAGE 20 % (Tris Borate EDTA PolyAcrylamide Gel Electrophoresis). 2 nmoles de chaque échantillon sont chargés supplémenté par du bleu de charge (Fermentas R0611). L'électrophorèse s'effectue en conditions non dénaturante à 45 V pendant environ 27 heures. Le gel est démoulé, coloré (5 min maximum) au bleu de toluidine, puis rincé à l'eau.

Remarque : en absence d'appareil PCR disponible, dans un becher, on maintient de l'eau à 90 °C, on y plonge les tubes à hybrider pendant 10 minutes, puis on transfère le tout (becher + tubes) dans une boîte en polystyrène hermétique que l'on dépose à 4 °C durant la nuit.

## Bibliographie

- Aida, Y. and G. Matsuda (2009). "Role of Vpr in HIV-1 nuclear import: therapeutic implications." Curr.HIV.Res. **7**(2): 136-143.
- Al-Mawsawi, L. Q. and N. Neamati (2007). "Blocking interactions between HIV-1 integrase and cellular cofactors: an emerging anti-retroviral strategy." Trends in Pharmacological Sciences **28**(10): 526-535.
- Anthony, N. J. (2004). "HIV-1 integrase: a target for new AIDS chemotherapeutics." Curr.Top.Med.Chem. **4**(9): 979-990.
- Ariumi, Y., F. Serhan, et al. (2006). "The integrase interactor 1 (INI1) proteins facilitate Tat-mediated human immunodeficiency virus type 1 transcription." Retrovirology **3**.
- Arthur, L. O., J. W. Bess Jr, et al. (1992). "Cellular proteins bound to immunodeficiency viruses: Implications for pathogenesis and vaccines." Science **258**(5090): 1935-1938.
- Baker, T. S. and J. E. Johnson (1996). "Low resolution meets high: towards a resolution continuum from cells to atoms." Current Opinion in Structural Biology **6**(5): 585-594.
- Barré-Sinoussi, F., J. C. Chermann, et al. (1983). "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)." Science (New York, N.Y.) **220**(4599): 868-871.
- Bernard, P. and M. Couturier (1992). "Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes." Journal of Molecular Biology **226**(3): 735-745.
- Boese, A., P. Sommer, et al. (2004). "Ini1/hSNF5 is dispensable for retrovirus-induced cytoplasmic accumulation of PML and does not interfere with integration." FEBS Letters **578**(3): 291-296.
- Bouyac-Bertoia, M., J. D. Dvorin, et al. (2001). "HIV-1 infection requires a functional integrase NLS." Mol.Cell. **7**(5): 1025-1035.
- Bowerman, B., P. O. Brown, et al. (1989). "A nucleoprotein complex mediates the integration of retroviral DNA." Genes & development **3**(4): 469-478.
- Brenner, S. and R. W. Horne (1959). "A negative staining method for high resolution electron microscopy of viruses." Biochimica Et Biophysica Acta **34**: 103-110.
- Bru, C., E. Courcelle, et al. (2005). "The ProDom database of protein domain families: more emphasis on 3D." Nucl. Acids Res. **33**(suppl\_1): D212-215-D212-215.
- Bujacz, G., J. Alexandratos, et al. (1996). "The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant." FEBS Lett. **398**(2-3): 175-178.
- Bujacz, G., M. Jaskolski, et al. (1995). "High-resolution structure of the catalytic domain of avian sarcoma virus integrase." J.Mol.Biol. **253**(2): 333-346.
- Bujacz, G., M. Jaskolski, et al. (1996). "The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations." Structure. **4**(1): 89-96.
- Bukovsky, A. and H. Gottlinger (1996). "Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease." J.Virol. **70**(10): 6820-6825.
- Bukrinsky, M. I. and O. K. Haffar (1997). "HIV-1 nuclear import: in search of a leader." Front Biosci. **2**:d578-87.: d578-d587.
- Burke, C. J., G. Sanyal, et al. (1992). "Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase." J.Biol.Chem. **267**(14): 9639-9644.
- Busschots, K., J. Vercaemmen, et al. (2005). "The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding." Journal of Biological Chemistry **280**(18): 17841-17847.
- Busschots, K., A. Voet, et al. (2007). "Identification of the LEDGF/p75 binding site in HIV-1 integrase." J.Mol.Biol. **365**(5): 1480-1492.

- Carlson, M. and B. C. Laurent (1994). "The SNF/SWI family of global transcriptional activators." Current Opinion in Cell Biology **6**(3): 396-402.
- Carteau, S., S. C. Batson, et al. (1997). "Human immunodeficiency virus type 1 nucleocapsid protein specifically stimulates Mg<sup>2+</sup>-dependent DNA integration in vitro." J.Virol. **71**(8): 6225-6229.
- Carteau, S., R. J. Gorelick, et al. (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**(8): 6670-6679.
- CCP4-6.1 (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallographica. Section D, Biological Crystallography **50**(Pt 5): 760-763.
- Cereseto, A., L. Manganaro, et al. (2005). "Acetylation of HIV-1 integrase by p300 regulates viral integration." EMBO J. **24**(17): 3070-3081.
- Chen, A., I. T. Weber, et al. (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**(7): 4173-4182.
- Chen, J. C., J. Krucinski, et al. (2000). "Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding." Proc.Natl.Acad.Sci.U.S.A. **97**(15): 8233-8238.
- Cheng, S. W., K. P. Davies, et al. (1999). "c-MYC interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function." Nat Genet **22**: 102-105.
- Cherepanov, P., A. L. Ambrosio, et al. (2005). "Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75." Proc.Natl.Acad.Sci.U.S.A. **102**(48): 17308-17313.
- Cherepanov, P., E. Devroe, et al. (2004). "Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase." Journal of Biological Chemistry **279**(47): 48883-48892.
- Cherepanov, P., E. Devroe, et al. (2004). "Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase." J.Biol.Chem. **279**(47): 48883-48892.
- Cherepanov, P., G. Maertens, et al. (2003). "HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells." Journal of Biological Chemistry **278**(1): 372-381.
- Cherepanov, P., G. Maertens, et al. (2003). "HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells." J.Biol.Chem. **278**(1): 372-381.
- Cherepanov, P., Z. Y. Sun, et al. (2005). "Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75." Nat.Struct.Mol.Biol. **12**(6): 526-532.
- Chow, S. A. (1997). "In Vitro Assays for Activities of Retroviral Integrase." Methods **12**(4): 306-317.
- Chow, S. A., K. A. Vincent, et al. (1992). "Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus." Science. **255**(5045): 723-726.
- Ciuffi, A., M. Llano, et al. (2005). "A role for LEDGF/p75 in targeting HIV DNA integration." Nat.Med. **11**(12): 1287-1289.
- Clapham, P. R. and R. A. Weiss (1997). "Spoilt for choice of co-receptors." Nature **388**(6639): 230-231.
- Corpet, F., F. Servant, et al. (2000). "ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons." Nucl. Acids Res. **28**(1): 267-269.
- Craig, E., Z. K. Zhang, et al. (2002). "A masked NES in INI1/hSNF5 mediates hCRM1-dependent nuclear export: Implications for tumorigenesis." EMBO Journal **21**(1-2): 31-42.
- Craigie, R. (2001). "HIV integrase, a brief overview from chemistry to therapeutics." J.Biol.Chem. **276**(26): 23213-23216.
- Craigie, R., K. Mizuuchi, et al. (1991). "A rapid in vitro assay for HIV DNA integration." Nucleic Acids Research **19**(10): 2729-2734.
- Crise, B., L. Buonocore, et al. (1990). "CD4 is retained in the endoplasmic reticulum by the human immunodeficiency virus type 1 glycoprotein precursor." Journal of Virology **64**(11): 5585-5593.

- Dagert, M. and S. D. Ehrlich (1979). "Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells." Gene **6**(1): 23-28.
- Daniels, T., J. Zhang, et al. (2005). "Antinuclear autoantibodies in prostate cancer: Immunity to LEDGF/p75, a survival protein highly expressed in prostate tumors and cleaved during apoptosis." Prostate **62**(1): 14-26.
- Das, S., J. Cano, et al. (2009). "Multimerization and DNA Binding Properties of INI1/hSNF5 and Its Functional Significance." Journal of Biological Chemistry **284**(30): 19903-19914.
- De Clercq, E. (2009). "Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV." International Journal of Antimicrobial Agents **33**(4): 307-320.
- Delelis, O., V. Parissi, et al. (2007). "Efficient and specific internal cleavage of a retroviral palindromic DNA sequence by tetrameric HIV-1 integrase." PloS One **2**(7): e608-e608.
- Deprez, E., P. Tauc, et al. (2000). "Oligomeric States of the HIV-1 Integrase As Measured by Time-Resolved Fluorescence Anisotropy." Biochemistry **39**(31): 9275-9284.
- Deprez, E., P. Tauc, et al. (2001). "DNA binding induces dissociation of the multimeric form of HIV-1 integrase: a time-resolved fluorescence anisotropy study." Proceedings of the National Academy of Sciences of the United States of America **98**(18): 10090-5.
- DeSoultrait, V., A. Caumont, et al. (2002). "A novel short peptide is a specific inhibitor of the human immunodeficiency virus type 1 integrase." J.Mol.Biol. **318**(1): 45-58.
- Dietz, F., S. Franken, et al. (2002). "The family of hepatoma-derived growth factor proteins: Characterization of a new member HRP-4 and classification of its subfamilies." Biochemical Journal **366**(2): 491-500.
- Doranz, B. J., J. Rucker, et al. (1996). "A dual-tropic primary HIV-1 isolate that uses fusin and the  $\beta^2$ -chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors." Cell **85**(7): 1149-1158.
- Drelich, M., M. Haenggi, et al. (1993). "Conserved residues Pro-109 and Asp-116 are required for interaction of the human immunodeficiency virus type 1 integrase protein with its viral DNA substrate." J. Virol. **67**(8): 5041-5044.
- Drelich, M., M. Haenggi, et al. (1993). "Conserved residues Pro-109 and Asp-116 are required for interaction of the human immunodeficiency virus type 1 integrase protein with its viral DNA substrate." J.Virol. **67**(8): 5041-5044.
- Dubochet, J., M. Adrian, et al. (1988). "Cryo-electron microscopy of vitrified specimens." Quarterly Reviews of Biophysics **21**(2): 129-228.
- Dubochet, J., F. P. Booy, et al. (1981). "Low temperature electron microscopy." Annu Rev Biophys Bioeng **10**: 133-49.
- Dyda, F., A. B. Hickman, et al. (1994). "Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases." Science. **266**(5193): 1981-1986.
- Dyda, F., A. B. Hickman, et al. (1994). "Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases." Science (New York, N.Y.) **266**(5193): 1981-1986.
- Emerman, M. and M. H. Malim (1998). "HIV-1 regulatory/accessory genes: Keys to unraveling viral and host cell biology." Science **280**(5371): 1880-1884.
- Emiliani, S., A. Mousnier, et al. (2005). "Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication." Journal of Biological Chemistry **280**(27): 25517-25523.
- Emiliani, S., A. Mousnier, et al. (2005). "Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication." J.Biol.Chem. **280**(27): 25517-25523.
- Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr.D.Biol.Crystallogr. **60**(Pt 12 Pt 1): 2126-2132.
- Engelman, A. and R. Craigie (1992). "Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro." Journal of Virology **66**(11): 6361-6369.

- Engelman, A. and R. Craigie (1995). "Efficient magnesium-dependent human immunodeficiency virus type 1 integrase activity." J.Virol. **69**(9): 5908-5911.
- Engelman, A., A. B. Hickman, et al. (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**(9): 5911-5917.
- Feng, Y., C. C. Broder, et al. (1996). "HIV-1 entry cofactor: Functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor." Science **272**(5263): 872-877.
- Fernandjian, S., R. S. Maroun, et al. (2001). "Self-association of an amphipathic helix peptide inhibitor of HIV-1 integrase assessed by electro spray ionization mass spectrometry in trifluoroethanol/water mixtures." Rapid Commun.Mass Spectrom. **15**(5): 320-324.
- Finn, R. D., J. Mistry, et al. (2010). "The Pfam protein families database." Nucl. Acids Res. **38**(suppl\_1): D211-222-D211-222.
- Foti, M., A. Mangasarian, et al. (1997). "Nef-mediated clathrin-coated pit formation." Journal of Cell Biology **139**(1): 37-47.
- Franke, E. K., H. E. H. Yuan, et al. (1994). "Specific incorporation of cyclophilin A into HIV-1 virions." Nature **372**(6504): 359-362.
- Freed, E. O., G. Englund, et al. (1995). "Role of the basic domain of human immunodeficiency virus type 1 matrix in macrophage infection." Journal of Virology **69**(6): 3949-3954.
- Gallay, P., T. Hope, et al. (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**(18): 9825-9830.
- Ganapathy, V. and C. A. Casiano (2004). "Autoimmunity to the Nuclear Autoantigen DFS70 (LEDGF): What Exactly Are the Autoantibodies Trying to Tell Us?" Arthritis and Rheumatism **50**(3): 684-688.
- Gao, K., S. L. Butler, et al. (2001). "Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes." EMBO J **20**(13): 3565-3576.
- Gao, K., S. L. Butler, et al. (2001). "Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes." EMBO J. **20**(13): 3565-3576.
- Garnier, J., D. J. Osguthorpe, et al. (1978). "Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins." Journal of Molecular Biology **120**(1): 97-120.
- Ge, H., Y. Si, et al. (1998). "Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation." EMBO Journal **17**(22): 6723-6729.
- Goff, S. P. (1990). "Retroviral reverse transcriptase: Synthesis, structure, and function." Journal of Acquired Immune Deficiency Syndromes **3**(8): 817-831.
- Golas, M. M., B. Sander, et al. (2003). "Molecular architecture of the multiprotein splicing factor SF3b." Science **300**(5621): 980-4.
- Goldgur, Y., R. Craigie, et al. (1999). "Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design." Proc.Natl.Acad.Sci.U.S.A. **96**(23): 13040-13043.
- Goldgur, Y., F. Dyda, et al. (1998). "Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium." Proceedings of the National Academy of Sciences of the United States of America **95**(16): 9150-4.
- Goldgur, Y., F. Dyda, et al. (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**(16): 9150-9154.
- Grandgenett, D. P., S. Bera, et al. (2009). "Biochemical and biophysical analyses of concerted (U5/U3) integration." Methods (San Diego, Calif.) **47**(4): 229-236.
- Greenberg, M. E., S. Bronson, et al. (1997). "Co-localization of HIV-1 Nef with the AP-2 adaptor protein complex correlates with Nef-induced CD4 down-regulation." EMBO Journal **16**(23): 6964-6976.

- Greenwald, J., V. Le, et al. (1999). "The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity." *Biochemistry*. **38**(28): 8892-8898.
- Grobler, J. A., K. Stillmock, et al. (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**(10): 6661-6666.
- Guiot, E. (2006). "Relationship between the Oligomeric Status of HIV-1 Integrase on DNA and Enzymatic Activity." *Journal of Biological Chemistry* **281**(32): 22707-22719.
- Hare, S. and P. Cherepanov (2009). "The Interaction Between Lentiviral Integrase and LEDGF: Structural and Functional Insights." *Viruses* **1**: 780-801.
- Hare, S., N. F. Di, et al. (2009). "Structural basis for functional tetramerization of lentiviral integrase." *PLoS.Pathog.* **5**(7): e1000515.
- Hare, S., S. S. Gupta, et al. (2010). "Retroviral intasome assembly and inhibition of DNA strand transfer." *Nature* **464**(7286): 232-236.
- Hare, S., M. C. Shun, et al. (2009). "A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75." *PLoS.Pathog.* **5**(1): e1000259.
- Harper, A. L., L. M. Skinner, et al. (2001). "Use of patient-derived human immunodeficiency virus type 1 integrases to identify a protein residue that affects target site selection." *J.Virol.* **75**(16): 7756-7762.
- Hazuda, D. J., P. Felock, et al. (2000). "Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells." *Science*. **287**(5453): 646-650.
- Heger, A., C. A. Wilton, et al. (2005). "ADDA: a domain database with global coverage of the protein universe." *Nucl. Acids Res.* **33**(suppl\_1): D188-191-D188-191.
- Hehl, E. A., P. Joshi, et al. (2004). "Interaction between human immunodeficiency virus type 1 reverse transcriptase and integrase proteins." *J.Virol.* **78**(10): 5056-5067.
- Herrmann, C. H. and A. P. Rice (1995). "Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: Candidate for a Tat cofactor." *Journal of Virology* **69**(3): 1612-1620.
- Hjelmeland, L. M. (1980). "A nondenaturing zwitterionic detergent for membrane biochemistry: design and synthesis." *Proceedings of the National Academy of Sciences of the United States of America* **77**(11): 6368-6370.
- Hogg, R. S., D. R. Bangsberg, et al. (2006). "Emergence of drug resistance is associated with an increased risk of death among patients first starting HAART." *PLoS Medicine* **3**(9): e356-e356.
- Hoxie, J. A., J. D. Alpers, et al. (1986). "Alterations in T4 (CD4) protein and mRNA synthesis in cells infected with HIV." *Science* **234**(4780): 1123-1127.
- Inoue, H., H. Nojima, et al. (1990). "High efficiency transformation of Escherichia coli with plasmids." *Gene* **96**(1): 23-28.
- Invitrogen™ (2003). "Gateway® Technology A universal technology to clone DNA sequences for functional analysis and expression in multiple systems." *Manual Catalog nos. 12535-019 and 12535-027*(version D).
- Jaskolski, M., J. N. Alexandratos, et al. (2009). "Piecing together the structure of retroviral integrase, an important target in AIDS therapy." *FEBS J* **276**(11): 2926-2946.
- Jenkins, T. M., A. Engelman, et al. (1996). "A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization." *J.Biol.Chem.* **271**(13): 7712-7718.
- Jenkins, T. M., D. Esposito, et al. (1997). "Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking." *EMBO J* **16**(22): 6849-6859.
- Jenkins, T. M., A. B. Hickman, et al. (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**(13): 6057-6061.

- Jing, N., W. Xiong, et al. (2002). "Potassium-dependent folding: a key to intracellular delivery of G-quartet oligonucleotides as HIV inhibitors." *Biochemistry*, **41**(17): 5397-5403.
- Jing, N. and X. Xu (2001). "Rational drug design of DNA oligonucleotides as HIV inhibitors." *Curr. Drug Targets. Infect. Disord.* **1**(2): 79-90.
- Johnson, A. A., J. M. Sayer, et al. (2006). "Effect of DNA modifications on DNA processing by HIV-1 integrase and inhibitor binding: role of DNA backbone flexibility and an open catalytic site." *J. Biol. Chem.* **281**(43): 32428-32438.
- Jowett, J. B. M., V. Planelles, et al. (1995). "The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle." *Journal of Virology* **69**(10): 6304-6313.
- Kalpana, G. V., S. Marmon, et al. (1994). "Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5." *Science* **266**: 2002-2006.
- Kalpana, G. V., S. Marmon, et al. (1994). "Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5." *Science*, **266**(5193): 2002-2006.
- Kalpana, G. V., S. Marmon, et al. (1994). "Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5." *Science* **266**(5193): 2002-2006.
- Karki, R. G., Y. Tang, et al. (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**(12): 739-760.
- Katz, R. A. and A. M. Skalka (1994). "The retroviral enzymes." *Annual Review of Biochemistry* **63**: 133-173.
- LaFemina, R. L., C. L. Schneider, et al. (1992). "Requirement of active human immunodeficiency virus type 1 integrase enzyme for productive infection of human T-lymphoid cells." *J. Virol.* **66**(12): 7414-7419.
- Le Gall, S., L. Erdtmann, et al. (1998). "Nef interacts with the  $\hat{1}/4$  subunit of clathrin adaptor complexes and reveals a cryptic sorting signal in MHC I molecules." *Immunity* **8**(4): 483-495.
- Lecompte, O., J. D. Thompson, et al. (2001). "Multiple alignment of complete sequences (MACS) in the post-genomic era." *Gene* **270**(1-2): 17-30.
- Lee, S. P. and M. K. Han (1996). "Zinc stimulates Mg<sup>2+</sup>-dependent 3'-processing activity of human immunodeficiency virus type 1 integrase in vitro." *Biochemistry*, **35**(12): 3837-3844.
- Lewinski, M. K. and F. D. Bushman (2005). "Retroviral DNA integration--mechanism and consequences." *Adv. Genet.* **55**: 147-181.
- Linding, R., L. J. Jensen, et al. (2003). "Protein disorder prediction: implications for structural proteomics." *Structure (London, England: 1993)* **11**(11): 1453-1459.
- Linding, R., R. B. Russell, et al. (2003). "GlobPlot: exploring protein sequences for globularity and disorder." *Nucl. Acids Res.* **31**(13): 3701-3708.
- Llano, M., D. T. Saenz, et al. (2006). "An essential role for LEDGF/p75 in HIV integration." *Science*, **314**(5798): 461-464.
- Llano, M., D. T. Saenz, et al. (2006). "An essential role for LEDGF/p75 in HIV integration." *Science* **314**(5798): 461-464.
- Llano, M., M. Vanegas, et al. (2004). "LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes." *Journal of Virology* **78**(17): 9524-9537.
- Llano, M., M. Vanegas, et al. (2004). "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* **78**: 9524-9537.
- Lo Conte, L., S. E. Brenner, et al. (2002). "SCOP database in 2002: refinements accommodate structural genomics." *Nucl. Acids Res.* **30**(1): 264-267.
- Lodi, P. J., J. A. Ernst, et al. (1995). "Solution structure of the DNA binding domain of HIV-1 integrase." *Biochemistry*, **34**(31): 9826-9833.
- Lubkowski, J., F. Yang, et al. (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**(9): 4831-4836.

- Lutzke, R. A. and R. H. Plasterk (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**(6): 4841-4848.
- Lutzke, R. A., C. Vink, et al. (1994). "Characterization of the minimal DNA-binding domain of the HIV integrase protein." Nucleic Acids Res. **22**(20): 4125-4131.
- Maertens, G., P. Cherepanov, et al. (2004). "Identification and characterization of a functional nuclear localization signal in the HIV-1 integrase interactor LEDGF/p75." Journal of Biological Chemistry **279**(32): 33421-33429.
- Maertens, G., P. Cherepanov, et al. (2003). "LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells." J.Biol.Chem. **278**(35): 33528-33539.
- Maignan, S., J. P. Guilloteau, et al. (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**(2): 359-368.
- Malim, M. H., L. S. Tiley, et al. (1990). "HIV-1 structural gene expression requires binding of the rev trans-activator to its RNA target sequence." Cell **60**(4): 675-683.
- Marchand, C., K. Maddali, et al. (2009). "HIV-1 IN inhibitors: 2010 update and perspectives." Current Topics in Medicinal Chemistry **9**(11): 1016-1037.
- Margottin, F., S. P. Bour, et al. (1998). "A novel human WD protein, h- $\hat{I}^2$ TrCP, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif." Molecular Cell **1**(4): 565-574.
- Maroun, M., O. Delelis, et al. (2006). "Inhibition of early steps of HIV-1 replication by SNF5/Ini1." Journal of Biological Chemistry **281**(32): 22736-22743.
- Maroun, M., O. Delelis, et al. (2006). "Inhibition of early steps of HIV-1 replication by SNF5/Ini1." J.Biol.Chem. **281**(32): 22736-22743.
- Mély, Y., H. de Rocquigny, et al. (1995). "Binding of the HIV-1 nucleocapsid protein to the primer tRNA(3Lys), in vitro, is essentially not specific." The Journal of Biological Chemistry **270**(4): 1650-1656.
- Merkel, G., M. D. Andrade, et al. (2008). "Oligonucleotide-based assays for integrase activity." Methods In Press, Corrected Proof.
- Merkel, G., M. D. Andrade, et al. (2009). "Oligonucleotide-based assays for integrase activity." Methods (San Diego, Calif.) **47**(4): 243-248.
- Michel, F., C. Crucifix, et al. (2009). "Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor." The EMBO Journal.
- Molteni, V., J. Greenwald, et al. (2001). "Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain." Acta Crystallogr.D.Biol.Crystallogr. **57**(Pt 4): 536-544.
- Moore, J. P. (1997). "Coreceptors: Implications for HIV pathogenesis and therapy." Science **276**(5309): 51-52.
- Morozov, A., E. Yung, et al. (1998). "Structure-function analysis of integrase interactor 1/hSNF5L1 reveals differential properties of two repeat motifs present in the highly conserved region." Proceedings of the National Academy of Sciences of the United States of America **95**(3): 1120-1125.
- Morozov, A., E. Yung, et al. (1998). "Structure-function analysis of integrase interactor 1/hSNF5L1 reveals differential properties of two repeat motifs present in the highly conserved region." Proceedings of the National Academy of Sciences of the United States of America **95**(3): 1120-1125.
- Mousnier, A., N. Kubat, et al. (2007). "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. **104**(34): 13615-13620.
- Mulder, L. C., L. A. Chakrabarti, et al. (2002). "Interaction of HIV-1 integrase with DNA repair protein hRad18." J.Biol.Chem. **277**(30): 27489-27493.

- N.I.A.I.D. (2010). "The Evidence That HIV Causes AIDS." from <http://www3.niaid.nih.gov/topics/HIVAIDS/Understanding/howHIVCausesAIDS/HIVcausesAIDS.htm>.
- Needleman, S. B. and C. D. Wunsch (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins." *Journal of Molecular Biology* **48**(3): 443-453.
- Nishizawa, Y., J. Usukura, et al. (2001). "Spatial and temporal dynamics of two alternatively spliced regulatory factors, lens epithelium-derived growth factor (LEDGF/p75) and p52, in the nucleus." *Cell and Tissue Research* **305**(1): 107-114.
- Nollstadt, K. H., Y. D. Karkhanis, et al. (1989). "Potential of the sulfobetaine detergent Zwittergent 3-12 as a desorbing agent in biospecific and bioselective affinity chromatography." *Journal of Chromatography* **497**: 87-100.
- Nominé, Y., T. Ristriani, et al. (2001). "Formation of soluble inclusion bodies by hpv e6 oncoprotein fused to maltose-binding protein." *Protein Expression and Purification* **23**(1): 22-32.
- Ohno, M., M. Fornerod, et al. (1998). "Nucleocytoplasmic transport: The last 200 nanometers." *Cell* **92**(3): 327-336.
- ONUSIDA (2008). "Rapport sur l'épidémie mondiale de sida 2008."
- Ota, T., Y. Suzuki, et al. (2004). "Complete sequencing and characterization of 21,243 full-length human cDNAs." *Nat Genet* **36**(1): 40-45.
- Ouali, M., C. Laboulais, et al. (2000). "Modeling of the inhibition of retroviral integrases by styrylquinoline derivatives." *J.Med.Chem.* **43**(10): 1949-1957.
- Pandey, K. K., S. Sinha, et al. (2007). "Transcriptional coactivator LEDGF/p75 modulates human immunodeficiency virus type 1 integrase-mediated concerted integration." *Journal of Virology* **81**(8): 3969-79.
- Pettersen, E. F., T. D. Goddard, et al. (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." *Journal of Computational Chemistry* **25**(13): 1605-1612.
- Piguet, V., Y. L. Chen, et al. (1998). "Mechanism of Nef-induced CD4 endocytosis: Nef connects CD4 with the  $\hat{1}/4$  chain of adaptor complexes." *EMBO Journal* **17**(9): 2472-2481.
- Plewniak, F., L. Bianchetti, et al. (2003). "PipeAlign: A new toolkit for protein family analysis." *Nucleic Acids Research* **31**(13): 3829-3832.
- Podtelezhnikov, A. A., K. Gao, et al. (2003). "Modeling HIV-1 integrase complexes based on their hydrodynamic properties." *Biopolymers*. **68**(1): 110-120.
- Pollard, V. W. and M. H. Malim (1998). "The HIV-1 Rev protein." *Annual Review of Microbiology* **52**: 491-532.
- Pommier, Y., A. A. Johnson, et al. (2005). "Integrase inhibitors to treat HIV/AIDS." *Nat.Rev.Drug Discov.* **4**(3): 236-248.
- Prilusky, J., C. E. Felder, et al. (2005). "FoldIndex(C): a simple tool to predict whether a given protein sequence is intrinsically unfolded." *Bioinformatics* **21**(16): 3435-3438.
- Pumfery, A., L. Deng, et al. (2003). "Chromatin remodeling and modification during HIV-1 Tat-activated transcription." *Current HIV Research* **1**(3): 343-362.
- Qiu, C., K. Sawada, et al. (2002). "The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds." *Nature Structural Biology* **9**(3): 217-224.
- Radermacher, M. (1988). "Three-Dimensional reconstruction of single particles from random and nonrandom tilt series." *Journal of Electron Microscopy Technique* **9**(4): 359-394.
- Re, F., D. Braaten, et al. (1995). "Human immunodeficiency virus type 1 Vpr arrests the cell cycle in G2 by inhibiting the activation of p34(cdc2)-cyclin B." *Journal of Virology* **69**(11): 6859-6864.
- Reines, D., J. W. Conaway, et al. (1996). "The RNA polymerase II general elongation factors." *Trends in Biochemical Sciences* **21**(9): 351-355.
- Retroviruses (1997). *Retroviruses*, Cold Spring Harbor Laboratory Press.
- Rice, W. G., J. G. Supko, et al. (1995). "Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS." *Science* **270**(5239): 1194-1197.

- Roberts, C. W. M. and S. H. Orkin (2004). "The SWI/SNF complex--chromatin and cancer." Nature Reviews. Cancer **4**(2): 133-142.
- Rogel, M. E., L. I. Wu, et al. (1995). "The human immunodeficiency virus type 1 vpr gene prevents cell proliferation during chronic infection." Journal of Virology **69**(2): 882-888.
- Rost, B., G. Yachdav, et al. (2004). "The PredictProtein server." Nucl. Acids Res. **32**(suppl\_2): W321-326-W321-326.
- Schubert, U., L. C. Antón, et al. (1998). "CD4 glycoprotein degradation induced by human immunodeficiency virus type 1 Vpu protein requires the function of proteasomes and the ubiquitin- conjugating pathway." Journal of Virology **72**(3): 2280-2288.
- Sharma, P., D. P. Singh, et al. (2000). "Activation of LEDGF gene by thermal- and oxidative-stresses." Biochemical and Biophysical Research Communications **276**(3): 1320-1324.
- Shinohara, T., D. P. Singh, et al. (2002). "LEDGF, a survival factor, activates stress-related genes." Progress in Retinal and Eye Research **21**(3): 341-358.
- Singh, D. P., N. Fatma, et al. (2001). "LEDGF binds to heat shock and stress-related element to activate the expression of stress-related genes." Biochemical and Biophysical Research Communications **283**(4): 943-955.
- Singh, D. P., N. Ohguro, et al. (2000). "Lens epithelium-derived growth factor: Effects on growth and survival of lens epithelial cells, keratinocytes, and fibroblasts." Biochemical and Biophysical Research Communications **267**(1): 373-381.
- Sinha, S. and D. P. Grandgenett (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**(13): 8208-8216.
- Smolov, M., M. Gottikh, et al. (2006). "Kinetic study of the HIV-1 DNA 3'-end processing." The FEBS Journal **273**(6): 1137-1151.
- Sorin, M., E. Yung, et al. (2006). "HIV-1 replication in cell lines harboring IN1/hSNF5 mutations." Retrovirology **3**.
- Sova, P. and D. J. Volsky (1993). "Efficiency of viral DNA synthesis during infection of permissive and nonpermissive cells with vif-negative human immunodeficiency virus type 1." Journal of Virology **67**(10): 6322-6326.
- Stec, I., S. B. Nagl, et al. (2000). "The PWWP domain: A potential protein-protein interaction domain in nuclear proteins influencing differentiation?" FEBS Letters **473**(1): 1-5.
- Suhre, K., J. Navaza, et al. (2006). "NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps." Acta Crystallographica. Section D, Biological Crystallography **62**(Pt 9): 1098-1100.
- Tama, F., O. Miyashita, et al. (2004). "Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis." J.Mol.Biol. **337**(4): 985-999.
- Telesnitsky, A. and S. P. Goff (1997). "Reverse transcriptase and the generation of retroviral DNA." Retroviruses: 121-160.
- Temesgen, Z., D. Warnke, et al. (2006). "Current status of antiretroviral therapy." Expert.Opin.Pharmacother. **7**(12): 1541-1554.
- Thompson, J. D., D. G. Higgins, et al. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Research **22**(22): 4673-4680.
- Thompson, J. D., A. Muller, et al. (2006). "MACSIMS: multiple alignment of complete sequences information management system." BMC Bioinformatics **7**: 318-318.
- Thompson, J. D., F. Plewniak, et al. (1999). "A comprehensive comparison of multiple sequence alignment programs." Nucleic Acids Research **27**(13): 2682-2690.
- Thompson, J. D., F. Plewniak, et al. (2001). "Towards a reliable objective function for multiple sequence alignments." Journal of Molecular Biology **314**(4): 937-951.
- Thompson, J. D., V. Prigent, et al. (2004). "LEON: multiple aLignment Evaluation Of Neighbours." Nucl. Acids Res. **32**(4): 1298-1307.

- Thompson, J. D., J. C. Thierry, et al. (2003). "RASCAL: rapid scanning and correction of multiple sequence alignments." *Bioinformatics* **19**(9): 1155-1161.
- Tsuchiya, H., T. Iseda, et al. (1996). "Identification of a novel protein (VBP-1) binding to the von Hippel-Lindau (VHL) tumor suppressor gene product." *Cancer Res.* **56**(13): 2881-2885.
- Turelli, P., V. Doucas, et al. (2001). "Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: Interference with early steps of viral replication." *Molecular Cell* **7**(6): 1245-1254.
- Turlure, F., E. Devroe, et al. (2004). "Human cell proteins and human immunodeficiency virus DNA integration." *Frontiers in bioscience : a journal and virtual library.* **9**: 3187-3208.
- Turlure, F., E. Devroe, et al. (2004). "Human cell proteins and human immunodeficiency virus DNA integration." *Front Biosci.* **9**:3187-208.: 3187-3208.
- Turner, B. G. and M. F. Summers (1999). "Structural biology of HIV." *Journal of Molecular Biology* **285**(1): 1-32.
- Van Heel, M. (1987). "Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction." *Ultramicroscopy* **21**(2): 111-123.
- van Heel, M. and J. Frank (1981). "Use of multivariate statistics in analysing the images of biological macromolecules." *Ultramicroscopy* **6**(2): 187-194.
- van Heel, M., M. Schatz, et al. (1992). "Correlation functions revisited." *Ultramicroscopy* **46**(1-4): 307-316.
- Van Maele, B., K. Busschots, et al. (2006). "Cellular co-factors of HIV-1 integration." *Trends in Biochemical Sciences* **31**(2): 98-105.
- Vandegraaff, N., E. Devroe, et al. (2006). "Biochemical and genetic analyses of integrase-interacting proteins lens epithelium-derived growth factor (LEDGF)/p75 and hepatoma-derived growth factor related protein 2 (HRP2) in preintegration complex function and HIV-1 replication." *Virology* **346**(2): 415-426.
- Vanegas, M., M. Llano, et al. (2005). "Identification of the LEDGF/p75 HIV-1 integrase-interaction domain and NLS reveals NLS-independent chromatin tethering." *Journal of Cell Science* **118**(8): 1733-1743.
- Wang, J. Y., H. Ling, et al. (2001). "Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein." *EMBO J.* **20**(24): 7333-7343.
- Wang, W., J. Cote, et al. (1996). "Purification and biochemical heterogeneity of the mammalian SWI-SNF complex." *EMBO Journal* **15**(19): 5370-5382.
- Wang, Y., H. Klock, et al. (2005). "Homogeneous high-throughput screening assays for HIV-1 integrase 3beta-processing and strand transfer activities." *Journal of Biomolecular Screening: The Official Journal of the Society for Biomolecular Screening* **10**(5): 456-462.
- Waterhouse, A. M., J. B. Procter, et al. (2009). "Jalview Version 2--a multiple sequence alignment editor and analysis workbench." *Bioinformatics* **25**(9): 1189-1191.
- Wei, P., M. E. Garber, et al. (1998). "A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA." *Cell* **92**(4): 451-462.
- Wei, X., S. K. Ghosh, et al. (1995). "Viral dynamics in human immunodeficiency virus type 1 infection." *Nature* **373**(6510): 117-122.
- Whitcomb, J. M. and S. H. Hughes (1992). "Retroviral reverse transcription and integration: Progress and problems." *Annual Review of Cell Biology* **8**: 275-306.
- Wicker, N., D. Dembele, et al. (2002). "Density of points clustering, application to transcriptomic data analysis." *Nucl. Acids Res.* **30**(18): 3992-4000.
- Wicker, N., G. R. Perrin, et al. (2001). "Secator: A Program for Inferring Protein Subfamilies from Phylogenetic Trees." *Mol Biol Evol* **18**(8): 1435-1441.
- Wiskerchen, M. and M. A. Muesing (1995). "Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells." *J.Virol.* **69**(1): 376-386.

- Wiskerchen, M. and M. A. Muesing (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**(1): 597-601.
- Wright, P. E. and H. J. Dyson (1999). "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm." Journal of Molecular Biology **293**(2): 321-331.
- Wu, X., H. Liu, et al. (1999). "Human immunodeficiency virus type 1 integrase protein promotes reverse transcription through specific interactions with the nucleoprotein reverse transcription complex." J.Virol. **73**(3): 2126-2135.
- Yang, L. W. and I. Bahar (2005). "Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes." Structure. **13**(6): 893-904.
- Yang, Z. R., R. Thomson, et al. (2005). "RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins." Bioinformatics (Oxford, England) **21**(16): 3369-3376.
- Yung, E., M. Sorin, et al. (2001). "Inhibition of HIV-1 virion production by a transdominant mutant of integrase interactor 1." Nature Medicine **7**(8): 920-6.
- Yung, E., M. Sorin, et al. (2001). "Inhibition of HIV-1 virion production by a transdominant mutant of integrase interactor 1." Nature Medicine **7**(8): 920-926.
- Yung, E., M. Sorin, et al. (2004). "Specificity of Interaction of IN1/hSNF5 with Retroviral Integrases and Its Functional Significance." Journal of Virology **78**(5): 2222-2231.
- Zahm, J. A., S. Bera, et al. (2008). "Mechanisms of human immunodeficiency virus type 1 concerted integration related to strand transfer inhibition and drug resistance." Antimicrobial Agents and Chemotherapy **52**(9): 3358-68.
- Zheng, R., T. M. Jenkins, et al. (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**(24): 13659-13664.
- Zouhiri, F., J. F. Mouscadet, et al. (2000). "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. **43**(8): 1533-1540.

## Résumé

Le génome viral inversement transcrit du VIH-1 (Virus de l'Immunodéficience Humaine type 1) est intégré dans le génome de la cellule cible par l'intégrase, une protéine virale de 288 acides aminés. L'intégrase est partie intégrante d'un important complexe nucléoprotéique, le complexe de pré-intégration (CPI). Ce complexe est formé après la reverse transcription dans le cytoplasme de la cellule infectée. Il migre par l'intermédiaire des microtubules jusqu'à l'enveloppe nucléaire où il est transporté dans le noyau à travers le pore nucléaire. Le CPI est composé d'ADN et de protéines virales dont l'intégrase. D'autres protéines virales (VPR, NC, CA) et des protéines cellulaires participent à ce complexe. Parmi les protéines cellulaires identifiées s'illustrent notamment LEDGF (Lens Epithelium-Derived Growth Factor) et INI1 (INtegrase Interactor 1). Cette protéine aussi appelée SNF5 est une composante du complexe SWI-SNF de remodelage de la chromatine. Mon travail de thèse s'articule autour de la résolution de structures tridimensionnelles de l'intégrase du VIH-1 en complexe avec LEDGF et INI1. Mon sujet consistait à réaliser une stratégie permettant d'aboutir à la formation *in vitro* de ce complexe. Par une analyse bioinformatique de la protéine INI1, j'ai défini les limites de différents domaines potentiellement stables d'INI1 contenant le domaine de liaison à l'intégrase (IBD). Après clonage et expression dans le système bactérien *E.coli*. J'ai mis au point la production à grande échelle et la purification des fragments ayant la meilleure solubilité. Après avoir obtenu un complexe stable, J'ai résolu la structure du complexe ternaire entre l'intégrase pleine longueur type sauvage, LEDGF et l'IBD d'INI1 par microscopie électronique. Les structures cristallographiques connues des domaines de l'intégrase ont été positionnées dans la carte de densité électronique.

J'ai mis au point des tests fonctionnels originaux basés sur les techniques d'anisotropie de fluorescence qui m'ont permis de tester l'affinité du complexe pour l'ADN ainsi que l'activité enzymatique de 3'*processing*. Les résultats indiquent qu'au sein du complexe ternaire, l'IBD de INI1 réduit significativement l'affinité pour l'ADN cible et dans une moindre mesure l'affinité pour l'ADN viral. De plus, l'IBD d'INI1 inhibe totalement la réaction de 3'*processing*.

La structure du complexe ternaire nous permet de visualiser les positions relatives des différentes protéines dans le complexe. On montre que le complexe est formé de 4 molécules d'intégrase, deux molécules de LEDGF et deux molécules d'INI1. La combinaison de ces résultats avec les tests fonctionnels nous permettent de proposer un modèle et un rôle pour INI1 qui serait d'empêcher l'auto intégration de l'ADN viral sur lui-même quand le CPI se trouve dans le noyau cellulaire.

---

The viral genome reverse transcribed from the HIV-1 (Human Immunodeficiency Virus type I) is integrated in the target cell genome by the integrase, a viral protein of 288 amino acids. The integrase is a part of an important nucleoproteic complex, the preintegration complex (PIC). This complex is formed after the reverse transcription in the cytoplasm of infected cells. It migrates along the microtubules up to the nuclear envelope where it is carried in the nucleus through the nuclear pore. The PIC is composed of DNA, and viral proteins including integrase. Other viral proteins (VPR, NC, CA) and cellular proteins are part of this complex. Among the cellular proteins, two of them are the object of this work: LEDGF (Lens Epithelium-Derived Growth Factor) and INI1 (INtegrase Interactor 1). This last protein, also called SNF5, is a component of the SWI-SNF chromatin remodeling complex. The object of my PhD work is the understanding of the structure function relationship with the 3D structure solving of the complex between HIV-1 integrase in complex with LEDGF and INI1. I set up a strategy to reconstitute this complex *in vitro*. Using a bioinformatic approach, I defined the limits of domains, containing the region interacting with integrase, potentially stable of INI1. After cloning and expression in *E. coli*, I set up the production and purification of the more soluble fragments. After obtaining a stable complex I solved by electron microscopy the structure of the complex between the integrase, LEDGF and INI1. The known crystallographic structures of the integrase domains were fitted in the EM map.

I set up functional tests based on fluorescence anisotropy, which permit us to test the affinity of the complex for DNA and the 3'*processing* activity. The results show that INI1 reduces the affinity for the target DNA and for the viral DNA. Moreover INI1 inhibits totally the 3' *processing* reaction.

The structure of the ternary complex shows the relative position of the different proteins in the complex. We show that the complex is made of 4 integrase molecules, 2 LEDGF and 2 INI1 molecules. The comparison of these results with the functional tests allows us to propose a model and a role for INI1 which would be to prevent the auto integration of the viral DNA on itself when the PIC is in the nucleus.