N°d'ordre NNT : 2017LYSE1035

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N° 205
Ecole Doctorale Interdisciplinaire Sciences-Santé

Spécialité de doctorat : Biochimie

Soutenue publiquement le 08/03/2017, par :
Halima YAJJOU-HAMALIAN

Etudes moléculaires et structurales - d'un nouveau mode de dimérisation des intégrases rétrovirales pour développer des modulateurs d'oligomérisation ET - de l'intégrase porcine de PERV-A/C en complexe avec son cofacteur cellulaire humain Brd2 dans un contexte de xénotransplantation

---

Devant le jury composé de :

MIELE Adriana	PR	Université Claude Bernard Lyon 1	Présidente
MORERA Solange	DR	CNRS, Saclay	Rapporteure
PARISSI Vincent	DR	CNRS, Université de Bordeaux	Rapporteur
RUFF Marc	DR	CNRS, Illkirch	Rapporteur
GOUET Patrice	PR	Université Claude Bernard Lyon 1	Directeur de thèse
RONFORT Corinne	CR	INRA, Lyon	Co-directrice de thèse

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	**M. le Professeur Frédéric FLEURY**
Président du Conseil Académique	M. le Professeur Hamda BEN HADID
Vice-président du Conseil d'Administration	M. le Professeur Didier REVEL
Vice-président du Conseil Formation et Vie Universitaire	M. le Professeur Philippe CHEVALIER
Vice-président de la Commission Recherche	M. Fabrice VALLÉE
Directrice Générale des Services	Mme Dominique MARCHAND

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur G.RODE
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : Mme la Professeure C. BURILLON
Faculté d'Odontologie	Directeur : M. le Professeur D. BOURGEOIS
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : Mme la Professeure C. VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. X. PERROT
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : Mme la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. F. DE MARCHI
Département Biologie	Directeur : M. le Professeur F. THEVENARD
Département Chimie Biochimie	Directeur : Mme C. FELIX
Département GEP	Directeur : M. Hassan HAMMOURI
Département Informatique	Directeur : M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur : M. le Professeur G. TOMANOV
Département Mécanique	Directeur : M. le Professeur H. BEN HADID
Département Physique	Directeur : M. le Professeur J-C PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y.VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : M. B. GUIDERDONI
Polytech Lyon	Directeur : M. le Professeur E.PERRIN
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur C. VITON
Ecole Supérieure du Professorat et de l'Education	Directeur : M. le Professeur A. MOUGNIOTTE
Institut de Science Financière et d'Assurances	Directeur : M. N. LEBOISNE

Remerciements

Après trois années d'aventures scientifiques et humaines, j'aimerai remercier toutes les personnes qui ont rendu cette expérience si unique et qui m'ont conduit à ce jour mémorable. Mes remerciements s'adressent en premier lieu à l'Agence Nationale de Recherche sur le Sida et les hépatites virales ANRS, qui m'a fait confiance en m'attribuant une allocation de recherche.

Je souhaite remercier le Dr Christophe Geourjon, directeur de l'Institut de Biologie et Chimie des Protéines (IBCP Lyon-Gerland) et le Dr Jean-Michel Jault qui m'a accueilli dans son Unité Mixte de Recherche UMR 5086.

Je remercie également le Pr. Adriana Miele, le Dr Solange Morera, le Dr Vincent Parissi et le Dr Marc Ruff qui m'ont fait l'honneur de participer au jury de soutenance de thèse. Je les remercie pour le temps qu'ils ont consacré à la lecture de mon manuscrit, mais également pour les remarques pertinentes qui visent à améliorer sa qualité.

J'adresse mes sincères remerciements à mon directeur de thèse, le Pr. Patrice Gouet. Je lui suis grandement reconnaissante pour la confiance qu'il m'a accordée à de multiples reprises, son soutien et sa disponibilité. Je remercie également le Dr Corinne Ronfort, co-directrice de ma thèse, qui a su être présente malgré les changements professionnels qu'elle a vécu durant les deux dernières années.

Durant ce travail doctoral, j'ai pu utiliser différents équipements de la plateforme *Protein Science Facility* PSF (SFR Biosciences, UMS 3444). Ainsi, je souhaite remercier Adeline Page, Isabelle Zanela-Cléon, Frédéric Delolme (spectrométrie de masse), Roland Montserret (dichroïsme circulaire) et Virginie Gueguen-Chaignon (cristallogenèse).

Une partie de mon travail de thèse a été réalisé au sein de l'ancienne équipe « Rétrovirus, Intégration Virale et Vaccins », à l'INRA Lyon-Gerland. Je remercie Kathy Gallay pour la formation biomol, les discussions sur la « construction de fou », les débats plus personnels. Rencontrer des personnes comme toi redonne confiance en l'humain, merci. Je remercie également Catherine Luengo-Guyonnot pour les nombreuses tentatives de manip' et les multiples échecs, pour les échanges philosophiques et le réconfort que tu m'as apporté. Merci à Carine Maisse-Paradisi avec qui j'ai passé deux mois, entre Lyon-Sud et l'INRA. Nous n'étions pas gâtées par la science mais ça nous a rapprochés. Merci également à Cyrielle Vituret et à Emma Reungoat, pour les déjeuners au soleil et les évènements de vie partagés ensemble.

Je veux également exprimer toute ma gratitude au Dr Vincent Parissi et à Delphine Lapaillerie pour leur accueil durant mon séjour à Bordeaux, le partage de leur savoir scientifique et leur joie de vivre.

Je remercie les équipes collaboratrices du Dr Maurin (ENS Lyon) et du Dr Yannick Blanchard (ANSES Ploufragan) qui ont participé à ce travail de recherche. Je remercie sincèrement Maëlle Mosser pour son engagement dans ce travail doctoral. J'adresse plus particulièrement mes remerciements au Dr Yannick Blanchard pour sa gentillesse et son implication dans ce projet.

Je tiens à remercier tous les membres de l'équipe « Biocristallographie et Biologie Structurale des Cibles Thérapeutiques ». Merci à Nushin pour sa bienveillance. Merci à Lionel pour les fous rires et les Magaaa en tout genre, à Seb pour le partage de bons plans balades (Center Park) et les pauses sans

science, à Fred (*alias* Abdel Bachir) pour ses conseils techniques et ses belles descriptions du désert (jmèl va !). Un grand merci à Loïc Wawique (*alias* Pilou) avec qui j'ai tant partagé depuis la L1. Toutes nos sessions révision, entrecoupées de leçons de conduite en Pilippe-mobil, valaient le coup.

Je remercie également les anciens membres. Merci à Magali pour les potins et les rigolades, à Allison pour ses précieux conseils sur l'intégrase. Merci à mon formateur Romain, une âme capillairement généreuse qui m'a transmis ses connaissances en science et en canidé. Sans toi, les techniques de qualité n'ont plus la même saveur.

À ces remerciements, j'associe bien évidemment les membres de mon équipe. Merci au chef Patrice. Merci à CHRISTOPHE (parce que la taille c'est important) pour les jeux de mots, les corrections qui détendent l'atmosphère et les conseils manip que tu m'as donné durant ces trois ans. Merci à Xavier pour son soutien tant scientifique que personnel, les moments philosophiques sur le parvis de l'institut et son regard perfectionniste sur mes présentations. Merci à Marie pour sa douceur et ses mots rassurants. Merci à Francesca, notre maman italienne, pour son ouverture d'esprit (je crois que nous avons abordé tous les sujets possibles ensemble) et ses succulents desserts. Merci à Margaux pour ses critiques cinématographiques, sa bonne humeur et ses western-blot de qualité ?! Enfin je remercie ma petite réunionnaise, Stelle, pour son soutien sans faille et sa présence, même dans les moments peu joyeux : « I'm still standing » et autres chansons écoutées sur la route du synchrotron, dans le labo ou en soirée resteront à jamais gravées dans ma mémoire.

Mon quotidien à l'IBCP a également été égayé par différentes personnes, ainsi je souhaite remercier Célia, Morgane, Thomas, Charlotte, JB, Arthur, Jul, Stéph, (Sylvain Vallet)², Franky, Anne-So et Capu pour les sessions Synchrotron, les pauses thé, les soirées au Ninkasi ou au Bureau et/ou les congrès. Je remercie également les gestionnaires Souad, Dorothée et Christine pour leur gentillesse et leur professionnalisme.

Mon entourage « hors labo » m'a également été d'une grande aide durant ces trois années. Je pense bien entendu à ma précieuse amie Marie qui me soutient depuis presque 18 ans. Merci pour cette amitié fraternelle sincère. Merci à Maryse, Nono et Lili avec qui j'ai de beaux souvenirs. Un grand merci à mes amis rencontrés à l'école et toujours bien présents dans ma vie : Elo, Alice, Manu, Taki, Cha, Vaness mais aussi la FunTeam avec Matou, O2 et Lilou. Qu'il est bon de passer du temps avec vous. À mon ami et frère Mouss qui continue à me supporter et à être présent malgré la distance, tabarnak !

Enfin, je tiens à remercier ces belles personnes sans qui rien n'aurait été possible : ma famille. À mes parents, qui ont toujours cru en moi et dont les encouragements ont été une force pour m'aider à avancer. À mes sœurs jumelles Oumnia et Najet qui m'aident à grandir et à croire en moi. À Ilham qui me manque tellement chaque jour. À mes merveilleuses nièces Ana et Eva. À Cyrille, pour tout le bien qu'il apporte à Oum. Les circonstances actuelles nous poussent à évoluer mais également à montrer nos sentiments, sachez que je vous aime de tout mon cœur.

Je terminerai en remerciant mon époux et meilleur ami Razmig, pour son soutien indéfectible au quotidien. Le temps a renforcé nos liens et je suis fière de l'équipe que nous formons. Bientôt dix ans. Quelles belles années de partage, de complicité et d'amour. Puissent-elles être encore nombreuses.

Soyez le changement que vous voulez voir dans le monde. Ghandi

Sommaire

Remerciements……3
Sommaire……5
Abréviations……9
Curriculum vitae……12

Introduction bibliographique……15

I. Rétrovirus……16
1. Généralités……16
2. Taxonomie……17
3. Organisation génomique et structurale……20
3.1. Le génome rétroviral……20
3.1.1. Les LTR……21
3.1.2. Les gènes rétroviraux……22
3.1.3. Cas des rétrovirus complexes……23
3.1.4. Cas des rétrovirus oncogènes……24
3.2. La morphologie de la particule virale……25
4. Cycle réplicatif rétroviral……25
4.1. La phase pré-intégrative……25
4.2. La phase post-intégrative……26
4.3. Inhibiteurs antirétroviraux……28
4.3.1. Ciblage de la fusion et de l'entrée……29
4.3.2. Ciblage de la transcription inverse……29
4.3.3. Ciblage de la protéase……30
4.3.4. Ciblage de l'intégrase……30

II. Intégrase……31
1. Complexe de pré-intégration……31
2. Aspects fonctionnels……34
2.1. Intégrase et transposases……34
2.2. Intégration concertée chez les rétrovirus……36
2.3. Fonctions secondaires……37
3. Aspects structuraux……40
3.1. L'intégrase seule……40
3.1.1. Le domaine amino-terminal……41
3.1.2. Le domaine central catalytique……42
3.1.3. Le domaine carboxy-terminal……43
3.1.4. Structures des fragments à deux domaines……44
3.1.5. Oligomérisation de l'intégrase……45
3.1.5.1. Modèles tétramériques……46
3.1.5.2. Modèles octamériques……49
3.1.6. Interaction intasome-nucléosome……50
3.2. Les partenaires cellulaires essentiels à l'intégration productive……51
3.2.1. LEDGF/p75……51
3.2.2. La famille BET……53

Objectifs du projet doctoral……57

Matériels et méthodes……61

I. Techniques de biologie moléculaire……62
1. Constructions plasmidiques……62
2. Construction « multiTEV »……64
2.1. Amorces utilisées……65
2.2. Fabrication du plasmide recombinant « multiTEV »……65
2.2.1. Amplification des gènes d'intérêt……65
2.2.2. Formation de l'insert……67
2.2.3. Clonage avec le système In-Fusion®……67
3. Transformation de bactéries compétentes……68
4. Extraction de l'ADN……69
5. Dosage de l'ADN……69
6. Séquençage de l'ADN……70
7. Souches bactériennes utilisées……70
7.1. Stockage des bactéries transformées avec l'ADN d'intérêt……70
7.2. Surproduction des protéines recombinantes……71

II. Surproduction et purification des protéines d'intérêt……72
1. Culture bactérienne et induction……72
2. Lyse bactérienne……73
3. Purification de la protéine d'intérêt……73
3.1. Chromatographie d'affinité pour le nickel……74
3.2. Chromatographie d'affinité pour la streptavidine……75
3.3. Chromatographie d'exclusion de taille……76
4. Dosage des protéines……77
5. Concentration des protéines……77
6. Echange de tampons par dessalage……78
7. Clivage par la protéase du TEV……78

III. Analyse qualitative et quantitative des protéines d'intérêt……79
1. Techniques biochimiques……79
1.1. Electrophorèse discontinue en conditions dénaturantes……79
1.2. *Western-blot*……80
1.3. Pontage covalent……81
1.3.1. Pontage au BS^3……81
1.3.2. Pontage au GTA……82
2. Techniques biophysiques……82
2.1. Spectrométrie de masse……82
2.2. Diffusion dynamique de la lumière……82
2.3. SEC-MALS……83
2.4. Dichroïsme circulaire……84

IV. Etude structurale des protéines d'intérêt……86
1. Cristallographie par diffraction aux rayons X……86
1.1. Cristallogenèse……86
1.1.1. Méthode de diffusion de vapeur……86
1.1.2. Technique de la goutte assise……87
1.1.3. Technique de la goutte suspendue……87
1.2. Diffraction aux rayons X……88
2. Diffusion des rayons X aux petits angles……89
3. Modélisation par homologie de séquence……91

V. Mesures d'interactions……92
1. *Histone array*……92
2. Criblage d'inhibiteurs par intégration concertée *in vitro*……93

Résultats et discussion……95

I. CCD isolés de RAV-1 et VIH-1……96
1. CCD sauvage et H103C de l'IN de RAV-1……96
1.1. Surproduction et purification……96
1.2. Essais de solubilisation……98
2. CCD mutés de l'IN de VIH-1……99
2.1. Surproduction et purification……99
2.2. Essais de solubilisation……100

II. Etude de l'IN de RAV-1 sauvage entière……102
1. Surproduction et purification……102
2. Clivage de l'étiquette de purification……104
3. Analyse par diffusion dynamique de la lumière (DLS)……105
4. Essais de cristallisation……105
5. Criblage d'inhibiteurs anti IN RAV-1……106
5.1. Conception des molécules……106
5.2. Solubilisation des composés……109
5.3. Intégration concertée *in vitro*……109
5.3.1. Optimisation de la concentration d'IN de RAV-1……109
5.3.2. Tests d'intégration concertée en présence des composés potentiellement inhibiteurs……110

III. Modélisation par homologie de l'intasome de PERV-A/C……112

IV. Etude moléculaire et structurale de l'IN de PERV-A/C en complexe avec le cofacteur cellulaire humain Brd2……114
1. Etude du CTD isolé de l'IN de PERV-A/C……115
1.1. Surproduction et purification……115
1.2. Clivage de l'étiquette de purification……117
1.3. Essais de cristallisation……118
1.4. Criblage d'interactions avec des histones par *histone array*……118
2. Etude du CTD isolé de Brd2……120
2.1. Surproduction et purification……120
2.2. Clivage de l'étiquette poly-histidine……121
3. Formation du complexe protéique avec les CTD isolés……122
4. Stratégie « multiTEV »……123
4.1. Surproduction et purification……123
4.2. Identification du complexe protéique par *western-blot*……126
4.3. Caractérisation du complexe IN_{CTD}-$Brd2_{CTD}$ en solution……126
4.3.1. Etat oligomérique en solution……126
4.3.1.1. Chromatographie d'exclusion de taille (SEC)……126
4.3.1.2. Pontage covalent au BS^3 et au GTA……127
4.3.1.3. SEC-MALS……128
4.3.2. Etude structurale du complexe IN_{CTD}-$Brd2_{CTD}$……130
4.3.2.1. Dichroïsme circulaire……130
4.3.2.2. Essais de cristallisation……130

4.3.2.3. Diffusion des rayons X aux petits angles (SAXS).131

Conclusion et perspectives……....134

Publication……..140

Porcine endogenous retrovirus-A/C: biochemical properties of its integrase and susceptibility to raltegravir……...140

Références bibliographiques……148

Abréviations

A_{260}/A_{280}	rapport d'absorbance à 260 nm et 280 nm
AI	*AutoIntegration*
ALLINI	*ALLosteric Integrase Inhibitors*
APOBEC3G	*APOlipoprotein B mRNA editing Enzyme, Catalytic polypeptide-like 3G*
APS	*Ammonium PerSulfate*
ARNt	ARN de transfert
ARV	AntiRétroViraux
ASLV	*Avian Sarcoma and Leukemia Virus*
AZT	3'-AZidoThymidine
BAF	*Barrier to Autointegration Factor*
BCIP	5-Bromo-4-Chloro-3-Indolyl Phosphate
BD	BromoDomaine
BET	*Bromodomain and Extra Terminal domain*
BLAST	*Basic Local Alignment Search Tool*
Brd2	*Bromodomain containing protein 2*
BS^3	Bis(SulfoSuccinimidyl) Suberate
CA	CApside
CAPS	acide 3-(cyclohexylamino)-1-propanesulfonique
CCD	domaine central catalytique
CCR5	*C-C Chemokine Receptor type 5*
CD	*Circular Dichroism*
CD4	*Cluster Differenciation 4*
CHAPS	3-[(3-cholamidopropyl)diméthylammonio]-1-propanesulfonate
CI	*Cleaved Intasome*
CTD	domaine carboxy-terminal CTD
CTF	*CAAT binding Transcription Factors*
CXCR4	*C-X-C Chemokine Receptor type 4*
ddNTP	didésoxyribonucléotides
DLS	*Dynamic Light Scattering*
D_{max}	Distance interatomique maximale
DMSO	DiMéthySulfOxyde
dNTP	désoxyribonucléotides
DTT	DiThioThréitol
EBM	*ET Binding Motif*
E. coli	*Escherichia coli*
EDTA	*EthyleneDiamine Tetraacetic Acid*
EED	*Embryonic Ectoderm Development*
EMBL	*European Molecular Biology Laboratory*
env	glycoprotéine d'enveloppe
ESRF	*European Synchrotron Radiation Facility*
FDA	*Food and Drug Administration*
FPLC	*Fast Protein Liquid Chromatography*
FSI	*Full-Site Integration*

gag	*group specific antigen*
GST	Glutathion-S-Transférase
GTA	Glutaraldéhyde
HAART	*Highly Active AntiRetroviral Therapy*
HABA	acide benzoïque 2-(4-hydroxy-azobenzène)
HAT	Histone AcétylTransférase
HERV	*Human Endogenous RetroViruses*
HMG	*High Mobility Group*
HMGA1	*High Mobility Group chromosomal protein*
HRP	*Hepatoma derived growth factor Related Proteins*
HIS	*Half-Site Integration*
HTLV	*Human T-cell Leukemia/Lymphoma Virus*
IBD	*Integrase Binding Domain*
ICTV	*International Commitee on Taxonomy of Viruses*
IMAC	*Immobilized Metal Ion Affinity Chromatography*
IN	Intégrase
INI1	*INtegrase Interactor 1*
INSTI	*INtegrase Strand Transfer Inhibitors*
IPTG	IsoPropylThioGalactoside
LAV	*Lymphoadenopathy Associated Virus*
LB	Luria-Bertani
LEDGF/p75	*Lens Epithelium-Derived Growth Factor*
LTR	*Long Terminal Repeat*
m/z	rapport masse/charge
MA	MAtrice
MALDI-TOF	*Matrix Assisted Laser Desorbtion Ionisation-Time Of Flight*
MALS	*Multi-Angle Light Scattering*
MES	acide 2-[*N*-morpholino]-éthane-sulfonique
MLV	*Murine Leukemia Virus*
MM	Marqueurs de masses moléculaires connues
MMTV	*Mouse Mammary Tumor Virus*
NBT	NitroBlue Tetrazolium
NC	NucléoCapside
NCBI	*National Center for Biotechnology Information*
NED	*N-terminal Extension Domain*
Nef	*Negative regulatory factor*
NHS	N-HydroxySulfosuccinimide
NLS	*Nuclear Localization Sequence*
NNRTI	*Non Nucleosides and nucleotides Reverse Transcriptase Inhibitors*
NRTI	*Nucleosides and nucleotides Reverse Transcriptase Inhibitors*
NTA	*NitriloTriacetic Acid*
NTD	domaine amino-terminal
OMS	Organisation Mondiale de la Santé
PAGE	*PolyAcrylamide Gel Electrophoresis*
PCR	*Polymerase Chain Reaction*
PDB	*Protein Data Bank*

PdI	coefficient de polydispersité
PEG	PolyEthylène Glycol
PERV	*Porcine Endogenous RetroVirus*
PES	PolyEtherSulfone
PFV	*Prototype Foamy Virus*
PIC	*Pre Integration Complex*
pol	*polymérase*
PR	PRotéase
RAV	*Rous Associated Virus*
Rev	*Regulator of expression of virion proteins*
R_g	Rayon de giration
R_h	Rayon hydrodynamique
RING3	*Really Interesting New Gene 3*
RMN	Résonance Magnétique Nucléaire
RMSD	*Root Mean Square Deviation*
RNase H	RiboNucléase H
RSV	*Rous Sarcoma Virus*
RT	*Reverse Transcriptase*
RTC	*Reverse Transcriptase Complex*
SAXS	*Small Angle X-ray Scattering*
SDS	*Sodium Dodecyl Sulfate*
SEC	*Size Exclusion Chromatography*
SIDA	Syndrome de l'ImmunoDéficience Acquise
STC	*Strand Transfer Complex*
SU	Glycoprotéines de surface
SWI/SNF	*SWItch/Sucrose Non-Fermentable*
Tat	*Transactivator of transcription*
TB	*Terrific Broth*
TBE	Tris-Borate-EDTA
TCC	*Target Capture Complex*
TEMED	N,N,N',N'-tétraméthyl-éthylènediamine
TEV	*Tobacco Etch Virus*
TFIID	*Transcription Factor II D*
TM	glycoprotéines transmembranaires
U3	*Unique sequence at 3' end of genome*
U5	*Unique sequence at 5' end of genome*
UI	*Uncleaved intasome*
Vif	*Virion infectivity factor*
VIH	Virus de l'Immunodéficience Humaine
v/v	volume/volume
w/v	poids/volume
w/w	poids/poids

Curriculum Vitae

Halima YAJJOU-HAMALIAN

29 ans

Adresse professionnelle :

Institut de Biologie et Chimie des Protéines IBCP- CNRS FR 3302
« *Molecular Microbiology and Structural Biochemistry* » MMSB
CNRS UMR 5086 - Université Lyon 1 UCBL1
UMS3444 BioSciences Gerland - Lyon Sud
Équipe « Rétrovirus et Biochimie Structurale »
7, Passage du Vercors – 69367 LYON cedex 07
Tél. : 04.72.72.26.28 – Fax : 04.72.72.26.16 – email : halima.yajjou@ibcp.fr

Adresse personnelle :

Résidence « Le Rubis », 2 place du traité de Rome 69007 Lyon
Tél. : 06.23.54.53.02 – email : halima.hamalian@gmail.com

Cursus

2014-2017 :

❖ **Thèse de Doctorat en Biochimie, financement ANRS**

Lyon, France

IBCP, CNRS/UCBL1-« Rétrovirus et Biochimie Structurale » (UMR 5086) et « Rétrovirus, intégration rétrovirale et vaccins » (UMR 754)

Projet de recherche : **« Etudes moléculaires et structurales d'un nouveau mode de dimérisation des intégrases rétrovirales pour développer des modulateurs d'oligomérisation anti VIH-1 & de l'intégrase porcine PERV-A/C en complexe avec son cofacteur cellulaire humain Brd2 dans un contexte de xénotransplantation »**

Directeurs de thèse : Pr. Patrice GOUET (patrice.gouet@ibcp.fr)
Dr Corinne RONFORT (corinne.ronfort@inra.fr)

2013 :

(Octobre-Décembre)

❖ **Assistant ingénieur**

Lyon, France

INRA, UCBL1

Activité de recherche : **« Inhibition de l'étape d'intégration par des siRNA en utilisant la technique de RT-qPCR »**

Directrice : Dr Corinne RONFORT (corinne.ronfort@inra.fr)

2012-2013 :

(Octobre - Juillet)

❖ **Stage de fin d'étude**

Lyon, France

IBCP, CNRS/UCBL1-« Rétrovirus et Biochimie Structurale » (UMR 5086) et « Rétrovirus, intégration rétrovirale et vaccins » (UMR 754)

Sujet de recherche : **« Etude cristallographique du domaine de liaison de la protéine humaine Brd2 à l'intégrase du rétrovirus endogène porcin PERV »**

Responsable de stage : Pr. Patrice GOUET (patrice.gouet@ibcp.fr)

2011-2013 :

- **Master Recherche Biochimie Structurale et Fonctionnelle, mention bien**

Université Claude Bernard Lyon 1, France

2008-2011 :

- **Licence Biochimie, mention bien**

Université Claude Bernard Lyon 1, France

Participation à des communications orales et/ou posters

2016 : (Octobre)

- **21ème Journée de l'Ecole Doctorale Interdisciplinaire Sciences-Santé EDISS**

Campus de la Doua, Villeurbanne – 13 octobre 2016

Poster : « Caractérisation du complexe protéique formé par l'intégrase rétrovirale porcine et son cofacteur cellulaire humain Brd2 »

Yajjou, H., Chahpazoff, M., Gallay, K., Ronfort, C., Gouet, P.

Lauréate du prix des meilleurs posters

2016 : (Avril)

- **Séminaire interne IBCP**

IBCP Gerland – 4 avril 2016

Exposé oral : « Structural study of the porcine integrase in complex with its human cellular cofactor Brd2 »

Yajjou, H., Chahpazoff, M., Gallay, K., Ronfort, C., Gouet, P.

2016 : (Février)

- **Journée Scientifique Sidaction 2016**

Mairie de Paris – 12 février 2016

Poster : « Etude d'un nouvel assemblage dimérique des intégrases rétrovirales et recherche d'inhibiteurs anti VIH-1 »

Yajjou, H., Merceron, R., Robert, X., Gallay, K., Ronfort, C., Gouet, P.

2015 : (Octobre)

- **Université des Jeunes Chercheurs Sidaction 2015,**

Carry-le-Rouet – 10-16 octobre 2015

Poster : « Etude moléculaire et structurale d'un nouveau mode d'association des intégrases rétrovirales et recherche d'inhibiteurs anti VIH-1 »

Yajjou, H., Merceron, R., Robert, X., Gallay, K., Ronfort, C., Gouet, P.

2015 :

(Juillet)

❖ **Séminaire INRA**

INRA Lyon – 9 juillet 2015

Exposé oral : « Etude structurale de l'interaction entre la protéine humaine Brd2 et l'intégrase du rétrovirus endogène porcin PERV-A/C » Yajjou, H., Chahpazoff, M., Gallay, K., Ronfort, C., Gouet, P.

2014 :

(Novembre)

❖ **Congrès « New strategies for macromolecular complexes analysis**

Institut de Génétique et Biologie Moléculaire et Cellulaire IGBMC, Illkirch, France – 24-28 novembre 2014

Poster : « Rational research of oligomerization modulators against retroviral integrases »

Yajjou, H., Merceron, R., Robert, X., Gallay, K., Ronfort, C., Gouet, P.

Publications

2015 :

(Avril)

❖ **Article en co-premier auteur, *Journal of General Virology* :**

Titre : Porcine Endogenous Retrovirus (PERV)-A/C: biochemical properties of its integrase and susceptibility to Raltegravir.

Liste des auteurs : Antonin Demange; Halima Yajjou-Hamalian ; Kathy Gallay; Catherine Luengo; Véronique Beven; Aurélie Leroux; Marie-Pierre Confort; Elsy Al Andary; Patrice Gouet; Karen Moreau; Corinne Ronfort; Yannick Blanchard.

Introduction bibliographique

I. Rétrovirus

1. Généralités

Les rétrovirus sont des virus enveloppés de la famille des *Retroviridae* qui infectent principalement les vertébrés. Il s'agit de particules virales globulaires avec un diamètre moyen de 100 nm (Kurth *et al.*, 2010). Ils possèdent un génome composé de deux brins d'ARN monocaténaire à polarité positive. Ils sont constitués de protéines structurales et enzymatiques. Certains rétrovirus peuvent également contenir des protéines virales additionnelles (protéines régulatrices et accessoires).

Lors du cycle réplicatif des rétrovirus, l'ARN rétroviral est converti en un ADN bicaténaire. Cette réaction est catalysée par une enzyme virale de type ADN polymérase ARN-dépendante, la transcriptase inverse (ou rétrotranscriptase). L'ADN rétroviral nouvellement synthétisé peut ensuite être intégré dans le génome de la cellule infectée, au cours de l'étape d'intégration catalysée par une autre enzyme virale : l'intégrase.

En 1908, la première description d'une maladie causée par un rétrovirus fait suite aux travaux menés par deux chercheurs danois sur des leucémies chez le poulet (Weiss *et al.*, 2011). À travers cette étude, les auteurs ont montré que des ultrafiltrats cellulaires de poulets malades pouvaient transmettre la leucémie à des poulets sains. Malgré cette découverte, il a fallu attendre les travaux de Peyton Rous pour démontrer l'origine virale de ces maladies néoplasiques. Ce dernier mit également en évidence l'existence de l'agent infectieux responsable, le *Rous Sarcoma Virus* ou RSV (Rous, 1911). La conversion de son génome ARN en un ADN viral a été découverte en 1970 (Temin *et al.*, 1970). Les rétrovirus porteurs d'oncogènes tel que RSV, responsables de cancers et de leucémies, sont appelés oncorétrovirus.

Historiquement, le premier rétrovirus humain à avoir été isolé fut, en 1980, le rétrovirus T-lymphotropique humain de type I ou HTLV-1 (*Human T-cell Leukemia/Lymphoma Virus type I*). Deux maladies sont associées à ce rétrovirus : la leucémie lymphoïde T de l'adulte et la paraparésie spastique tropicale (Poiesz *et al.*, 1980). Le HTLV-2, qui présente une séquence proche du HTLV-1 mais qui est moins pathogène, fut découvert deux ans plus tard (Kalyanaraman *et al.*, 1982).

En 1981, certains membres de la communauté homosexuelle aux États-Unis, touchés par le Syndrome de l'ImmunoDéficience Acquise ou SIDA, développaient des maladies

opportunistes et des cancers. En 1983, un rétrovirus fut isolé à partir d'une biopsie ganglionnaire d'un patient atteint de lymphoadénopathie (stade pré-SIDA). Découvert par Luc Montagnier, Jean-Claude Chermann et Françoise Barré-Sinoussi de l'Institut Pasteur de Paris, il fut nommé virus associé à une lymphoadénopathie (*Lymphoadenopathy Associated Virus* ou LAV, Barre-Sinoussi *et al.*, 1983). Ce rétrovirus avait la capacité de cibler et de se répliquer dans des cellules du système immunitaire, les lymphocytes T CD4+. Les études menées sur des sérums de patients atteints ont permis de définir ce virus comme étant l'agent infectieux causatif du SIDA. En 1986, le terme LAV est remplacé par Virus de l'Immunodéficience Humaine de type I ou VIH-1. La même année, un nouveau type nommé VIH-2 fut isolé à son tour par l'équipe de L. Montagnier (Clavel *et al.*, 1986). Les virus VIH-1 et VIH-2 sont les seuls rétrovirus découverts chez l'homme et appartenant au genre *Lentivirus*.

Avec plus de 36 millions de personnes infectées à ce jour et 1,2 millions de décès dans le monde en 2014, le SIDA représente un problème majeur de santé publique (Organisation Mondiale de la Santé OMS, 2015).

2. Taxonomie

Les virus ont initialement été classés en fonction de la nature de leur génome, selon la classification générale des virus définie par David Baltimore (Baltimore, 1971). La famille des *Retroviridae* appartenait au groupe VI (ARN associé à une activité transcriptase inverse). De par la découverte de nouveaux rétrovirus, il a fallu modifier cette taxonomie en utilisant des critères de classification différents.

Ce nouveau système universel de classification a été établi par le Comité International de Taxonomie des Virus (*International Commitee on Taxonomy of Viruses* ou ICTV, King *et al.*, 2011) et est basé sur la proximité des séquences du gène viral *pol*. L'ICTV actualise régulièrement la taxonomie des rétrovirus et la rend disponible sur le Web (http://www.ictvonline.org/virustaxonomy.asp).

Au sein de cette classification, la famille des *Retroviridae* est divisée en deux sous-familles : les *Spumaretroviridae* et les *Orthoretroviridae*. La sous-famille des *Orthoretroviridae* regroupe six genres : les *Alpharetrovirus*, *Betaretrovirus*, *Deltaretrovirus*, *Epsilonretrovirus* et *Gammaretrovirus* sont dotés d'un potentiel oncogénique et peuvent induire des tumeurs ; tandis que les *Lentivirus* sont non oncogènes mais capables d'induire des maladies

chroniques à évolution lente. Le genre *Spumaretrovirus*, appartenant à la sous-famille des *Spumaretroviridae*, est retrouvé dans différentes espèces animales, mais son implication pathologique n'est pas connue à ce jour (Figure 1).

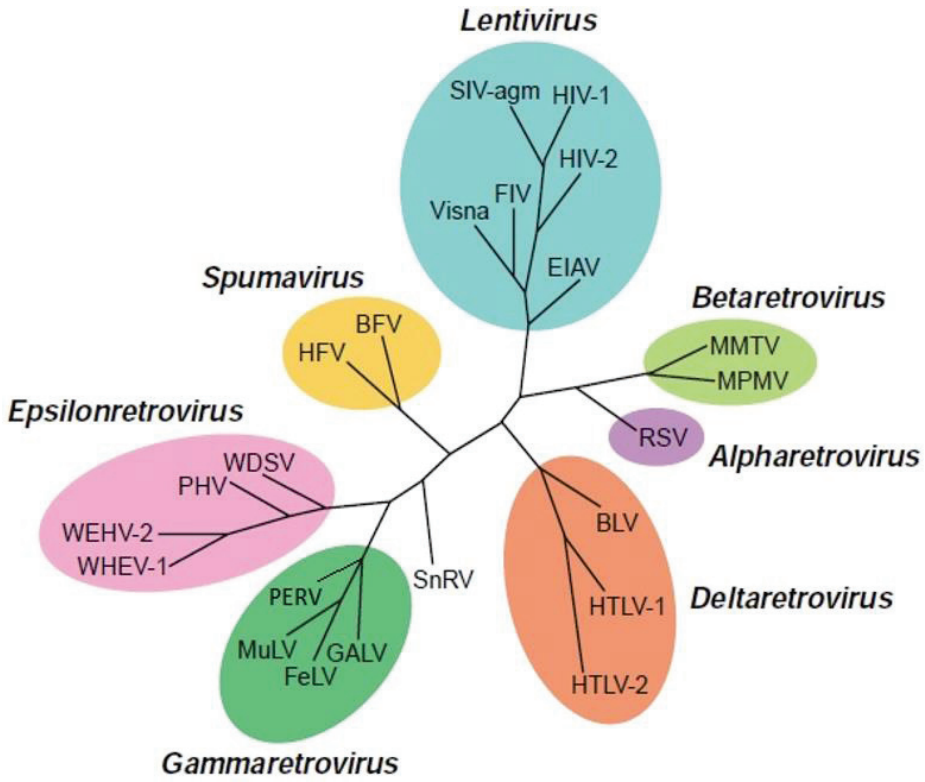


Figure 1. Taxonomie des rétrovirus basée sur l'analyse phylogénétique des régions conservées du gène viral *pol* (d'après King *et al.*, 2011).

La complexité du génome rétroviral peut également servir de critère de classification des rétrovirus. Lorsque ce génome ne contient que les gènes codant les protéines enzymatiques et structurales, le rétrovirus est dit « simple » (exemple des *Avian Sarcoma and Leukemia Virus* ou ASLV). En plus de ces gènes, s'il porte les gènes codant des protéines régulatrices et accessoires, le rétrovirus est qualifié de « complexe » (cas de VIH-1) (Coffin *et al.*, 1997).

Les rétrovirus peuvent également être divisés en fonction de leur caractère endogène ou exogène. Les agents infectieux ciblant les vertébrés sont généralement des rétrovirus exogènes. Il s'agit de particules infectieuses extracellulaires, dont la transmission peut être horizontale (d'un organisme à l'autre) ou verticale (transmission mère-enfant lors de l'allaitement ou de l'accouchement). Lorsqu'un rétrovirus exogène parvient à infecter une cellule de la lignée germinale, il devient partie intégrante du génome de l'hôte : il s'agit du processus d'endogénisation. Le rétrovirus endogène est ainsi transmis à la descendance selon les lois mendéliennes (transmission verticale) (Figure 2).

Pour la majeure partie des rétrovirus endogènes, le processus d'endogénisation a eu lieu il y a plusieurs millions d'années. Le *Koala endogenous Retrovirus* (KoRV) est un exemple de rétrovirus en cours d'endogénisation. Il est capable d'infecter des koalas au nord-est de l'Australie (forme exogène) et d'entraîner ainsi l'apparition de lymphomes et de leucémies. Il est également retrouvé dans la lignée germinale de certains koalas (forme endogène) (Tarlinton *et al.*, 2006).

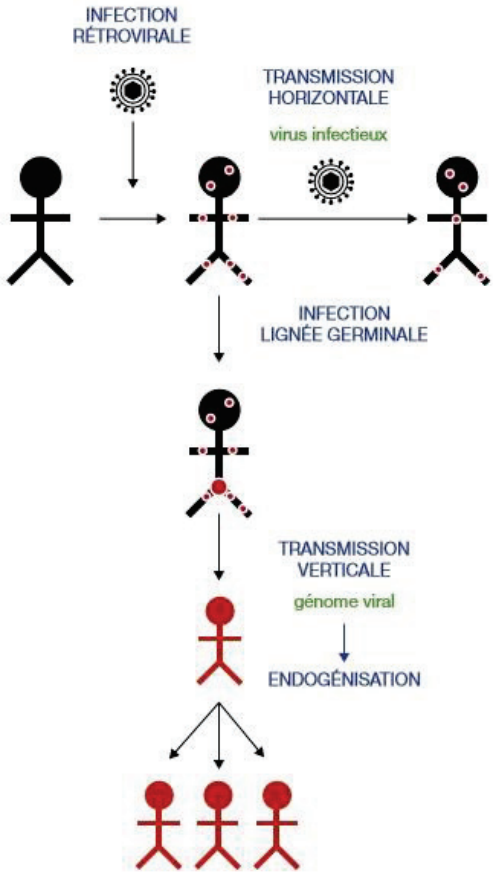


Figure 2. Principe d'endogénisation des rétrovirus (d'après Dewannieux *et al.*, 2010).

Découverts en 1967 chez le poulet sain, les rétrovirus endogènes n'ont pas à générer de nouvelles particules infectieuses pour se multiplier. Chez l'homme, ces vestiges d'anciennes infections nommés *Human Endogenous RetroViruses* ou HERV, représentent 8% du génome (Dewannieux *et al.*, 2010).

La plupart des rétrovirus endogènes ont subi des mutations ponctuelles, de délétions ou de recombinaisons au niveau génomique (par rapport au rétrovirus exogène dont ils descendent), qui les ont rendus défectifs pour la réplication (Weiss, 2006). En revanche, certains rétrovirus endogènes ont conservé leur phase ouverte de lecture et pourraient participer à des processus

tant physiologiques (exemple de la syncytine, *cf.* § suivant) que pathologiques (diabète, sclérose, cancer) (Balada *et al.*, 2010, Ruprecht *et al.*, 2008).

La famille des rétrovirus endogènes humains HERV-W code une protéine d'enveloppe nommée syncytine (Blond *et al.*, 2000). Cette protéine d'origine virale joue un rôle primordial dans la formation du syncytiotrophoblaste, un tissu organique du placenta qui assure les échanges entre la mère et le fœtus. Chez le modèle murin, l'absence de syncytine perturbe la structure de l'interface materno-fœtale et conduit à la mort de l'embryon *in utero* (Mi *et al.*, 2000).

Chez le porc, le rétrovirus endogène porcin (*Porcine Endogenous RetroVirus* ou PERV), du genre *Gammaretrovirus,* a conservé sa capacité de réplication mais aucune pathologie ne lui a été associée à ce jour. Cependant, une transmission de ce virus du porc à l'homme a été mise en évidence dans des cellules *in vitro* (Patience *et al.*, 1997). Cette découverte soulève le risque de transmission de PERV à l'homme lors d'une xénotransplantation. Le terme xénotransplantation définit la transplantation de tissus, d'organes ou de cellules de l'animal à l'homme. Cette procédure représente une alternative face au déficit mondial croissant de donneurs d'organes. De par son élevage aisé et maîtrisé, la taille et le poids de ses organes, ainsi que sa compatibilité physiologique, le porc représente l'animal de choix pour les xénogreffes (Cooper *et al.*, 2002, Sachs, 1994). Cependant, par le biais de telles transplantations, il ne peut être exclu que le rétrovirus PERV puisse engendrer l'apparition de pathologies diverses chez l'homme, soit en l'infectant, soit en recombinant avec des rétrovirus résidents (endogènes ou exogènes).

3. Organisation génomique et structurale

1. Le génome rétroviral

Le génome des rétrovirus est composé de deux brins d'ARN monocaténaires homologues de 7 à 12 kb, à polarité positive et liés entre eux de façon non covalente (King *et al.*, 2011). Il porte différentes modifications post-traductionnelles : une coiffe à l'extrémité 5', certaines adénosines méthylées et un signal de polyadénylation à l'extrémité 3'. Comme pour l'ARN messager eucaryote, la coiffe est impliquée dans l'initiation de la traduction ; alors que la queue polyadénylée protègerait l'ADN viral intégré de la dégradation (Hawkins, 1996).

Le génome rétroviral est composé de deux régions non codantes présentes à chaque extrémité nommées *Long Terminal Repeat* ou LTR. La région codante, située entre ces deux régions non codantes, est composée au minimum de trois gènes de structure communs à tous les rétrovirus : les gènes *gag* (*group specific antigen*), *pol* (polymérase) et *env* (glycoprotéine d'enveloppe). Contrairement aux rétrovirus simples (comme les virus de type ASLV), le génome des rétrovirus complexes possède des gènes supplémentaires qui codent des protéines régulatrices et accessoires.

1. Les LTR

L'ADN viral intégré ou « provirus » est encadré par deux séquences LTR identiques issues de l'étape de transcription inverse. Ces séquences jouent un rôle essentiel dans l'intégration et dans l'expression des gènes proviraux. Les LTR sont divisées en trois parties : une région U3 suivie d'une région R puis d'une région U5 (Figure 3).

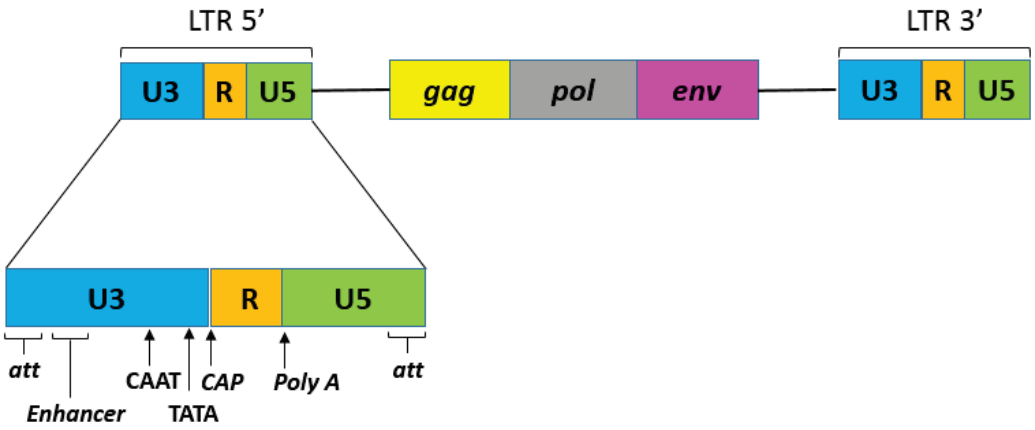


Figure 3. Organisation générale du génome proviral.

La région U3 (*Unique sequence at 3' end of genome*) présente à son extrémité 5' un site d'attachement *att* reconnu spécifiquement par l'intégrase. Elle contient aussi des séquences promotrices dont une séquence activatrice appelée *enhancer*, porteuse de sites de fixation de facteurs de transcription, capable d'agir sur les séquences virales et sur les séquences de l'hôte. Selon les gènes cellulaires activés, la cellule cible peut développer un état tumorigène. Dans la région U3, en amont du site d'initiation de la transcription, nous retrouvons également les boîtes CAAT et TATA (Figure 3). Il s'agit de séquences promotrices qui régulent le niveau d'expression des gènes viraux en interagissant avec des facteurs transcriptionnels de la cellule hôte. La boîte CAAT lie les CTF (*CAAT binding Transcription Factors*) tandis que la boîte TATA interagit avec le facteur TFIID (*Transcription Factor II D*). Ces facteurs de transcription

servent à la mise en place de l'ARN polymérase de type II sur l'ADN à transcrire (Drapkin *et al.*, 1993, Jones *et al.*, 1987).

La région R pour *Repeat*, la plus courte des trois régions, joue un rôle dans l'étape de transcription inverse. Elle contient les sites *CAP* et *PolyA* permettant respectivement l'ajout de la coiffe en 5' et du signal de polyadénylation en 3' de l'ARN messager viral. La transcription démarre dès la première base de la région R tandis que l'adénylation débute juste après sa dernière base.

La région U5 (*Unique sequence at 5' end of genome*) contient, comme la région U3, la séquence *att* reconnue spécifiquement par l'intégrase. Elle est, elle aussi, impliquée dans l'étape de transcription inverse. En aval de son extrémité 3' se fixe l'ARN de transfert couplé au résidu lysyl ou lysyl-ARNt, pour amorcer la réaction de transcription inverse.

2. Les gènes rétroviraux

Après traduction des ARN messagers viraux, deux précurseurs polyprotéiques nommés Gag et Gag-Pol sont produits. Dans 95% des cas, le codon stop en 3' du gène *gag* est correctement lu par le ribosome, ce qui provoque la production du précurseur Gag seul. Dans 5% des cas, un décalage du cadre de lecture a lieu pour VIH-1, ce qui engendre la production du précurseur Gag-Pol (Carter *et al.*, 2007, Jacks *et al.*, 1987). Ces deux polyprotéines Gag et Gag-Pol sont ensuite clivées par la protéase virale.

Le précurseur Gag clivé donne lieu à trois protéines matures : les protéines de MAtrice (MA), de CApside (CA) et de NucléoCapside (NC) qui sont impliquées dans l'organisation structurale du virus. La protéine MA tapisse la face interne de l'enveloppe virale et forme la matrice. Son signal de localisation nucléaire, ou NLS (*Nuclear Localization Sequence*), lui permet de participer à l'adressage de l'ADN viral au sein du complexe de pré-intégration ou PIC (*Pre Integration Complex*), complexe formé de protéines virales et cellulaires (Freed, 2001). La protéine MA interagit avec la membrane cellulaire et participe également à l'assemblage de la particule virale. La protéine CA est elle aussi impliquée dans l'étape d'assemblage. Constituant majeur de la capside virale, elle forme un cœur capsidique conique au centre du virus lors de la maturation des *Lentivirus* (Sundquist *et al.*, 2012). Enfin, la protéine NC est dotée d'une activité chaperonne vis-à-vis de l'ARN viral. Elle permet la formation de

conformations stables d'acides nucléiques, essentielles à la réplication virale (Levin *et al.*, 2010). Elle participe également à la dimérisation et à l'encapsidation du génome rétroviral.

Après clivage, le précurseur Gag-Pol permet d'obtenir les enzymes virales suivantes : la transcriptase inverse (*Reverse Transcriptase* ou RT) et l'intégrase (IN). Elles catalysent respectivement les réactions de transcription inverse subit par l'ARN viral et d'intégration de l'ADN viral rétro-transcrit. Contrairement à ces enzymes, la protéase (PR) est codée par le gène *gag* ou *pol*, ceci en fonction du type rétroviral (*gag* pour les ASLV, *pol* pour le VIH-1).

À la différence des précurseurs Gag-Pol et Pol qui sont clivés par la protéase virale, le précurseur Env est clivé par une protéase cellulaire après avoir été glycosylé. Il permet de générer deux types de glycoprotéines d'enveloppe, les glycoprotéines transmembranaires (TM) ou de surface (SU), qui sont adressées à la surface de la particule virale. Elles sont impliquées dans la reconnaissance et la fixation sur un récepteur spécifique à la surface de la cellule cible, puis dans l'entrée du virus dans la cellule résultant de la fusion des membranes virale et cellulaire.

3. Cas des rétrovirus complexes

Le génome des rétrovirus complexes possèdent, en plus des gènes communs à tous les rétrovirus, les gènes codant les protéines régulatrices et accessoires. C'est le cas du génome de VIH-1 qui contient six gènes supplémentaires, dont deux qui codent les protéines régulatrices Tat (*Transactivator of transcription*) et Rev (*Regulator of expression of virion proteins*) (Figure 4).

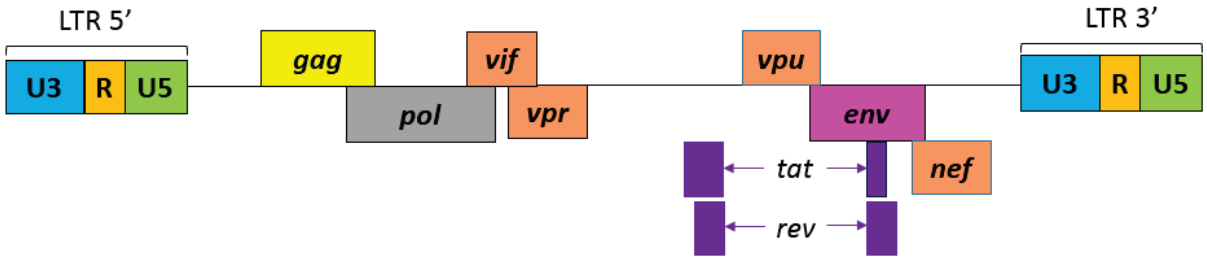


Figure 4. Structure génomique d'un rétrovirus complexe, le *Lentivirus* VIH-1.

La protéine Tat recrute l'ARN polymérase cellulaire, elle représente un puissant transactivateur de la transcription. La protéine Rev permet l'export des ARN messagers viraux, qui codent les protéines de structure, du noyau vers le cytoplasme.

Les quatre protéines accessoires retrouvées dans le virus VIH-1 sont les protéines Vpr, Nef, Vif et Vpu. La protéine Vpr est un activateur faible de la transcription qui participe au transport du PIC vers le noyau. La protéine Nef (*Negative regulatory factor*) augmente l'infectiosité et la réplication du virus et participe ainsi à la persistance de l'infection. La protéine Vif (*Virion infectivity factor*) sert à la propagation du virus en se fixant à l'enzyme cellulaire APOBEC3G (*APOlipoprotein B mRNA editing Enzyme, Catalytic polypeptide-like 3G*). APOBEC3G est un facteur de restriction de la réplication rétrovirale, grâce à son activité de catalyse de la désamination de cytidine en uridine. La fixation de Vif à APOBEC3 provoque la dégradation de cette dernière, qui ne peut plus bloquer le cycle de réplication rétroviral (Feng *et al.*, 2014). La protéine Vpu, quant à elle, assure un assemblage viral efficace et facilite le bourgeonnement des particules virales nouvellement formées.

4. Cas des rétrovirus oncogènes

Les oncovirus, virus et rétrovirus au potentiel oncogénique, peuvent induire des tumeurs variées et sont responsables de 22% des cancers humains (OMS, 2015). Un rétrovirus oncogène peut transformer la cellule qu'il infecte par l'un des trois mécanismes suivants : 1/ il peut s'intégrer à proximité d'un proto-oncogène cellulaire et déréguler son expression (exemple du rétrovirus *Murine Leukemia Virus* ou MLV, Hayward *et al.*, 1981); 2/ il peut également avoir intégré au sein de son génome un oncogène cellulaire (cas de RSV). Enfin, 3/ l'une des protéines rétrovirales peut altérer des facteurs cellulaires impliqués dans la réparation, l'apoptose ou le contrôle de la division cellulaire (cas de HTLV-1) (Assil *et al.*, 2011, Azran *et al.*, 2004).

Certains oncorétrovirus possèdent d'autres gènes viraux souvent similaires à des gènes cellulaires, en particulier à des proto-oncogènes. Ces gènes supplémentaires peuvent parfois induire des modifications des gènes *gag, pol* ou *env* par insertion voire par leur remplacement. Ces virus peuvent donc présenter des délétions dans ces gènes et devenir défectifs pour la réplication. C'est le cas de certaines souches de RSV qui ne possèdent pas de gène *env* et qui ne peuvent se répliquer qu'en présence de virus associés appelés RAV (*Rous Associated Virus*) comportant les trois gènes intacts (Martin, 2004). En infectant l'espère aviaire, les oncorétrovirus RSV et RAV provoquent respectivement des sarcomes (*Avian Sarcoma Virus* ou ASV) et des leucémies (*Avian Leukemia Virus* ou ALV). L'ensemble de ces oncorétrovirus infectant les espèces aviaires est regroupé sous le terme ASLV.

2. La morphologie de la particule virale

L'émergence de la technique de microscopie électronique a permis l'étude de la morphologie des particules virales. Les rétrovirus sont des particules sphériques enveloppées d'une centaine de nanomètres de diamètre. Après bourgeonnement, les particules virales présentent une enveloppe lipidique issue de la membrane plasmique de la dernière cellule infectée. Cette enveloppe est recouverte de spicules que sont les glycoprotéines TM et SU. La face interne de cette bicouche lipidique est tapissée d'une couche de protéines MA, qui renferme une deuxième enveloppe protéique constituée de protéines CA. Sous cette enveloppe CA se trouvent les protéines virales enzymatiques (PR, RT et IN) ainsi que le génome ARN viral diploïde condensé et protégé par la NC (Figure 5).

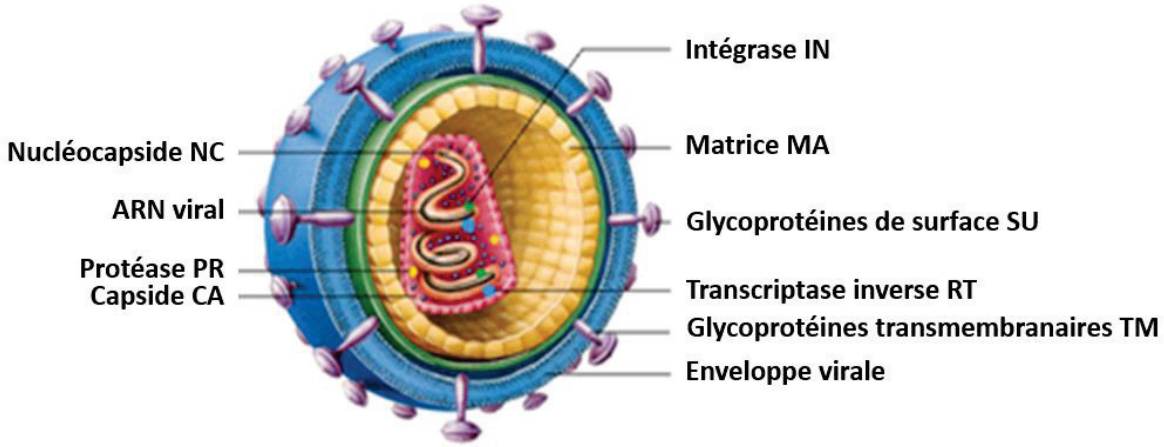


Figure 5. Morphologie d'une particule rétrovirale (d'après Freed, 2001).

4. Cycle réplicatif rétroviral

Le cycle réplicatif est nécessaire à la multiplication du rétrovirus dans les cellules vivantes. Il s'étend de l'entrée du virus dans la cellule hôte à la maturation du virion et est divisé en deux phases principales : la phase pré-intégrative et la phase post-intégrative (Figure 6).

1. La phase pré-intégrative

La phase pré-intégrative comprend les étapes suivantes : l'entrée du virus, la réaction de transcription inverse, l'import nucléaire et l'intégration du génome viral dans le génome cellulaire.

L'entrée du virus dans la cellule hôte nécessite en premier lieu l'interaction spécifique entre la glycoprotéine d'enveloppe virale SU et un récepteur membranaire de la cellule cible. Cette interaction à haute affinité induit un changement de conformation des protéines d'enveloppe,

ce qui amorce la fusion des membranes virale et cellulaire *via* la sous-unité TM. Dans le cas de VIH-1, le récepteur cellulaire appartient à la famille des immunoglobulines et s'appelle C*luster Differenciation 4* (CD4). Il interagit avec les glycoprotéines de l'enveloppe virale. Celles-ci interagissent également avec le co-récepteur *C-C Chemokine Receptor type 5* (CCR5) ou *C-X-C Chemokine Receptor type 4* (CXCR4) pour entraîner la fusion membranaire. Ce phénomène de fusion conduit à la libération de la capside virale dans le cytoplasme de la cellule hôte (Wilen *et al.*, 2012).

La réaction de transcription inverse, catalysée par la RT, permet l'obtention d'un ADN viral double brin à partir de la matrice ARN simple brin. Cette enzyme est contenue dans un complexe nucléoprotéique formé de deux molécules d'ARN viral, de protéines virales et cellulaires, appelé *Reverse Transcriptase Complex* ou RTC. La RT ne possède pas d'activité correctrice et peut introduire de nombreuses mutations en incorporant une base incorrecte lors de la transcription inverse. Dans le cas de VIH-1, la RT est responsable d'une variabilité génétique importante. Ce manque de fidélité explique en partie l'émergence de résistances aux traitements antirétroviraux et l'échappement du rétrovirus au système immunitaire de l'hôte (Le Grice, 2012).

L'ADN viral bicaténaire issu de l'étape de transcription inverse est transporté au noyau de la cellule hôte au sein du PIC. Dans le cas des *Lentivirus* capables d'infecter des cellules quiescentes et en division, l'import nucléaire du PIC se fait par transport actif. À l'inverse, pour les rétrovirus qui infectent uniquement les cellules en division (cas des G*ammaretrovirus*), le PIC entre dans le noyau seulement s'il y a cassure de la membrane nucléaire (Yamashita *et al.*, 2006). Au sein du noyau, l'ADN viral est intégré au génome de la cellule hôte par l'enzyme IN.

2. La phase post-intégrative

Lors de la phase post-intégrative, le provirus est transcrit de la jonction U3-R du LTR-5' à la jonction R-U5 du LTR-3' par l'ARN polymérase de type II cellulaire. Une partie des ARN messagers néosynthétisés sert de génome aux futures particules virales. L'autre partie sert de matrice pour la traduction des protéines structurales et enzymatiques et éventuellement des protéines régulatrices et accessoires dans le cas des rétrovirus complexes. Les précurseurs Gag et Gag-Pol sont adressés et s'assemblent au niveau de la membrane cellulaire, puis ils recrutent les protéines d'enveloppe.

Deux molécules d'ARN génomique sont encapsidées lors de cette étape d'assemblage. Pour la plupart des rétrovirus, l'assemblage et le bourgeonnement sont réalisés simultanément. Le bourgeonnement conduit au relargage des virions non infectieux.

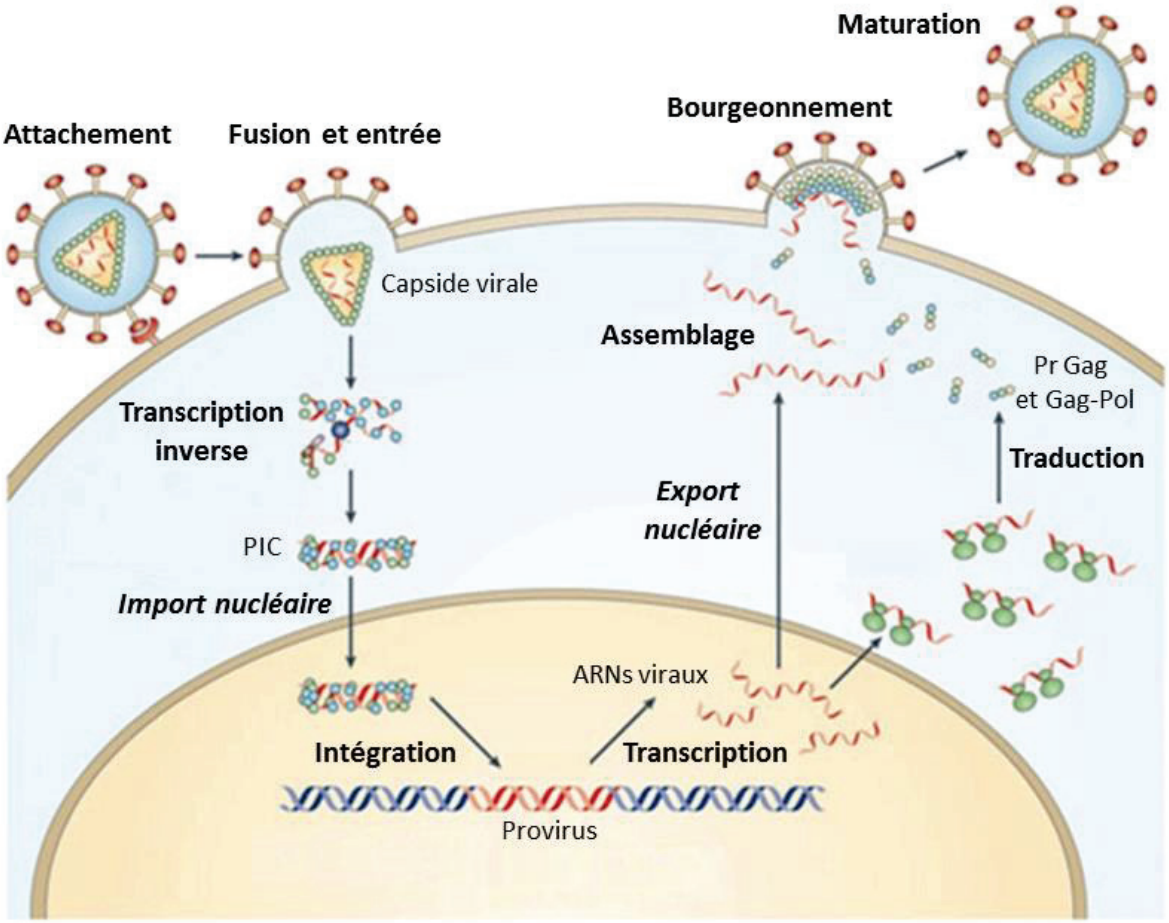


Figure 6. Schéma du cycle réplicatif des rétrovirus (Pr : précurseur) (d'après Nisole *et al.*, 2005).

Dans la majorité des cas, la particule virale nouvellement synthétisée subit une étape de maturation, à savoir le clivage des précurseurs Gag et Gag-Pol par la PR, ce qui rend le virion infectieux (Lee *et al.*, 2012). Cette étape entraîne un changement structural de la CA pour VIH-1 : initialement sphérique, elle devient conique. Des images de microscopie électronique révèlent que les particules immatures sont hétérogènes en taille et en forme et que le précurseur Gag est ancré à la membrane virale. Le cœur du virion immature est clair, alors que le virion mature apparaît condensé avec un contour clair et un cœur dense aux électrons (Figure 7).

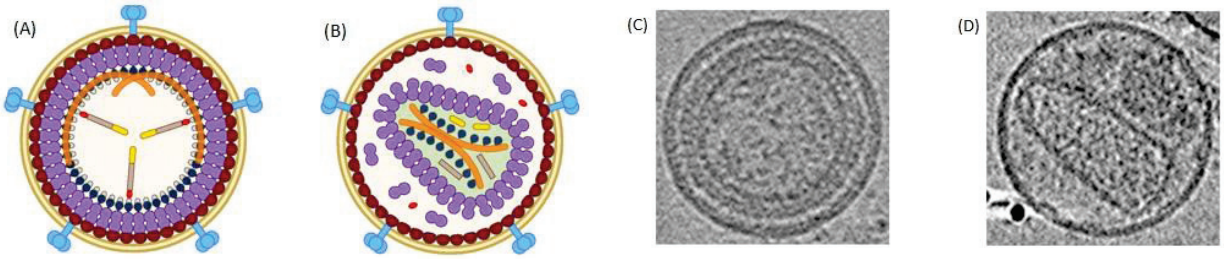


Figure 7. Schématisation et observation des particules virales non infectieuse (A, C) et infectieuse (D) de VIH-1 par microscopie électronique (d'après Balasubramaniam *et al.*, 2011).

3. Inhibiteurs antirétroviraux

Les infections rétrovirales peuvent entraîner l'apparition de pathologies sévères conduisant parfois à la mort de l'individu. Plus de 30 ans après la découverte du virus VIH-1, le SIDA reste un fléau responsable de la mort de 5700 personnes chaque jour dans le monde (données ONUSIDA). Pour tenter d'enrayer cette épidémie, différentes stratégies utilisant des composés AntiRétroViraux ou ARV ont été proposées. Cependant, des difficultés majeures sont rencontrées face au mécanisme d'action du *Lentivirus* VIH-1. Tout d'abord, le rétrovirus est intégré de façon stable dans le génome de la cellule infectée. Ainsi, il peut être transmis aux cellules filles lors de la mitose. Le VIH-1 a la particularité de pouvoir infecter des cellules du système immunitaire (macrophages, cellules dendritiques, lymphocytes T CD4) dont la destruction aboutira progressivement à une immunodéficience, favorable à l'apparition de maladies opportunistes. Les ARV n'ont pas accès à toutes ces cellules dispersées dans l'organisme. Certaines restent dans un état latent mais pourraient relâcher des particules virales à tout moment. Elles constituent ainsi un réservoir viral et forment une barrière à l'éradication de VIH-1. En plus de la transmission virale aux cellules filles et de l'existence de réservoirs viraux, il existe des variants génétiques nombreux. Issus du manque de fidélité de la RT, ils facilitent l'émergence et la persistance de résistances face aux ARV chez les patients infectés (Metifiot *et al.*, 2013).

Différentes stratégies sont actuellement en cours de développement pour bloquer l'infection rétrovirale. L'une d'entre elles consiste à concevoir des composés chimiques qui inhibent certaines enzymes virales et/ou étapes du cycle réplicatif des rétrovirus. Ces composés peuvent être combinés efficacement dans le cadre d'une multithérapie antirétrovirale hautement active ou HAART (*Highly Active AntiRetroviral Therapy*) (Arts *et al.*, 2012).

1. Ciblage de la fusion et de l'entrée

L'entrée du virus dans la cellule est une étape clé dans le cycle réplicatif rétroviral. Ce processus implique non seulement les protéines de surface du virus (SU et TM), mais également les récepteurs membranaires portés par la cellule cible.

Les inhibiteurs développés peuvent donc agir au niveau de la surface rétrovirale ou cellulaire. Ils agissent en bloquant la protéine SU ou en interagissant avec un co-récepteur. Approuvé en 2007 par la FDA (*Food and Drug Administration*), le selzentry ou maraviroc® est un inhibiteur chimique de l'entrée. Dans le cas de VIH-1, cette molécule antagoniste se lie au co-récepteur CCR5 présent à la surface des cellules immunitaires et empêche la pénétration du virus (Lieberman-Blum *et al.*, 2008).

D'autres inhibiteurs peptidiques sont en cours de développement et visent notamment à interagir avec la protéine SU pour empêcher au virus d'entrer dans la cellule.

2. Ciblage de la transcriptase inverse

Initialement utilisée comme traitement anti-cancer, la 3'-AZidoThymidine ou AZT a été la première molécule active sur VIH-1 et le premier inhibiteur de la RT (Wright, 1986). Cet analogue de la thymidine, commercialisé sous le nom de Zidovudine®, est utilisé dans le cadre d'une monothérapie. Bien qu'il inhibe la RT, cet antirétroviral présente des effets secondaires multiples et son efficacité s'est avérée insuffisante.

Les inhibiteurs de la RT représentent 54% des inhibiteurs approuvés par la FDA. Ils constituent ainsi la majorité des traitements antirétroviraux actuels. Ils sont classés en deux catégories : les inhibiteurs analogues et non analogues de nucléotides ou de nucléosides dont les abréviations respectives sont NRTI (*Nucleosides and nucleotides Reverse Transcriptase Inhibitors*) et NNRTI (*Non Nucleosides and nucleotides Reverse Transcriptase Inhibitors*). Les NRTI, dont l'AZT, sont des désoxynucléotides qui s'incorporent dans le brin ADN en cours de synthèse et qui bloquent son élongation. Le mécanisme d'action des NNRTI est différent puisqu'ils se fixent directement sur la RT en un site distinct du site actif de l'enzyme (De Clercq *et al.*, 2016, Greene *et al.*, 2008).

3. Ciblage de la protéase

Le développement de drogues anti-protéase est basé sur la connaissance de la structure de cette enzyme virale. La PR de VIH-1 est constituée de deux monomères qui, en s'assemblant, forment un tunnel abritant le site catalytique de l'enzyme (Navia *et al.*, 1989). Les inhibiteurs développés ont été conçus pour cibler cette poche catalytique et se lier à elle avec une haute affinité. Ils empêchent ainsi la fixation des substrats naturels de l'enzyme et inhibent son activité (Lv *et al.*, 2015, Wlodawer *et al.*, 1993).

Actuellement, une dizaine d'inhibiteurs de PR a été approuvée par la FDA pour une utilisation clinique (De Clercq *et al*., 2016). Toutefois, des mutations de résistance ont été observées dans la séquence de l'enzyme. Elles engendrent un élargissement du site actif entraînant une diminution de l'affinité entre la protéase et ses inhibiteurs.

4. Ciblage de l'intégrase

L'enzyme IN représente une cible thérapeutique de choix dans la lutte contre les infections rétrovirales. Inhiber l'activité de l'IN permettrait d'empêcher l'intégration de l'ADN viral dans l'ADN cellulaire et ainsi d'éviter la propagation du virus. L'IN ne possède pas d'homologue cellulaire direct chez l'homme, ce qui réduit considérablement les effets secondaires de ses inhibiteurs.

Trois inhibiteurs de l'IN sont actuellement autorisés pour traiter les patients. Il s'agit du raltégravir, du dolutégravir et de l'elvitégravir respectivement connus sous les noms commerciaux Isentress®, Tivicay® et Vitekta®. Le raltégravir et l'elvitégravir sont des dérivés dicétoniques, qui lient les ions magnésium du site actif. Ces molécules, appelées INSTI (*INtegrase Strand Transfer Inhibitors*), empêchent la réaction de transfert de brin de l'ADN rétroviral sur l'ADN cible (*cf.* § II.2.1). Cependant, comme avec les autres classes d'antirétroviraux, une résistance aux inhibiteurs de l'IN a rapidement été mise en évidence (Metifiot *et al.*, 2013). Elle est due à l'émergence, sous traitement anti-IN, de variants viraux portant des mutations de résistance sur le gène codant l'IN (Koh *et al.*, 2011). Le dolutégravir est un INSTI de nouvelle génération présentant une meilleure barrière génétique et actuellement utilisé en traitement HAART (Brenner *et al.*, 2016, Wensing *et al.*, 2014).

II. Intégrase

1. Complexe de pré-intégration

L'IN est une enzyme essentielle du cycle rétroviral et une cible thérapeutique privilégiée. Dans le cytoplasme de la cellule hôte, après transcription inverse, elle est associée à l'ADN viral rétrotranscrit pour former l'intasome au sein du PIC, qui comprend également d'autres protéines virales et cellulaires (Figure 8) (Farnet *et al.*, 1990).

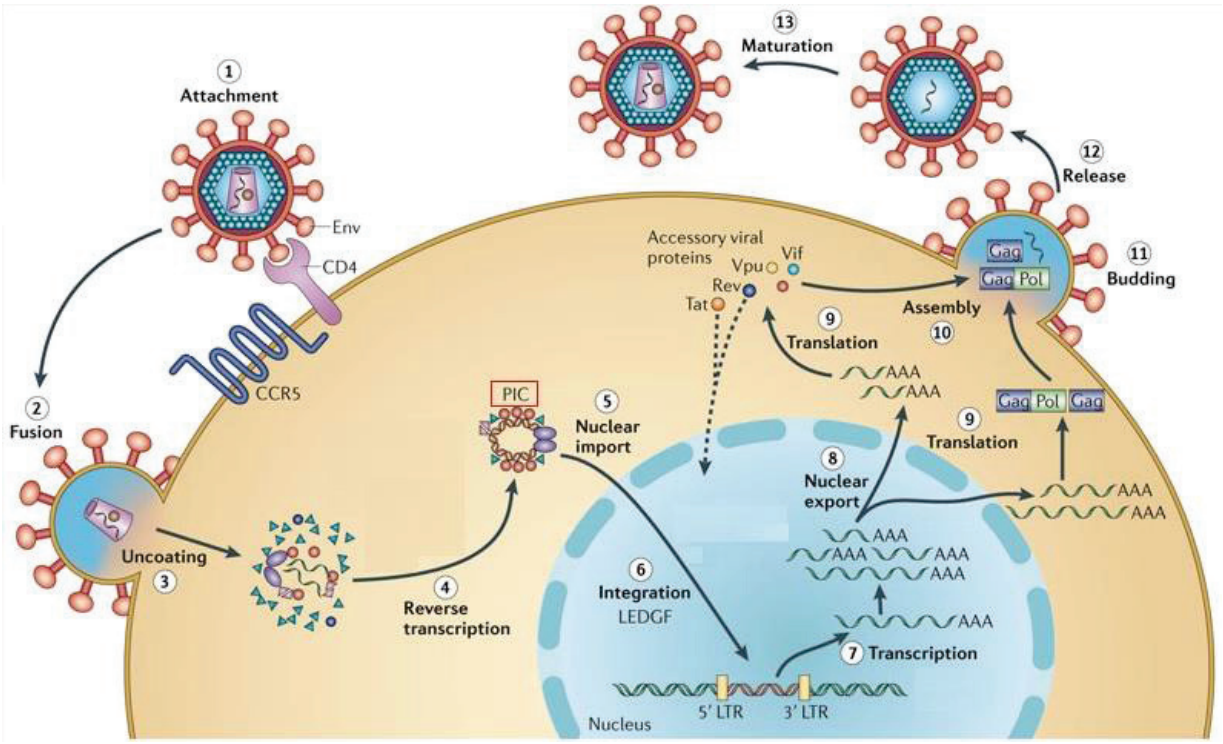


Figure 8. Formation et entrée du PIC dans le noyau lors de la réplication rétrovirale (Engelman *et al.*, 2012).

Dans le cas particulier de VIH-1, le PIC contient l'ADN viral rétrotranscrit linéaire, les protéines Vpr, MA, NC, CA, RT et IN ainsi que les partenaires cellulaires LEDGF/p75 (*Lens Epithelium-Derived Growth Factor*) dont l'IBD (*Integrase Binding Domain*), INI1 (*Integrase Interactor 1*), BAF (*Barrier to Autointegration Factor*) et HMGA1 (*High Mobility Group chromosomal protein A1*) (Bukrinsky, 2004, Krishnan *et al.*, 2012, Maillot *et al.*, 2013, Van Maele *et al.*, 2006) (Figure 9).

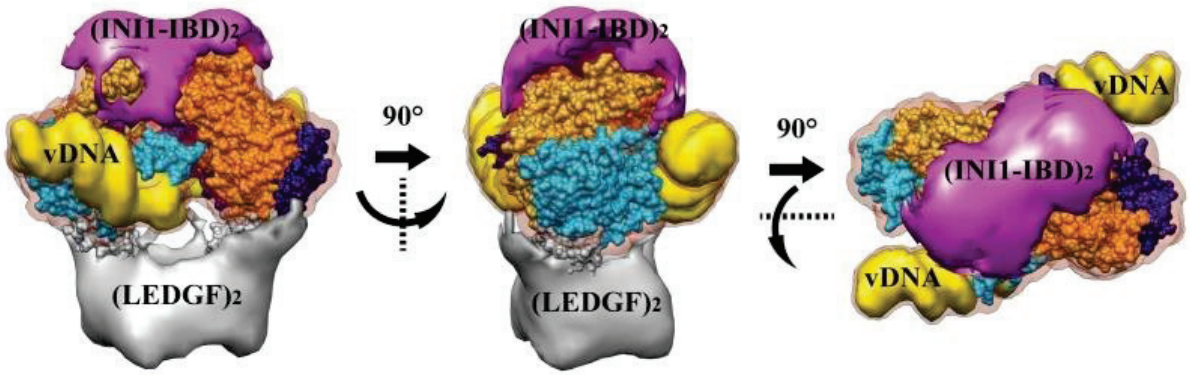


Figure 9. Reconstruction 3D par cryo-microscopie électronique du complexe IN/LEDGF/INI1-IBD/ADN viral (Maillot *et al.*, 2013).

La protéine cellulaire BAF empêche l'auto-intégration de l'ADN viral (Lee *et al.*, 1998). La protéine HMGA1 de la famille des protéines HMG (*High Mobility Group*) se lie préférentiellement aux régions riches en A/T dans l'ADN cellulaire. Elle peut ainsi moduler la structure de la chromatine et réguler la transcription (Bustin *et al.*, 1996).

D'autres protéines du PIC régulent positivement ou négativement l'activité de l'intégrase. *In vitro,* la RT est capable d'inhiber l'activité catalytique de l'IN (Oz *et al.*, 2002). La protéine INI1 est le premier cofacteur cellulaire de l'IN de VIH-1 à avoir été caractérisé (Kalpana *et al.*, 1994). Ce partenaire a un rôle très controversé dans l'étape d'intégration. *In vitro*, il active l'activité de transfert de brins (Kalpana *et al.*, 1994), mais il peut aussi exercer une action antirétrovirale en inhibant l'activité enzymatique de l'IN (Maillot *et al.*, 2013). L'IN pourrait être réactivée après transfert dans le compartiment nucléaire, ou lorsque le PIC interagit avec la chromatine. À l'inverse, d'autres protéines du PIC ont la capacité de stimuler l'activité enzymatique de l'IN *in vitro*. C'est le cas de la protéine virale NC et du partenaire cellulaire LEDGF/p75 (Carteau *et al.*, 1999, Cherepanov, 2007).

Dans le cas des *Lentivirus* qui infectent des cellules quiescentes, le PIC peut franchir la membrane nucléaire en passant par les pores nucléaires. Les NLS portés par les protéines IN, MA et Vpr permettent d'assurer ce ciblage nucléaire. Ces signaux sont reconnus par les protéines cellulaires importines et servent au transport actif de molécules du cytoplasme au noyau (Bukrinsky, 2004, Suzuki *et al.*, 2007). Une étude menée sur les NLS portés par la protéine MA a permis de mettre en évidence le rôle primordial de ces signaux, puisque leur mutation engendre un défaut de réplication de VIH-1 dans les macrophages (Haffar *et al.*, 2000).

Un des éléments impliqués dans l'adressage du PIC au noyau est le *DNA Flap central*. Il s'agit d'une séquence triple brin présente au centre de l'ADN viral et formée lors de la transcription inverse. L'absence de cette structure provoque une accumulation du génome viral à proximité de l'enveloppe nucléaire (Zennou *et al.*, 2000).

Après import nucléaire du PIC, les rétrovirus intègrent leur génome en des sites préférentiels du génome cellulaire. Chez les *Lentivirus*, dont VIH-1, l'intégration a lieu préférentiellement dans des régions codantes de gènes transcriptionnellement actifs (Schroder *et al.*, 2002). Dans le cas des *Gammaretrovirus* (comme MLV et PERV) et des *Spumaretrovirus* (comme PFV), l'ADN viral s'intègre dans l'ADN cellulaire à proximité des sites d'initiation de la transcription et au niveau des îlots CpG (Moalic *et al.*, 2006, Scherdin *et al.*, 1990, Trobridge *et al.*, 2006). Les génomes des *Alpha-*, *Beta-* et *Deltaretrovirus* semblent s'intégrer de manière quasi-aléatoire, même si les *Alpha-* et *Deltaretrovirus* ont une faible préférence pour les régions fortement transcrites (Mitchell *et al.*, 2004) (Figure 10).

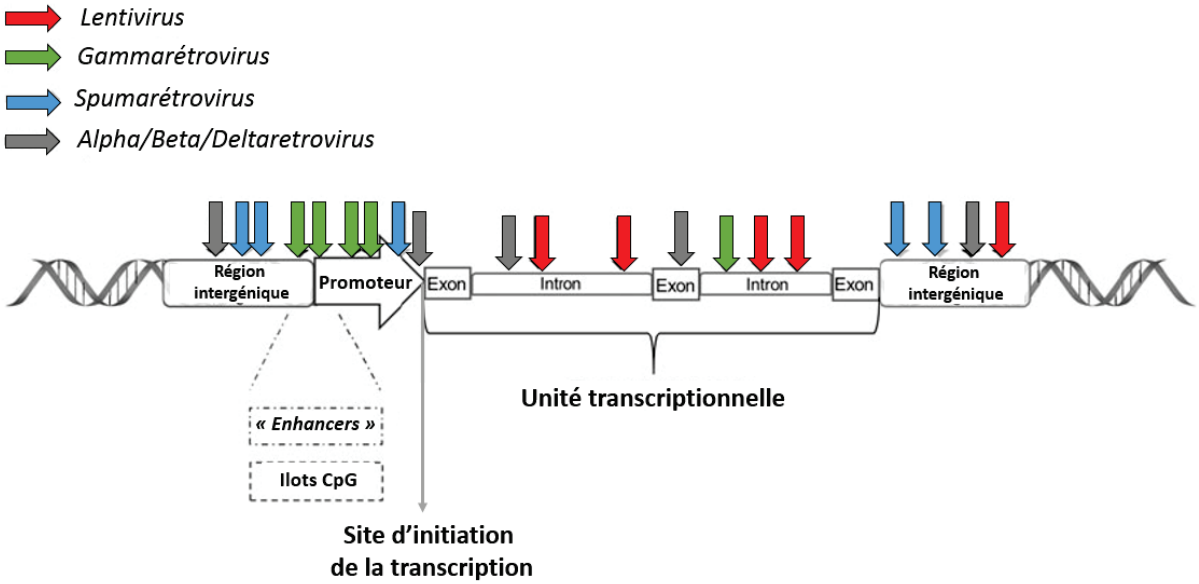


Figure 10. Sites d'intégration préférentiels de certains rétrovirus dans le génome de la cellule hôte (d'après Serrao *et al.*, 2016).

Certains constituants du PIC, principalement des cofacteurs cellulaires de l'IN, sont impliqués dans le ciblage des sites préférentiels d'intégration. Dans le cas de VIH-1, la protéine nucléaire Emerin associée au PIC *via* BAF permet l'interaction entre l'ADN viral et la chromatine (Jacque *et al.*, 2006). Le partenaire cellulaire LEDGF/p75 se lie à l'IN de VIH-1 et cible en particulier les régions fortement transcrites (Engelman *et al.*, 2008). L'absence du cofacteur LEDGF/p75 rend la réplication rétrovirale inefficace (Maertens *et al.*, 2003).

À l'inverse, l'ajout de LEDGF/p75 dans une réaction d'intégration *in vitro* favorise l'interaction entre l'IN et l'ADN cible et stimule l'intégration (Cherepanov, 2007). En plus des protéines Emerin, BAF et LEDGF/p75, le cofacteur cellulaire INI1 participe à la stabilisation du PIC et empêche l'auto-intégration de l'ADN viral (Maillot *et al.*, 2013). INI1 appartient au complexe SWI/SNF (*SWItch/Sucrose Non-Fermentable*). Ce complexe, très conservé chez les eucaryotes, est impliqué dans le remodelage de la chromatine (Lesbats *et al.*, 2011). Le partenaire INI1 se lie à l'IN et facilite son adressage et sa liaison au génome cellulaire. Les protéines de la famille HMG jouerait également un rôle dans le ciblage des sites d'intégration préférentiels puisqu'elles sont capables de stimuler l'intégration de l'ADN viral *in vitro* (Gao *et al.*, 2003, Hindmarsh *et al.*, 1999).

Les composants protéiques du PIC servent au transport dans le noyau et au ciblage des sites préférentiels d'intégration situés sur l'ADN cellulaire. Ils agissent aussi après l'étape d'intégration, lors de la réparation de l'ADN. En effet, après l'intégration du génome viral dans le génome cellulaire, des brèches apparaissent de part et d'autre du provirus. Il est nécessaire qu'elles soient réparées pour assurer la continuité entre l'ADN viral et l'ADN cellulaire. Différentes protéines cellulaires du PIC ont une fonction de réparation, dont les facteurs de type RAD qui ciblent les cassures double brin. Les études menées ont permis de mettre en évidence une interaction entre l'IN et les protéines humaines RAD18 et RAD51 impliquées dans la réparation de l'ADN (Desfarges *et al.*, 2006, Mulder *et al.*, 2002).

2. Aspects fonctionnels

Le rôle principal de l'IN est de catalyser l'intégration d'un génome substrat dans celui de la cellule hôte. Son site actif est constitué d'une triade catalytique appelée « motif D, D(35)E », car il est composé de deux résidus d'aspartate (D) et d'un résidu glutamate (E), séparé du second résidu aspartyl par 35 résidus d'acides aminés. Cette triade catalytique est conservée parmi les IN rétrovirales, de rétrotransposons et les transposases (Jaskolski *et al.*, 2009, Kulkosky *et al.*, 1992).

1. Intégrase et transposases

Les transposons et les rétrotransposons sont des séquences ADN mobiles qui, *via* l'activité d'une transposase, peuvent s'insérer dans une séquence ADN cible. La différence entre ces deux types d'éléments repose sur le processus d'intégration utilisé.

Pour les transposons, le processus est dit « direct » : l'ADN est excisé par une transposase à activité nucléase et ligase, avant d'être directement inséré dans l'ADN cible. À l'inverse, les rétrotransposons passent par un intermédiaire ARN : ils sont transcrits en ARN, puis retranscrits en ADN pour s'intégrer à l'ADN cible (Figure 11). Les rétrotransposons et plus précisément les rétrotransposons à LTR, ont un mécanisme d'intégration similaire à celui des rétrovirus. Ces rétrotransposons pourraient être les ancêtres des rétrovirus (Eickbush *et al.*, 2002). Cependant, les rétrotransposons ont un cycle totalement intracellulaire et, contrairement aux rétrovirus, ils ne peuvent être transmis aux cellules voisines (Cox *et al.*, 2011).

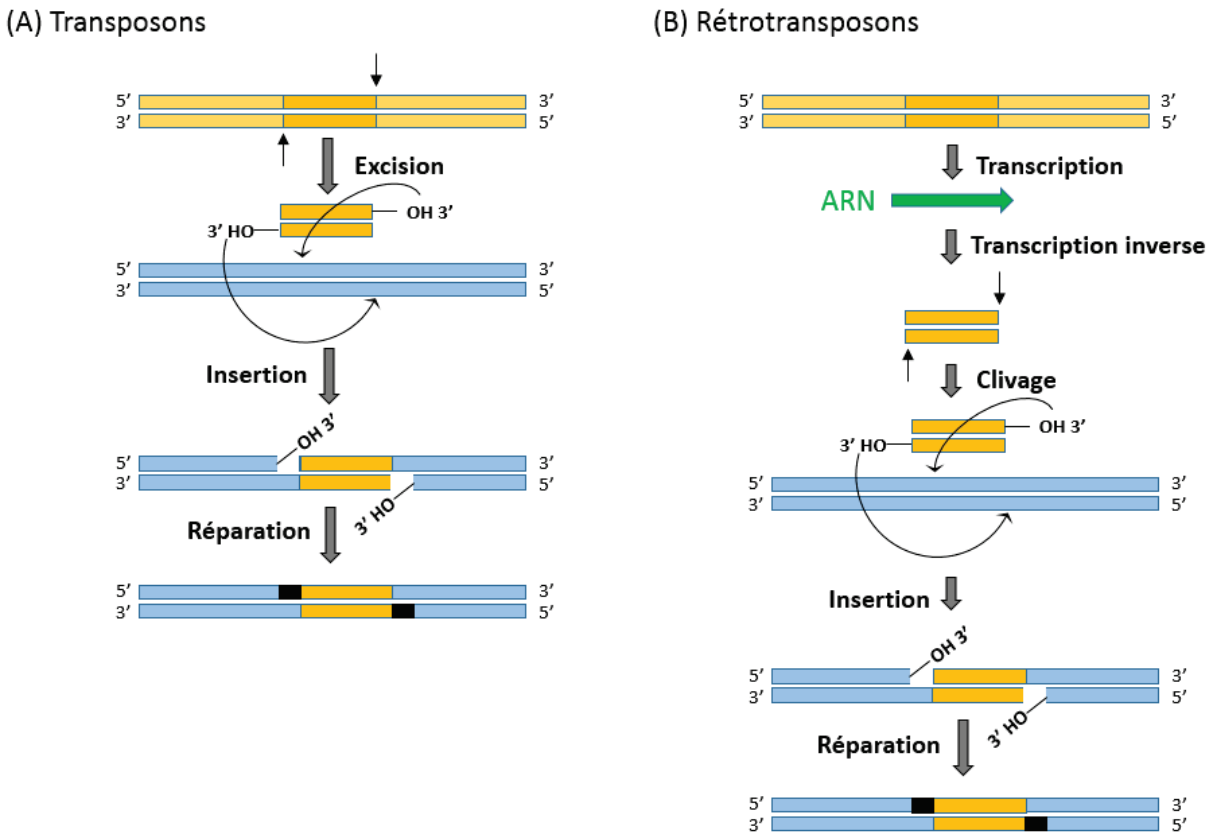


Figure 11. Mécanisme d'intégration des transposons (A) et des rétrotransposons (B) (d'après Cox *et al.*, 2011).

La protéine IN appartient au groupe des transposases DDE, puisqu'elle contient la triade catalytique DDE et un repliement de type RiboNucléase H ou RNase H au niveau du domaine catalytique (Kulkosky *et al.*, 1992). Cette classe de transposases est également retrouvée chez les bactériophages Tn5, Tn10 et Mu (Rice *et al.*, 2001). Bien qu'elles appartiennent au même groupe, elles présentent des différences dans les processus d'excision et d'insertion du transposon dans l'ADN cible.

2. Intégration concertée chez les rétrovirus

L'étape d'intégration rétrovirale est dite « concertée » car les deux extrémités de l'ADN viral s'insèrent en un même site de l'ADN cellulaire. Cette réaction est divisée en trois étapes successives : le clivage des extrémités ADN ou *3'-processing*, le transfert de brins ou *strand transfer* et la réparation de l'ADN. Les deux premières étapes sont catalysées par l'IN tandis que la dernière étape fait intervenir des enzymes de réparation cellulaires (Semenova *et al.*, 2008).

L'ADN viral est délimité à chaque extrémité par les LTR, contenant les séquences *att* aux extrémités des régions U3 et U5. Lors du *3'-processing* qui a lieu dans le cytoplasme, ces séquences d'attachement sont reconnues spécifiquement par l'IN qui se fixe à eux. En présence de cations divalents métalliques liés de type Mg^{2+} (probablement physiologiques) ou Mn^{2+} (Andrake *et al.*, 2009, Grobler *et al.*, 2002), l'IN catalyse une attaque nucléophile d'une molécule d'eau sur le squelette phosphodiester de l'ADN viral (Krishnan *et al.*, 2012) (Figure 12). Cette attaque aboutit au clivage des deux nucléotides terminaux des séquences inversées répétées (TT pour ASLV, GC pour VIH-1) sur les deux brins. Cette réaction permet la création d'extrémités 3' hydroxylées (ou 3'-OH) se terminant chez tous les rétrovirus par un dinucléotide CA (Brown *et al.*, 1989, Sherman *et al.*, 1990). L'ADN viral résultant est transporté au noyau au sein du PIC et sert de substrat pour la réaction de transfert de brins.

La deuxième étape, dite de transfert de brins, est également catalysée par l'IN et, comme lors du *3'-processing*, requiert la présence d'ions Mg^{2+} ou Mn^{2+}. Il s'agit d'une seconde attaque nucléophile des extrémités 3'-OH sur deux liaisons phosphodiesters distantes de l'ADN cible. Cette réaction a lieu simultanément et en un même site pour chaque extrémité de l'ADN viral (Engelman *et al.*, 1991). Les brins d'ADN cellulaire sont clivés avec un décalage de quatre à six paires de bases selon le type de rétrovirus (4 pb pour PERV, 5 pb pour VIH-1, 6 pb pour ASLV) (Figure 12). Cette réaction aboutit à une structure intermédiaire dans laquelle l'extrémité virale est liée à l'ADN cellulaire par un seul brin.

La troisième et dernière étape de l'intégration est la réparation de l'ADN cellulaire, catalysée par des enzymes de réparation cellulaires. Au cours de cette étape, une nucléase cellulaire clive les deux nucléotides terminaux en 5' des deux brins du génome viral. Les brèches formées au niveau de l'ADN cellulaire sont comblées par duplication d'une courte séquence chromosomique de part et d'autre du provirus (Figure 12).

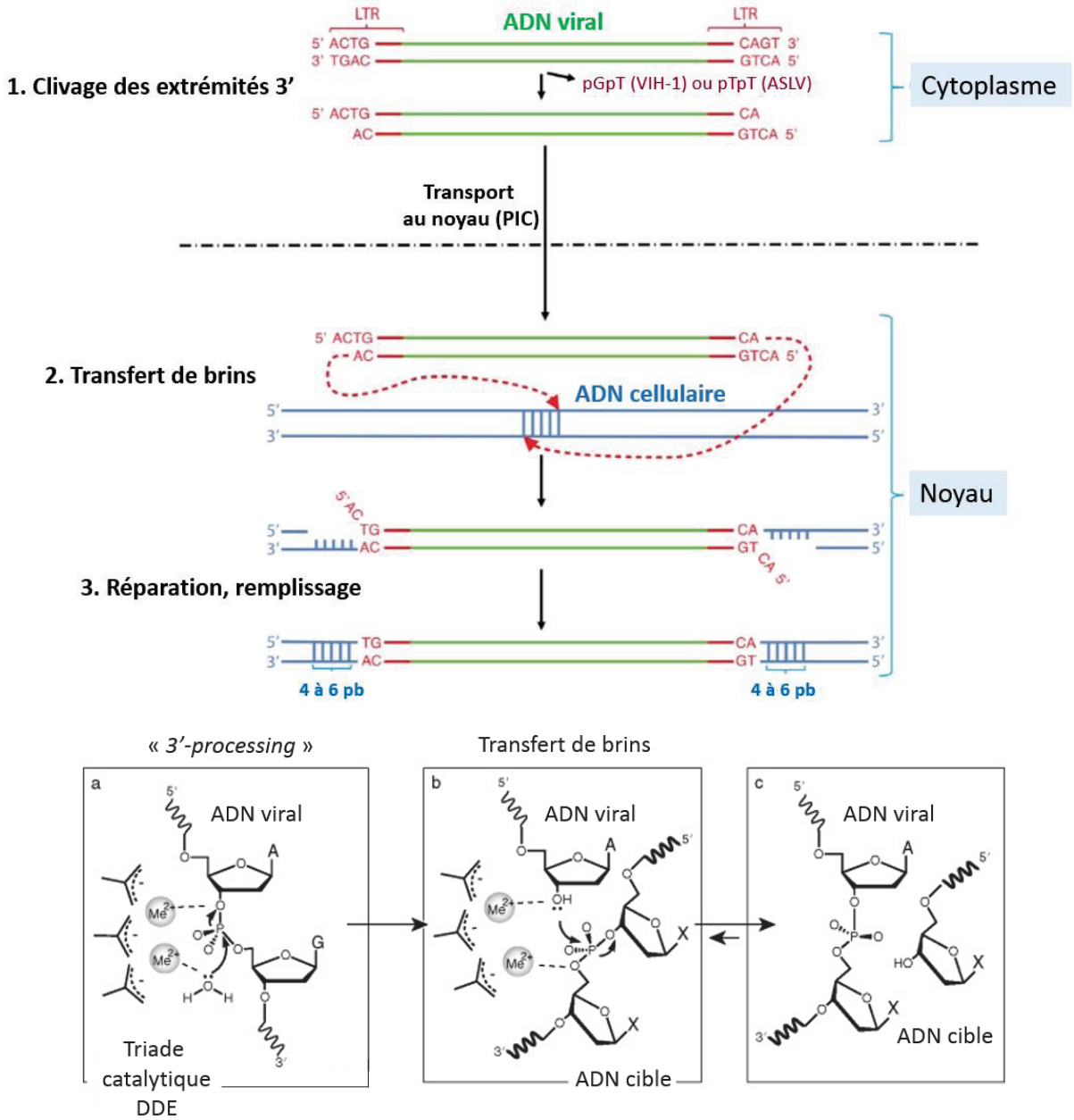


Figure 12. Haut. Représentation schématique de l'intégration rétrovirale (d'après Metifiot *et al.*, 2013). Bas. Mécanisme moléculaire de l'intégration rétrovirale avec les étapes de clivage (a) et de transfert de brins (b) catalysées par l'intégrase (d'après Semenova *et al.*, 2008). Me^{2+} représente les ions Mg^{2+} ou Mn^{2+}.

3. Fonctions secondaires

En dehors de son implication dans le processus d'intégration, l'IN remplit des fonctions secondaires, qui ne semblent pas avoir de rôle dans le processus d'intégration (Neamati *et al.*, 2011).

In vitro, en présence d'un ADN mimant le produit de la réaction de transfert de brins et de cations métalliques divalents, l'IN catalyse la réaction inverse du *strand transfer* : la désintégration (Chow *et al.*, 1992). Cette réaction enzymatique engendre la libération de

l'oligonucléotide mimant l'extrémité de l'ADN viral. Elle permet également la reformation de la liaison phosphodiester au niveau de l'ADN cellulaire (Figure 13). Contrairement aux étapes de clivage et de ligation qui nécessitent la présence de l'enzyme entière, la désintégration requiert au minimum le domaine central catalytique qui contient le site actif de l'IN (Gerton *et al.*, 1997).

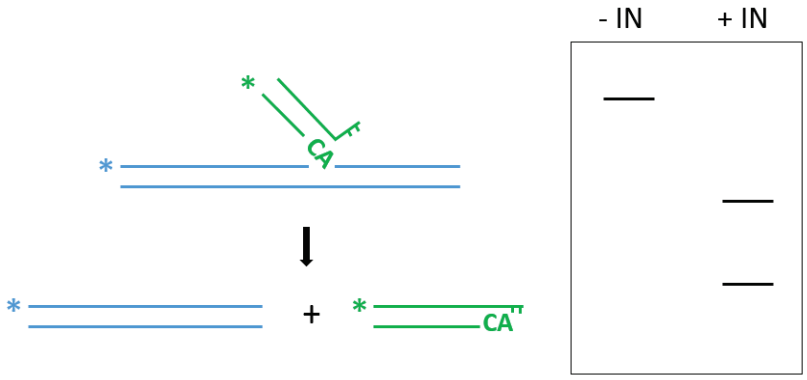


Figure 13. Réaction de désintégration (étoile : marquage radioactif, bleu : ADN cellulaire, vert : ADN viral) et analyse par électrophorèse des produits réactionnels.

Sous certaines conditions expérimentales, l'IN peut également présenter une activité nucléosidase non spécifique *in vitro*, en catalysant l'attaque nucléophile d'un atome d'oxygène provenant d'une molécule d'eau sur la liaison phosphodiester de l'ADN cible (Grandgenett *et al.*, 1978). L'IN peut aussi catalyser *in vitro* la réaction d'alcoolyse non spécifique sur un ADN cible. Comme pour les activités nucléosidase et de désintégration, l'IN catalyse une attaque nucléophile. Elle implique un groupement hydroxyle d'un alcool du milieu réactionnel, comme par exemple le glycérol. L'alcool se lie à l'ADN cible au niveau de son extrémité 5'-phosphate, à l'endroit où le clivage a eu lieu. Ces alcools incorporés peuvent jouer le rôle de la molécule d'eau dans la réaction de *3'-processing* en amorçant une attaque nucléophile (Katzman *et al.*, 1996). Enfin, l'IN joue un rôle dans le clivage des jonctions LTR-LTR ou 2LTR.

L'ensemble des fonctions secondaires de l'IN n'a jamais été observé *in vivo*. L'ADN viral partiellement rétrotranscrit ne peut pas s'intégrer dans le génome cellulaire. Cet ADN peut rester sous forme linéaire ou se circulariser avec une ou deux séquences LTR en tandem à l'intérieur de la cellule cible. Seuls les cercles à deux LTR peuvent être clivés *in vitro* par l'IN, au niveau de la jonction LTR-LTR (Delelis *et al.*, 2008, Mandal *et al.*, 2009).

En dehors de ces fonctions enzymatiques secondaires, l'IN est impliquée dans d'autres processus non enzymatiques de la réplication virale, mis en évidence par la création de mutants pléiotropiques (Engelman, 1999).

Les études menées montrent que l'IN est indispensable à la décapsidation de la particule virale, après son entrée dans la cellule hôte. En interagissant avec la protéine CA, l'IN assure une décapsidation optimale (Briones *et al.*, 2010).

À la suite de cette étape, l'ARN viral est utilisé comme substrat pour la réaction de transcription inverse. Les travaux de mutagenèse dirigée réalisés sur l'IN de VIH-1 décrivent un défaut de conversion du substrat ARN en ADN, mais aucune atteinte de l'intégration rétrovirale (Wilkinson *et al.*, 2011). Ils suggèrent une interaction entre la RT et l'IN pour une synthèse d'ADN réussie (Dobard *et al.*, 2007).

Un autre rôle de l'IN concerne son implication dans l'import nucléaire. L'équipe d'Ikeda a démontré que certains mutants de l'IN générés sont défectifs pour la reconnaissance de l'ADN viral issu de la transcription inverse. Ils affectent la formation du PIC mais également l'import nucléaire. L'IN joue donc un rôle important dans le transport de l'ADN viral au noyau (Ikeda *et al.*, 2004).

Enfin l'IN peut être impliquée, en fin de cycle réplicatif, dans les étapes d'assemblage et de relargage des virions. Malgré sa présence en faible proportion dans la particule virale, une IN mutée (mutations faux sens ou délétions) a le pouvoir de dérégler ces deux étapes (Freed, 2015).

L'IN possède donc des fonctions autres que celles nécessaires à l'intégration. *In vitro*, elle catalyse des réactions enzymatiques secondaires dont le rôle dans le cycle de réplication virale reste à démontrer, comme la désintégration, l'alcoolyse et le clivage des jonctions LTR-LTR dans les ADN non intégrés. Au niveau du cycle réplicatif, elle s'avère aussi indispensable pour la décapsidation, la transcription inverse, l'import nucléaire, l'assemblage et le relargage des virions.

3. Aspects structuraux

1. L'intégrase seule

Les intégrases rétrovirales sont codées par l'extrémité du gène *pol* et sont composées en moyenne de 290 résidus d'acides aminés soit ~32 kDa, bien que certaines aient des séquences plus longues. Par exemple, l'IN du *Prototype Foamy Virus* (PFV) possède 392 résidus d'acides aminés, tandis que celle des ASLV comprend 323 résidus puis 286 résidus après clivage post-traductionnel (Katz *et al.*, 1988). L'IN est constituée de trois domaines fonctionnels reliés entre eux par des boucles flexibles : (i) le domaine amino-terminal NTD, (ii) le domaine central catalytique CCD et (iii) le domaine carboxy-terminal CTD (Jaskolski *et al.*, 2009) (Figure 14). Ces trois domaines sont requis pour la formation d'un complexe stable entre l'IN et l'ADN viral, mais également pour les étapes de clivage et de transfert de brins lors de l'intégration (Vink *et al.*, 1994). Ils sont reliés entre eux par des segments flexibles qui ont rendu difficile l'étude cristallographique de cette enzyme. La structure de l'IN a d'abord été observée par fragments (NTD, CCD, CTD, NTD-CCD, CCD-CTD), comme décrit dans la suite de ce chapitre.



Figure 14. Structure primaire de l'IN de VIH-1 avec le NTD (orange) contenant le motif HHCC, le CCD (bleu) avec la triade catalytique DDE et le CTD (vert).

La structure primaire est peu conservée entre les IN des différents rétrovirus. Ainsi, l'IN de PERV-A/C partage respectivement 17%, 18% et 26% d'identité de séquence avec les IN de RAV-1, VIH-1 et PFV (Demange *et al.*, 2015). Au niveau des domaines isolés, les similarités de séquences et de structures sont plus importantes qu'à l'échelle de la protéine entière, bien que le CTD présente la séquence la moins conservée entre les IN, avec 14% d'identité entre les CTD des IN de PFV et PERV-A/C (Jaskolski *et al.*, 2009) (Figure 15).

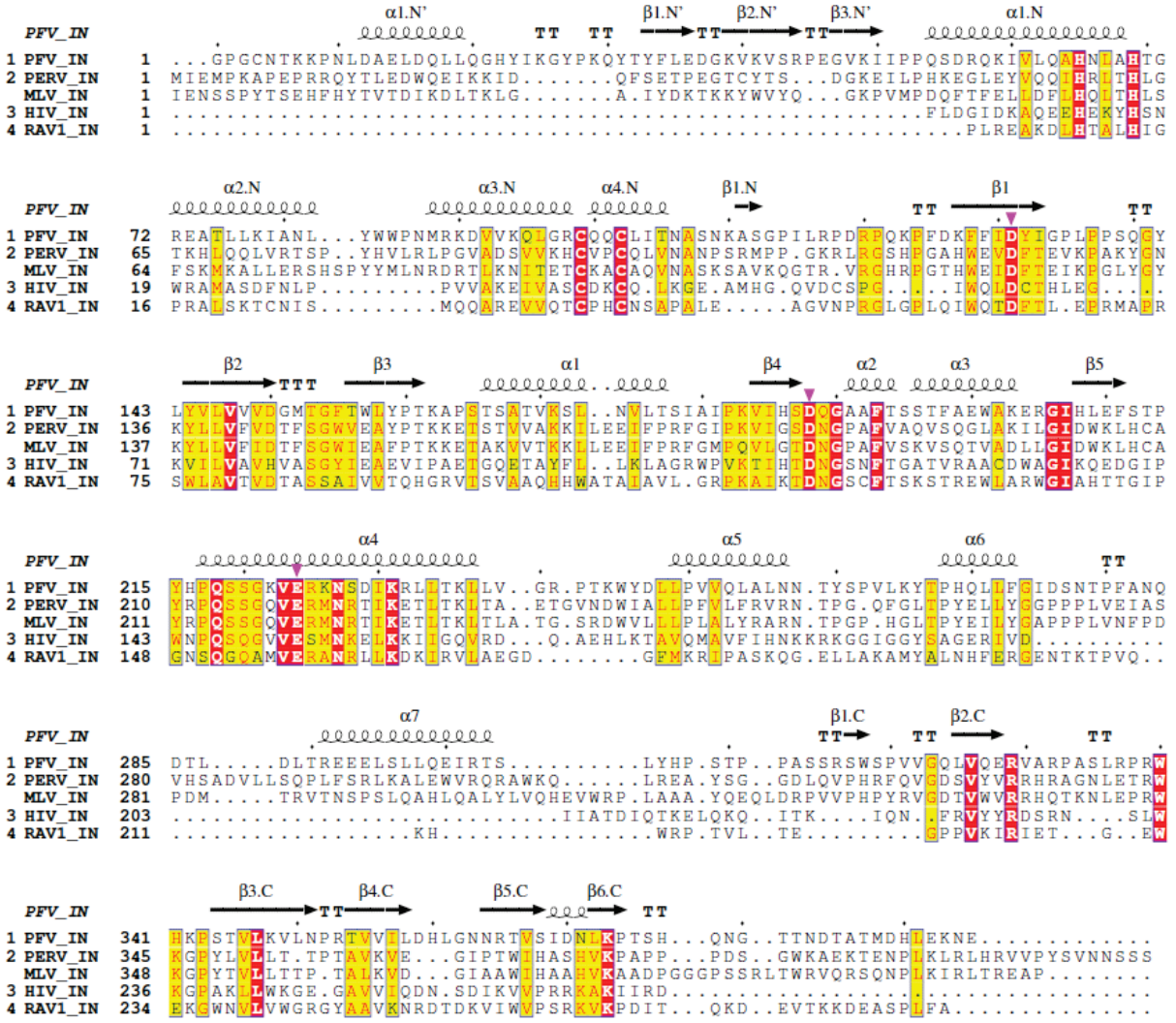


Figure 15. Alignement multiple des séquences des IN du *Spumavirus* PFV, des *Gammaretrovirus* PERV-A/C et MLV, du *Lentivirus* VIH-1 et de l'*Alpharetrovirus* RAV-1. Les éléments de structures secondaires sont schématisés au-dessus de l'alignement pour la structure cristalline de l'IN de PFV (Hare *et al.*, 2010) ; les extensions N', N et C matérialisent respectivement des éléments observés dans les domaines NED, NTD et CTD. Les résidus identiques et similaires (respectivement en rouge et en jaune) et ceux de la triade catalytique (triangles roses) sont également représentés. La figure a été construite avec ESPript (Gouet *et al.*, 2003).

1. Le domaine amino-terminal

Le NTD de l'intégrase rétrovirale contient une cinquantaine de résidus d'acides aminés. Sa structure résolue par Résonance Magnétique Nucléaire (RMN) pour les protéines du VIH-1 (Cai *et al.*, 1997) et VIH-2 (Eijkelenboom *et al.*, 1997) est répertoriée dans la *Protein Data Bank* (PDB). Elle présente un repliement conservé chez toutes les IN rétrovirales et les rétrotransposases (Figure 15). Elle est formée de quatre hélices α et d'un motif hélice-tour-hélice retrouvé dans les protéines liant les acides nucléiques.

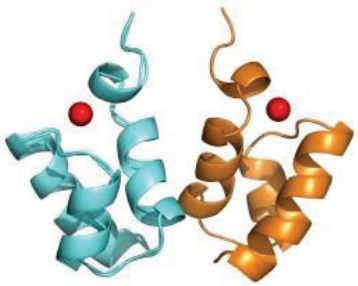


Figure 16. Structure du NTD de l'IN de VIH-1 dimérique en solution, liant des ions zinc (sphères rouges), résolue par RMN (code PDB 1WJC, Cai *et al.*, 1997).

Le NTD porte un motif HHCC constitué de deux résidus histidyl et de deux résidus cystéinyl, commun à toutes les intégrases. Ce motif est capable de lier un ion zinc bien qu'il ne corresponde pas aux repliements classiques de doigt à zinc (Eijkelenboom *et al.*, 1997).

Ce domaine protéique n'est pas directement impliqué dans la catalyse enzymatique (Bushman *et al.*, 1994) mais pourrait reconnaitre et positionner les extrémités de l'ADN viral lors du processus d'intégration (Yang *et al.*, 2001). La fixation de zinc sur le NTD permet de stabiliser la configuration de ses hélices α et de favoriser la multimérisation de l'IN, processus clé dans l'étape d'intégration (Zheng *et al.*, 1996).

Selon les rétrovirus, un domaine nommé NED pour *N-terminal Extension Domain*, situé en amont du NTD, peut être présent dans la structure primaire des IN rétrovirales. Ce domaine protéique supplémentaire est notamment présent dans l'IN des *Spumaretrovirus* et des *Gammaretrovirus* (Figure 15). La structure cristalline du complexe entre l'IN de PFV et l'ADN viral (intasome PFV) montre que ce domaine contribue à l'attachement de l'enzyme à l'ADN cible (Hare *et al.*, 2010).

2. Le domaine central catalytique

Plus d'une trentaine de structures de CCD de différentes intégrases rétrovirales, dont celle de VIH-1, ont été résolues par cristallographie aux rayons X. Dans le cas de VIH-1, elles ont été obtenues avec une protéine portant la mutation F185K, connue pour améliorer la solubilité de ce domaine (Jenkins *et al.*, 1996). Ainsi, les structures cristallines des formes *apo* et complexées aux cations divalents Mg^{2+} et Mn^{2+} (Figure 17) (Goldgur *et al.*, 1998, Maignan *et al.*, 1998) et à un inhibiteur (Goldgur *et al.*, 1999) du CCD de l'IN de VIH-1 sont retrouvées dans la PDB (Bujacz *et al.*, 1996, Dyda *et al.*, 1994).

Le CCD est observé sous forme dimérique en solution (Figure 17). Il présente un repliement de type RNase H retrouvé chez les rétrotransposases et les transposases DDE dont l'IN. Le CCD de l'IN de VIH-1 est formé d'un feuillet β à cinq brins (le second brin étant antiparallèle aux autres) et de six hélices α. Il est connecté au CTD par un segment flexible, qui apparaît plus ou moins structuré en hélice α selon les genres rétroviraux (hélice α7 chez PFV, Figure 15). L'interface de dimérisation du CCD implique des contacts hydrophobes entre l'hélice α1 et l'hélice α5 et est capable de lier des petites molécules (Molteni *et al.*, 2001). La structure tridimensionnelle du CCD de l'IN met en évidence la proximité des trois résidus de la triade catalytique. En effet, les résidus 140 à 149 forment une boucle flexible qui permet d'orienter le résidu E152 en direction des résidus D64 et D116. De par sa flexibilité, cette boucle dite catalytique joue un rôle central dans le processus intégratif (Greenwald *et al.*, 1999).

Le domaine CCD isolé a la capacité de catalyser la première réaction de l'étape d'intégration, le clivage des extrémités virales, mais avec une spécificité amoindrie en comparaison avec la protéine entière (Kulkosky *et al.*, 1995). Il est également capable de reconnaitre les extrémités virales (Agapkina *et al.*, 2006) et de catalyser seul la réaction de désintégration *in vitro*.



Figure 17. Structure cristalline du CCD dimérique de l'IN de VIH-1 (vert) lié à des ions Mg^{2+} (sphères jaunes) (code PDB 1BL3, Maignan *et al.*, 1998).

3. Le domaine carboxy-terminal

Le CTD est le domaine le moins conservé en termes de séquence parmi les IN rétrovirales. La structure du CTD seul de l'IN de VIH-1 obtenue par RMN décrit un dimère en solution (Lodi *et al.*, 1995). Chaque unité monomérique est formée de cinq brins β antiparallèles organisés en tonneau β, avec les brins β2, β3 et β4 impliqués dans l'interface de dimérisation (Figure 18). Cette topologie est comparable au domaine SH3 retrouvé chez les protéines impliquées dans des interactions avec d'autres protéines ou des acides nucléiques.

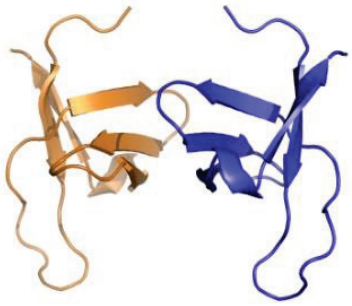


Figure 18. Structure du CTD de l'intégrase de VIH-1 sous forme dimérique résolue par RMN (code PDB 1IHV, Lodi *et al*., 1995).

Le CTD, également connu sous le nom de *DNA binding domain*, est impliqué dans la fixation de l'ADN viral (Esposito *et al.*, 1998) et de l'ADN cellulaire (Lutzke *et al.*, 1998). Il joue aussi un rôle important dans le processus d'oligomérisation (Kalpana *et al.*, 1999, Lutzke *et al*., 1998) et dans le repliement global de l'enzyme (Moreau *et al.*, 2003). Ce domaine est également impliqué dans les activités de clivage et de transfert de brins. En effet, la délétion de ce domaine entraîne une abolition totale de ces deux réactions. En plus de son importance dans le processus intégratif, le CTD lie la protéine cellulaire EED (*Embryonic Ectoderm Development*) et augmente ainsi l'activité de l'IN *in vitro* (Violot *et al.*, 2003).

4. Structures des fragments à deux domaines

En 2001, l'équipe de Wang a résolu la structure du fragment à deux domaines NTD-CCD de l'IN de VIH-1, par cristallographie aux rayons X. Cette structure permet d'observer une interface de dimérisation du NTD différente de celle retrouvée sur le domaine isolé (Wang *et al.*, 2001) et suggère que la structure quaternaire de l'IN peut adopter différents arrangements (Figure 19a).

Plus tard, la structure cristalline du fragment NTD-CCD de l'IN de VIH-2, en complexe avec un domaine de LEDGF/p75, a également été résolue. Dans cette structure, le NTD interagit avec l'IBD du cofacteur cellulaire LEDGF/p75. Un nouvel arrangement quaternaire est observé dans des NTD séparés (Hare *et al.*, 2009) (Figure 19b).

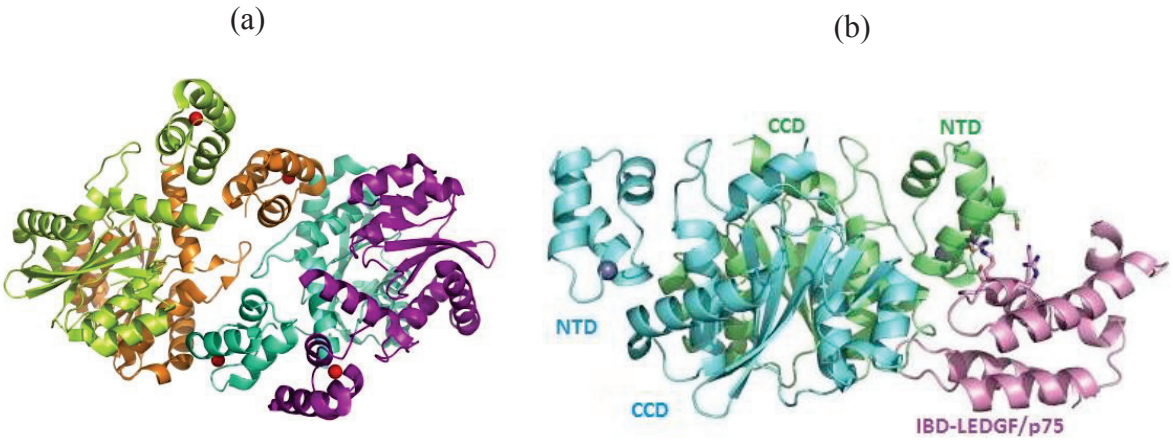


Figure 19. Structures cristallines du domaine NTD de l'IN de VIH-1 en complexe avec le CCD (a) (code PDB 1K6Y, Wang *et al*., 2001) et du NTD de l'IN de VIH-2 en complexe avec le CCD et l'IBD de LEDGF/p75 (b) (code PDB 3F9K, Hare *et al*., 2009).

Les structures cristallines du fragment à deux domaines CCD-CTD des IN de VIH-1 et de RSV ont été résolues par deux équipes différentes en 2000 (Chen *et al.*, 2000, Yang *et al.*, 2000). Si l'interface dimérique CCD-CCD reste préservée, ces deux structures présentent des différences d'arrangement significatives au niveau de leurs CTD. Dans le cas de VIH-1, les CTD sont séparés et positionnés de part et d'autre des domaines centraux (Figure 20a), tandis qu'une interface dimérique CTD-CTD est observée pour RSV (Figure 20b).

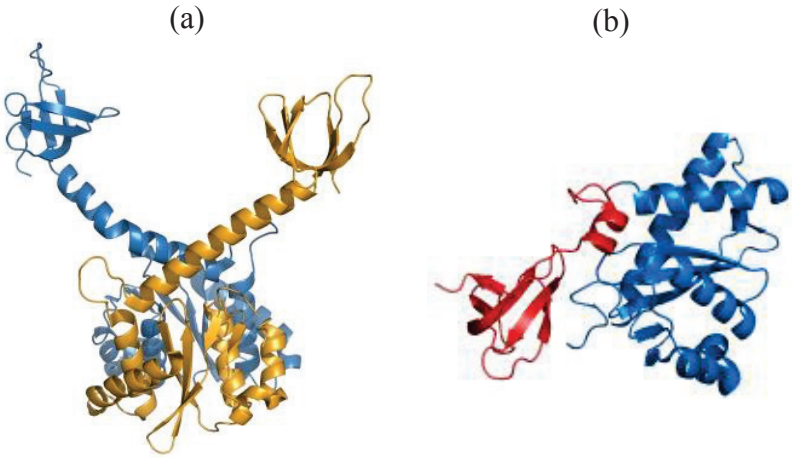


Figure 20. Structures cristallines des complexes CCD-CTD de l'intégrase de VIH-1 (a) (code PDB 1EX4, Chen *et al*., 2000) et de l'intégrase de RSV (b) (code PDB 1C0M, Yang *et al*., 2000).

5. Oligomérisation de l'intégrase

L'arrangement des domaines de l'IN semble donc pouvoir varier suivant les espèces rétrovirales. *In vitro*, l'IN entière adopte différents états oligomériques qui peuvent être caractérisés par chromatographie d'exclusion de taille et par ultracentrifugation analytique (Coleman *et al.*, 1999, Jones *et al.*, 1992). Elle a ainsi été observée sous forme de monomères,

dimères, tétramères et parfois d'octamères. De nombreux facteurs peuvent influencer l'état d'oligomérisation de cette enzyme rétrovirale, tels que les concentrations en protéine, en ADN substrat et en sels présents dans le tampon utilisé. La présence de détergents ou de cations divalents comme des ions Mg^{2+}, Mn^{2+} ou même Zn^{2+} (Lee *et al.*, 1997, Zheng *et al.*, 1996) peut également avoir un effet sur le degré d'oligomérisation de l'IN. D'autres éléments comme la température et la teneur en glycérol du milieu sont aussi capables d'induire un changement d'état oligomérique (Deprez *et al.*, 2000).

Afin de tenter de mieux comprendre le mécanisme d'action de cette enzyme, des premiers modèles de l'IN entière ont été proposés à partir de structures partielles (Wielens *et al.*, 2005), d'expériences de mutagenèse dirigée aléatoire (Calmels *et al.*, 2004) et de diffusion des rayons X aux petits angles ou SAXS (*Small Angle X-ray Scattering*) (Baranova *et al.*, 2007). Dans les dix dernières années, les premières structures expérimentales d'IN entières ont été publiées (Ballandras-Colas *et al.*, 2016, Maertens *et al.*, 2010, Michel *et al.*, 2009, Yin *et al.*, 2016). Elles montrent que lors de l'intégration, cette enzyme peut s'assembler soit sous forme tétramérique, soit sous forme octamérique comme détaillé ci-dessous.

1. Modèles tétramériques

La structure de l'IN entière, sous une forme tétramérique de dimère de dimères, a pu être observée pour la première fois par cryo-microscopie électronique pour l'intasome de VIH-1 en complexe avec son cofacteur LEDGF/p75 (Michel *et al.*, 2009). Elle a ensuite été observée par cristallographie aux rayons X pour l'intasome de PFV (code PDB 3OS2, figure 21) (Maertens *et al.*, 2010).

(a) (b)

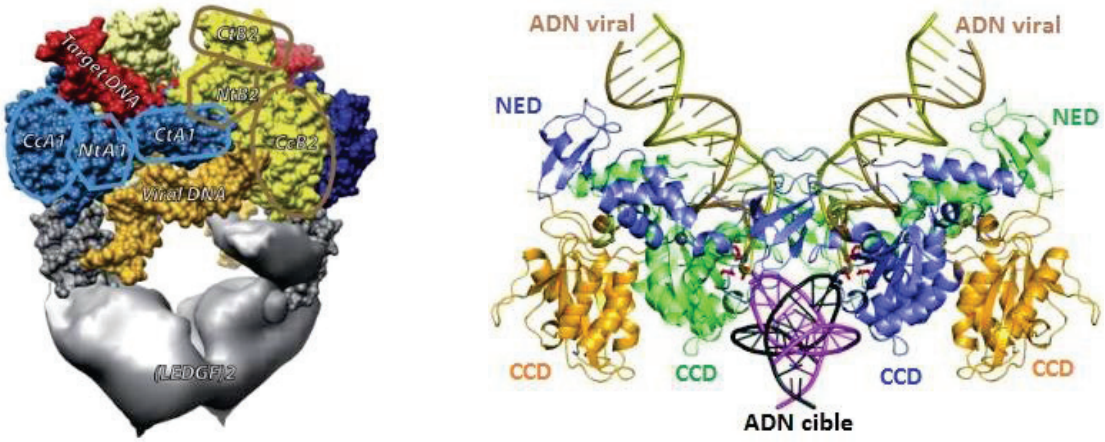


Figure 21. Structures expérimentales de l'IN entière sous forme tétramérique : (a) reconstruction 3D par cryo-microscopie électronique de l'intasome de VIH-1 en complexe avec l'ADN cellulaire et le cofacteur cellulaire LEDGF/p75 (Michel *et al.*, 2009); (b) structure cristalline de l'intasome de PFV complexé à l'ADN cible (code PDB 3OS2, Maertens *et al.*, 2010).

La reconstruction 3D de l'intasome de VIH-1 en présence de LEDGF/p75 et de l'ADN cible a été obtenue à une résolution de 14 Å. La connaissance des structures cristallines partielles de l'IN a servi au positionnement des domaines protéiques les uns par rapport aux autres. Au niveau stœchiométrique, ce modèle décrit quatre protéines IN pour deux partenaires LEDGF/p75. Ce cofacteur cellulaire joue un rôle de plateforme nécessaire à la tétramérisation de l'intégrase. L'organisation du modèle a été validée par des expériences de pontage covalent (Michel *et al.*, 2009).

L'étude cristallographique de PFV a, quant à elle, permis d'observer la structure de l'intasome lors de l'étape de *3'-processing* avec les extrémités d'ADN viral non-clivées (*Uncleaved intasome* ou UI) et clivées (*Cleaved Intasome* CI), ainsi que l'étape de transfert de brins avec l'ADN cellulaire en complexe (*Target Capture Complex* ou TCC) et lié de manière covalente à l'ADN viral brin à brin (*Strand Transfer Complex* ou STC) (Hare *et al.*, 2012) (Figure 22).

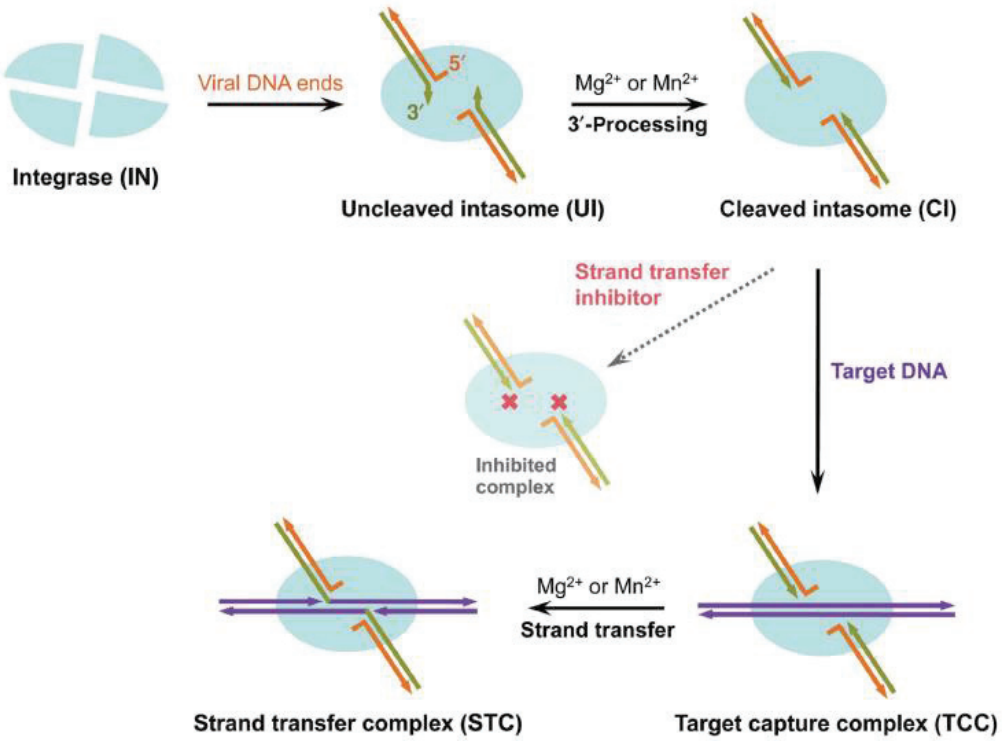


Figure 22. Intégration rétrovirale et intermédiaires réactionnels (Hare *et al.*, 2012).

Toutes ces structures montrant des états fonctionnels successifs sont tétramériques et présentent des repliements similaires. Les changements structuraux sont faibles et localisés au niveau des sites actifs. Ainsi, une comparaison structurale montre que des variations sur les positions des cofacteurs métalliques Mg^{2+} rendent les INSTI inefficaces contre la réaction de *3'-processing* (Marchand *et al.*, 2003).

De par des interactions étroites entre l'IN et l'ADN, la structure de PFV est plus compacte que le modèle de VIH-1. L'arrangement des domaines de l'IN est différent entre le modèle VIH et le modèle PFV. Dans le cas de PFV, le CTD s'intercale entre le NTD et le CCD, alors que le NTD est situé entre le CCD et le CTD dans l'intasome de VIH-1. L'absence du cofacteur LEDGF pourrait être à l'origine d'une différence d'arrangement spatial entre l'IN de PFV et l'IN de VIH-1. Il est important de noter que, du fait de l'agitation des domaines NTD et CTD, ces derniers ne sont visibles que dans l'un des monomères de chaque dimère de la structure de PFV. Cependant, la localisation de ces deux CTD au sein du tétramère de PFV a pu être déterminée par des études SAXS (Gupta *et al.*, 2012).

Par ailleurs, des expériences de SAXS et de pontage covalent ont permis l'obtention d'un modèle dimérique de l'IN entière de RSV sans ADN (Bojja *et al.*, 2011). Cette structure dimérique est proche de celle observée pour le dimère interne de l'intasome de PFV, avec deux

monomères entiers positionnés tête-bêche (Maertens *et al.*, 2010). Le domaine CTD situé entre les CCD et NTD semble ainsi jouer un rôle primordial dans le processus d'oligomérisation de l'enzyme.

2. *Modèles octamériques*

Deux modèles de l'intégrase octamérique ont été récemment publiés. Il s'agit de la reconstruction 3D par cryo-microscopie électronique de l'intasome d'un *Betaretrovirus*, nommé *Mouse Mammary Tumor Virus* (MMTV) (code PDB 3JCA, Ballandras-Colas *et al.*, 2016) (Figure 23) et de la structure cristalline de l'intasome de l'*Alpharetrovirus* RSV (code PDB 5EJK, figure 24) (Yin *et al.*, 2016).

L'intasome de MMTV est composé de huit monomères d'IN et contient une partie centrale (ou *core*) formée par deux dimères d'intégrase qui interagissent avec l'ADN viral (dimères A et B, Figure 23). Cette région est flanquée par deux autres dimères d'intégrase plus mobiles (dimères C et D, Figure 23).

Pour construire ce modèle, les auteurs ont résolu les structures des domaines CCD et CTD et du fragment NTD-CCD de l'intégrase de MMTV par cristallographie aux rayons X. Une modélisation de l'ADN viral selon la structure cristalline de l'intasome de PFV a été réalisée.

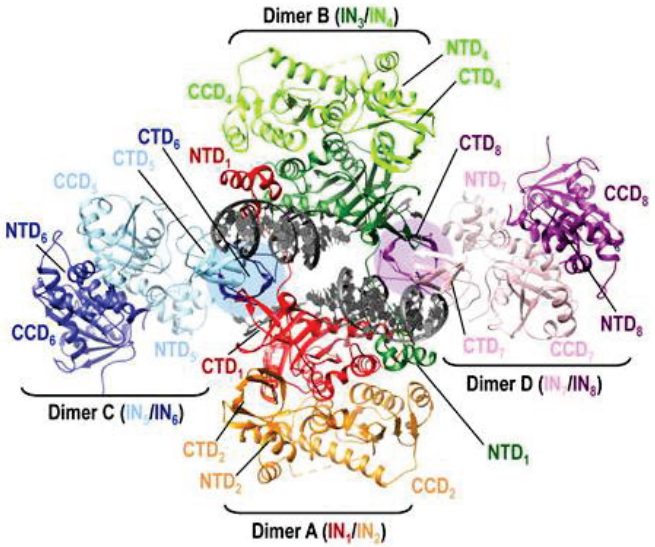


Figure 23. Reconstruction de l'intasome de MMTV par cryo-microscopie électronique (code PDB 3JCA, Ballandras-Colas *et al.*, 2016).

La structure de l'intasome de RSV a été résolue par cristallographie aux rayons X à 3,8 Å (Figure 24). Cet assemblage octamérique est formé par deux paires de dimères d'IN.

L'une d'entre elles, nommée *core* tétramérique, interagit avec les extrémités de l'ADN viral. L'autre paire sert de connecteur entre l'ADN viral et l'ADN cible.

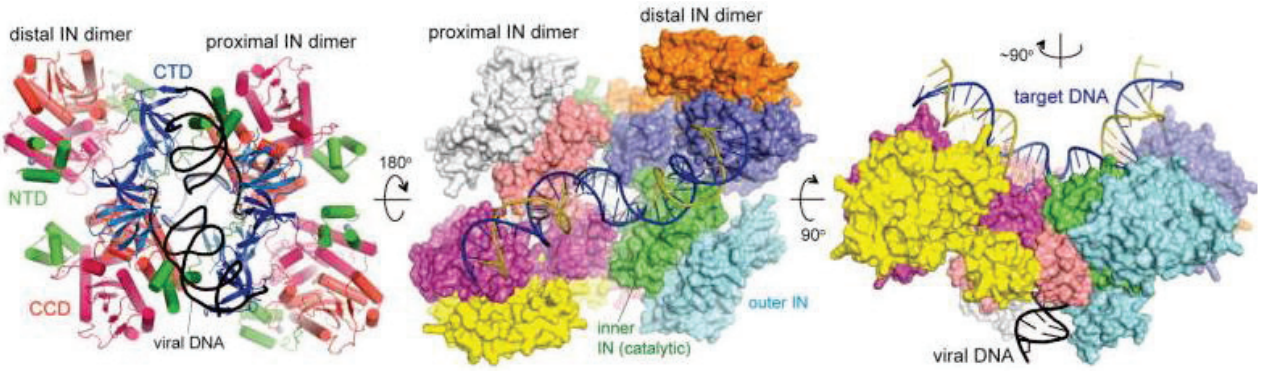


Figure 24. Structure de l'intasome de RSV résolue par cristallographie aux rayons X (code PDB 5EJK, Yin *et al.*, 2016).

En conclusion, il semblerait que l'IN s'assemble sous forme tétramérique dans les intasomes de *Lentivirus* et de *Spumaretrovirus* et sous forme octamérique dans ceux d'*Alpharetrovirus* et de *Betaretrovirus*. Cette différence serait principalement due à la longueur de la jonction entre les domaines CCD et CTD (*linker* CCD-CTD). Celui-ci comprend 50 résidus d'acides aminés dans l'IN de PFV et permet le positionnement des CTD au contact des extrémités de l'ADN viral et de l'ADN cible. En revanche, ce *linker* est beaucoup plus court dans l'IN de MMTV. Ce sont donc les CTD de deux dimères supplémentaires qui assureront cette fonction d'attachement aux extrémités ADN (code PDB 3JCA, Ballandras-Colas *et al.*, 2016).

6. Interaction intasome-nucléosome

Le nucléosome, unité de base de la chromatine, est un complexe nucléoprotéique. Il est composé de 146 pb d'ADN enroulé, chargé négativement, lié à un octamère d'histones contenant deux copies des histones H2A, H2B, H3 et H4 chargées positivement (Fischle *et al.*, 2003, Kornberg *et al.*, 1974). La structure atomique du nucléosome a été résolue par cristallographie aux rayons X (Davey *et al.*, 2002, Luger *et al.*, 1997). Différentes études réalisées *in vitro* démontrent que l'intégration a lieu préférentiellement sur l'ADN nucléosomique en des sites spécifiques, contrairement à ce qui est observé sur l'ADN cible nu (Pruss *et al.*, 1994, Pryciak *et al.*, 1992, Serrao *et al.*, 2016). Le modèle de l'intasome de PFV en complexe avec un mononucléosome a été obtenu par cryo-microscopie électronique à une résolution de 7,8 Å (Maskell *et al.*, 2015) (Figure 25). Il décrit l'implication de trois des quatre unités monomériques d'intégrase au sein de l'interface intasome-nucléosome. L'interaction a

lieu entre des résidus d'acides aminés de l'IN de PFV et l'hétérodimère H2A-H2B du nucléosome. La formation du complexe intasome-nucléosome implique tout particulièrement une interaction entre la queue amino-terminale de l'histone H2A et le domaine carboxy-terminal d'un monomère d'IN de PFV. En accord avec ce modèle structural, la délétion de la queue amino-terminale de H2A tout en conservant celle de H2B, affaiblit la liaison intasome-nucléosome et diminue l'efficacité de transfert de brins dans les nucléosomes (Maskell *et al.*, 2015).

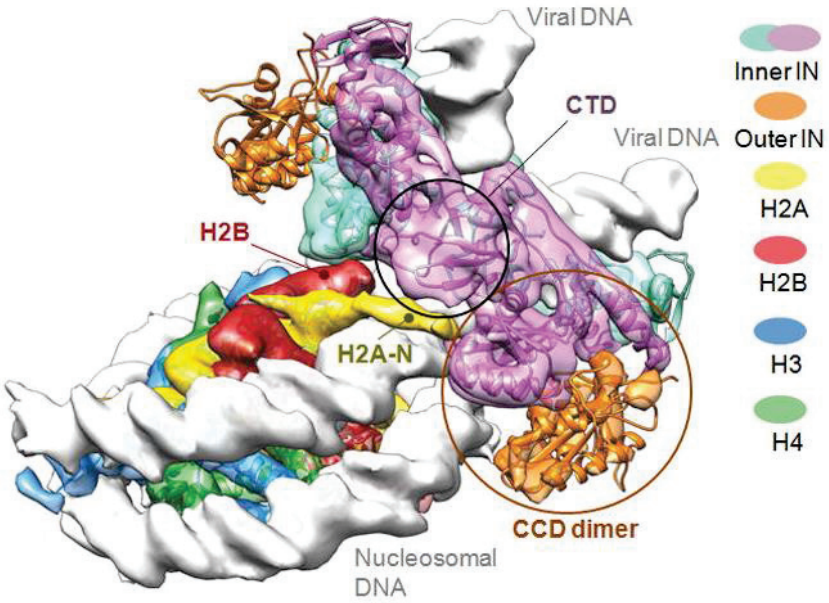


Figure 25. Modèle cryo-microscopique de l'intasome de PFV en complexe avec un mononucléosome (Maskell *et al.*, 2015).

2. Les partenaires cellulaires essentiels à l'intégration productive

Bien que l'IN rétrovirale soit suffisante *in vitro* pour catalyser l'intégration, de nombreux travaux ont mis en évidence le rôle clé de différents partenaires cellulaires *in vivo*. Ce chapitre traitera plus en détails des cofacteurs cellulaires LEDGF/p75 et Brd2, qui sont respectivement les partenaires des IN de VIH-1 et de PERV-A/C étudiées au laboratoire. Ces deux cofacteurs jouent des rôles similaires pour des espèces rétrovirales différentes, en dirigeant l'IN vers des régions promotrices ou transcriptionnellement actives de la cellule hôte.

1. LEDGF/p75

Le partenaire cellulaire LEDGF/p75, identifié indépendamment par plusieurs équipes, est spécifique aux *Lentivirus* (Cherepanov, 2007). Il s'agit d'un co-activateur général de transcription ubiquitaire de 75 kDa (Ge *et al.*, 1998), majoritairement retrouvé dans le noyau

cellulaire, où il est lié aux chromosomes. Il appartient à la famille HRP (*Hepatoma derived growth factor Related Proteins*) (Cherepanov *et al.*, 2004). Contrairement à sa dénomination (*Lens Epithelium-Derived Growth Factor*), LEDGF n'est ni un facteur de croissance, ni une protéine spécifique de l'épithélium du cristallin. Son interaction avec l'IN de VIH-1 a été mise en évidence par des expériences de double hybride et de co-immunoprécipitation (Emiliani *et al.*, 2005, Turlure *et al.*, 2004).

L'ARN messager codant LEDGF subit un épissage alternatif, ce qui conduit à la synthèse de deux isoformes nommées p75 et p52 (Singh *et al.*, 2000). L'implication de p52 dans l'activation de la transcription et dans l'épissage de différents ARN pré-messagers est clairement établie (Ge *et al.*, 1998). L'isoforme p75 semble également interagir avec la machinerie d'épissage (Singh *et al.*, 2015). Seule l'isoforme p75 possède une région IBD au niveau de sa région carboxy-terminale (Figure 26) qui permet de lier l'IN de VIH-1 (Cherepanov *et al.*, 2005, Michel *et al.*, 2009).

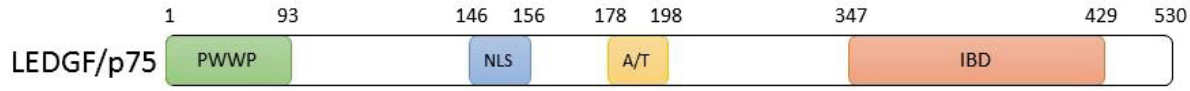


Figure 26. Représentation schématique de la structure primaire du cofacteur cellulaire LEDGF/p75 (d'après Reddy *et al.*, 2014).

En plus de l'IBD, LEDGF/p75 porte une séquence NLS et deux copies d'un motif nommé crochet A/T nécessaires à la liaison de régions riches en A/T de l'ADN. Le motif PWWP (Pro-Trp-Trp-Pro) présent à l'extrémité amino-terminale de ce cofacteur est retrouvé dans les protéines associées à la chromatine. Le crochet A/T et le motif PWWP sont indispensables à la liaison de LEDGF/p75 à la chromatine (Llano *et al.*, 2006, Turlure *et al.*, 2006) (Figure 27).

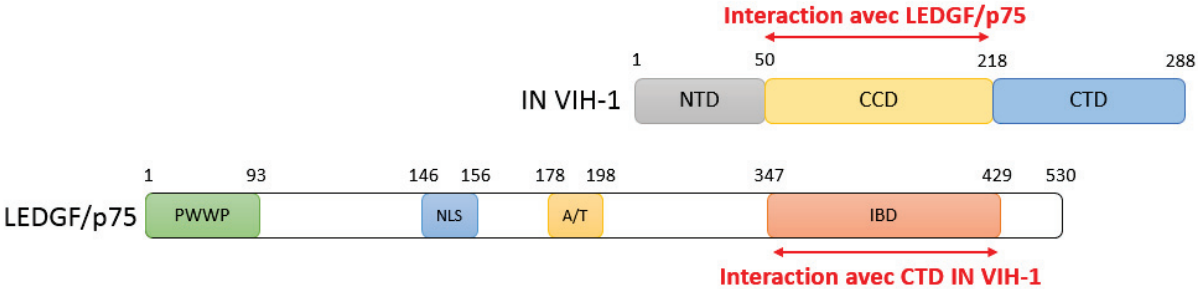


Figure 27. Zones d'interaction entre l'IN de VIH-1 et son cofacteur cellulaire LEDGF/p75 (d'après Reddy *et al.*, 2014).

Le cofacteur cellulaire LEDGF/p75 est associé aux processus pré-intégratif et d'intégration (Llano *et al.*, 2006). En effet, une inhibition de son expression par un petit ARN interférent (siRNA) bloque l'import nucléaire de l'intégrase de VIH-1, ainsi que son association à l'ADN

cellulaire (Maertens *et al.*, 2003). LEDGF/p75 participe à la formation du PIC (Vandegraaff *et al.*, 2006), augmente l'affinité de l'IN lentivirale pour l'ADN cible (Busschots *et al.*, 2005) et stimule l'étape de transfert de brins lors de l'intégration *in vitro* (Cherepanov *et al.*, 2003). Il est également impliqué dans le ciblage de sites spécifiques d'intégration de l'IN de VIH-1 (Emiliani *et al.*, 2005, Maertens *et al.*, 2003) et, plus précisément, de régions activement transcrites (Poeschla, 2008). En effet, le motif PWWP interagit avec la queue amino-terminale de l'histone H3K36me3 (histone H3 triméthylée sur le résidu lysyl en position 36) (Eidahl *et al.*, 2013). Cette histone est associée à la chromatine transcriptionnellement active (Bernstein *et al.*, 2007). Ce ciblage spécifique est en adéquation avec les sites d'intégration préférentiels des *Lentivirus* (Schroder *et al.*, 2002). Le cofacteur LEDGF/p75 permet ainsi une intégration efficace catalysée par l'IN de VIH-1 (Kvaratskhelia *et al.*, 2014). Ce partenaire cellulaire protège également l'IN de la dégradation par le protéasome en empêchant son ubiquitination (Llano *et al.*, 2004, Llano *et al.*, 2006). Des recherches sont en cours pour développer des inhibiteurs d'interaction protéine-protéine LEDGIN, qui ciblent l'association entre le CCD de l'IN de VIH-1 et l'IBD de LEDGF/p75 (Debyser *et al.*, 2015, Demeulemeester *et al.*, 2014).

2. La famille BET

La famille BET (*Bromodomain and Extra Terminal domain*) regroupe des protéines régulatrices retrouvées chez la levure, la drosophile et les mammifères, impliquées dans différents processus cellulaires comme la réplication, la transcription, la régulation de l'expression génique et le remodelage de la chromatine (LeRoy *et al.*, 2008, Mantovani *et al.*, 2010, Weidner-Glunde *et al.*, 2010, Wu *et al.*, 2007).

Elles sont porteuses d'un ou deux BromoDomaines BD du côté amino-terminal et d'un domaine extra-terminal ET proche de l'extrémité carboxy-terminale. Les BD sont des motifs de 110 résidus d'acides aminés capables de fixer des histones acétylées, tandis que le domaine ET est impliqué dans des interactions protéine-protéine.

Chez les mammifères, il existe quatre membres de la famille BET : Brd2, Brd3 et Brd4 sont ubiquitaires tandis que Brdt est spécifique du tissu testiculaire (Florence *et al.*, 2001) (Figure 28). Ces protéines interagissent avec des histones distinctes, ce qui suggère des affinités différentes pour la chromatine et le ciblage de régions spécifiques du génome (Dey *et al.*, 2003, Moriniere *et al.*, 2009, Umehara *et al.*, 2010). Il est intéressant de noter que Brd4 est capable d'interagir avec un résidu lysyl acétylé d'une sous-unité du facteur transcriptionnel NF-κB, par

l'intermédiaire de ses bromodomaines (Huang *et al.*, 2009). Cette interaction laisse supposer que les histones ne sont pas les seules partenaires acétylées reconnues par les bromodomaines des protéines BET.

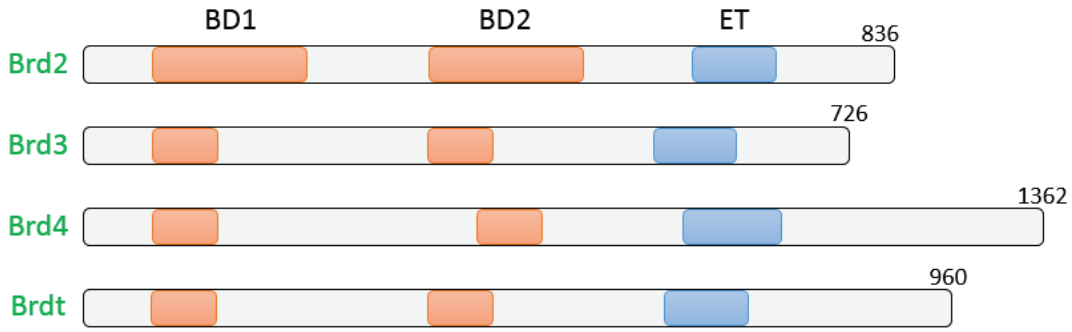


Figure 28. Représentation schématique de la structure primaire des protéines humaines BET. La séquence de Brd2 correspond à l'isoforme 2 utilisée dans nos travaux (d'après Nakamura *et al.*, 2007).

La protéine Brd2 (*Bromodomain containing protein 2*), également connue sous le nom de RING3 (*Really Interesting New Gene 3*), possède deux bromodomaines BD1 et BD2 responsables de l'interaction entre Brd2 et la chromatine acétylée. Ils interagissent spécifiquement avec les résidus lysyl 5, 8 et 12 acétylés de la queue d'histone H4 (H4K5ac, H4K8ac, H4K12ac) (Hnilicova *et al.*, 2013, Taverna *et al.*, 2007, Umehara *et al.*, 2010). Résolue par cristallographie aux rayons X à 2 Å de résolution, la structure du complexe formé par BD1 et la queue d'histone H4K12ac révèle un état dimérique de BD1, ainsi qu'une acétyl-lysine au niveau de l'interface dimérique (Umehara *et al.*, 2010).

Brd2 s'associe à la chromatine acétylée au cours de la mitose, contrairement aux autres protéines BET qui s'en détachent. Cette protéine porte également un NLS situé entre BD2 et le domaine ET, lui permettant d'accéder au noyau cellulaire. Entre les deux bromodomaines, il existe un domaine *kinase like* dont l'activité kinase est activée à la suite d'un signal mitogénique (Denis *et al.*, 2006, Nakamura *et al.*, 2007).

La protéine Brd2 est impliquée dans le processus de régulation de la transcription. Elle peut l'activer en se liant à des complexes protéiques contenant le facteur de transcription E2F. Cette liaison entraîne le recrutement d'Histones AcétylTransférases (HAT) et des facteurs de remodelage de la chromatine (Denis *et al.*, 2006), au niveau du promoteur du gène codant la cycline A (Sinha *et al.*, 2005). Brd2 joue le rôle de plateforme nécessaire au recrutement de ces différents facteurs et à l'activation de la transcription de ce gène.

À l'inverse, Brd2 peut avoir un effet inhibiteur sur l'expression de certains gènes. Dans le modèle murin, Brd2 recrute différents co-répresseurs transcriptionnels, entraînant ainsi la répression du gène PPARγ au niveau des adipocytes et l'inhibition de l'adipogenèse. L'absence de régulation de la transcription par Brd2 peut conduire à un profil d'obésité sévère chez la souris (Wang *et al.*, 2009).

Les protéines de la famille BET, plus particulièrement Brd2, 3 et 4, interagissent avec l'IN du *Gammaretrovirus* MLV (Gupta *et al.*, 2013). Cette interaction implique le domaine ET de ces partenaires cellulaires et le CTD de l'IN de MLV, associé à son ADN viral. Ces cofacteurs ont la capacité d'interagir de façon simultanée avec l'intasome de MLV et avec les queues d'histones acétylées. Cette interaction bimodale permet aux protéines BET de guider l'IN vers ses sites préférentiels d'intégration situés sur l'ADN cible.

Le cofacteur Brd4 est spécifique de l'IN de MLV. Il stimule son activité d'intégration de façon similaire à LEDGF/p75 sur l'IN de VIH-1. Cette interaction engage au minimum les domaines ET de Brd4 et EBM (*ET Binding Motif*) situé à l'extrémité carboxy-terminale du CTD de l'IN de MLV (Crowe *et al.*, 2016). Il s'agit de régions respectivement conservées dans les protéines BET et dans les IN gammarétrovirales. Des études RMN montrent que la séquence flexible EBM s'ordonne lorsqu'elle interagit avec le domaine ET pour former un feuillet de deux brins β antiparallèles (Aiyer *et al.*, 2014, Crowe *et al.*, 2016). Ainsi l'IN de MLV, en interaction avec Brd4, cible les sites d'initiation de la transcription pour une intégration efficace (Kvaratskhelia *et al.*, 2014). Cette interaction protéine-protéine joue également un rôle dans le recrutement du PIC nécessaire à l'import nucléaire de l'intégrase (Crowe *et al.*, 2016).

En présence de l'inhibiteur JQ1 qui cible les protéines BET, l'activité d'intégration de l'IN de MLV diminue *in vivo* et l'ADN viral est intégré hors des sites préférentiels. En effet, cette petite molécule inhibitrice se fixe à la poche hydrophobe présente dans l'un des bromodomaines des protéines BET, ce qui bloque l'interaction avec les queues d'histones acétylées (Filippakopoulos *et al.*, 2010). Cependant, JQ1 n'a aucun effet sur l'étape de transcription inverse lors du cycle réplicatif rétroviral (Gupta *et al.*, 2013).

Les équipes collaboratrices du Dr Ronfort (« Rétrovirologie, Intégration virale et Vaccins », INRA Lyon) et du Dr Blanchard (« Biosécurité virale », ANSES Ploufragan) ont mis en évidence une interaction similaire à celle décrite pour Brd4 et le CTD de l'IN de MLV, par méthode double-hybride et immunoprécipitation. Elle implique l'IN d'un autre

Gammaretrovirus PERV-A/C et la protéine humaine Brd2 qui permet une intégration productive des *Gammaretrovirus* (De Rijck *et al.*, 2013). La zone d'interaction localisée implique également le CTD de l'IN de PERV-A/C et une région contenant le domaine ET de Brd2 (article prochainement soumis à publication) (Figure 29).

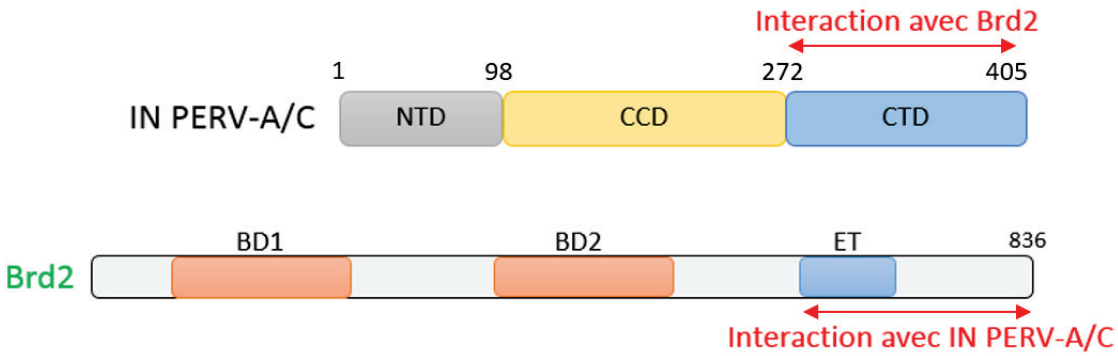


Figure 29. Interaction entre le CTD de l'IN de PERV-A/C et le CTD du cofacteur cellulaire Brd2.

Au cours de ma thèse, j'ai étudié les déterminants moléculaires et structuraux d'interaction entre l'IN de PERV-A/C et le cofacteur cellulaire Brd2, comme décrit dans le chapitre 'Résultats et Discussion'.

Objectifs du projet doctoral

Mon travail doctoral sur les intégrases rétrovirales a porté sur deux axes de recherche. Le premier concerne les études moléculaire et structurale d'un nouveau mode de dimérisation des IN rétrovirales et la recherche de modulateurs d'oligomérisation anti VIH-1. Le second axe est lui basé sur l'étude de l'IN de PERV-A/C en complexe avec son cofacteur cellulaire humain Brd2, dans un contexte de xénotransplantation.

La première partie des résultats de mon manuscrit s'inscrit dans la continuité des travaux de thèse réalisés par les Drs Allison Ballandras (2009-2012) et Romain Merceron (2010-2013).

A. Ballandras a résolu la structure cristalline du CCD de l'IN de RAV-1 à une résolution de 1,8 Å (Ballandras *et al.*, 2011). Un « nouvel » assemblage dimérique a ainsi pu être observé. La « nouvelle » interface dimérique engage trois paires d'hélices α contre deux paires d'hélices α dans les assemblages dimériques « canoniques » de CCD. Le passage entre les deux formes dimériques nécessite un glissement de 15 Å d'un monomère par rapport à l'autre. Cette translation entraîne l'apparition de deux crevasses, l'une d'entre elles renfermant une molécule de tampon MES (acide 2-[*N*-morpholino]-éthane-sulfonique) apportée par la solution de cristallisation (Figure 30).

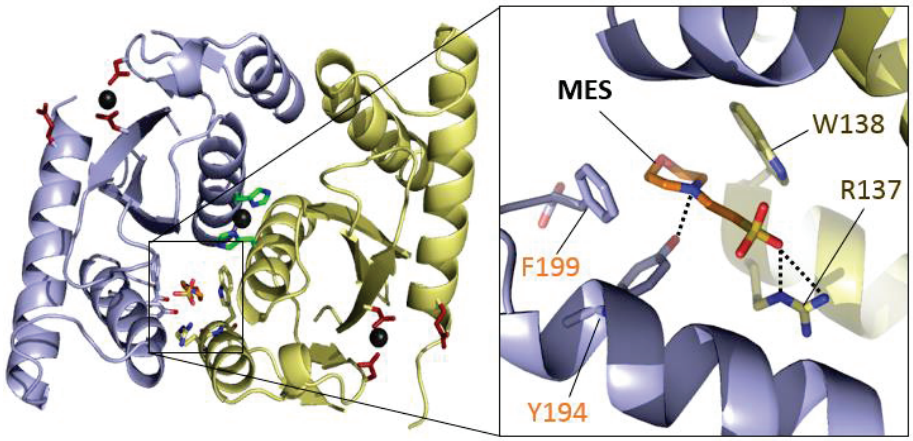


Figure 30. Structure cristallographique du « nouvel » assemblage dimérique du CCD de l'IN de RAV-1 (code PDB 3O4N, Ballandras *et al.*, 2011) en complexe avec une molécule de MES de la solution de cristallisation.

Afin de stabiliser ce « nouvel » assemblage en solution, un mutant CCD a été créé. Des études par modélisation moléculaire ont mis en évidence la possibilité de former un pont disulfure intermoléculaire par mutation des résidus histidyl en résidus cystéinyl en position 103 (Ballandras *et al.*, 2011). La mise au point des protocoles de surproduction et de purification du mutant CCD H103C de l'IN de RAV-1 a été réalisée par R. Merceron. En parallèle, il a également optimisé ces protocoles pour le CCD sauvage de l'IN de RAV-1.

Notre hypothèse de travail est basée sur les différences significatives observées entre l'interface « canonique » et la « nouvelle » interface. Cette dernière pourrait concerner l'IN entière et ainsi provoquer des réarrangements structuraux majeurs, menant à une conformation impropre à l'étape d'intégration. Le « nouvel » assemblage pourrait être stabilisé par des modulateurs d'oligomérisation. Ces petites molécules se fixeraient dans des sites éloignés du site actif et bloqueraient l'IN dans une conformation inactive. Elles pourraient être utilisées en combinaison avec des inhibiteurs de type INSTI, dans le cadre d'un traitement HAART, avec un minimum de risques d'apparition de résistances croisées.

Des composés analogues au MES, puis d'autres molécules prédites par *docking in silico* comme ayant une bonne affinité pour le CCD de l'IN de RAV-1, ont ainsi été synthétisés par l'équipe collaboratrice du Dr Maurin (« Chimie Bio-organique », ENS Lyon) et testés *in vitro* (par intégration concertée), comme décrit dans la suite du manuscrit.

En parallèle, je me suis intéressée au CCD de l'IN de VIH-1. En effet, les composés sélectionnés seront également testés *in vitro* et *in cellulo* sur le CCD de l'IN de VIH-1. Pour ce faire, j'ai travaillé sur la mise au point des étapes de surproduction et de purification des CCD de l'IN de VIH-1 F185K et Y99C/F185K. La mutation F185K est connue pour améliorer la solubilité de cette IN rétrovirale (Jenkins *et al.*, 1996), tandis que la mutation Y99C correspond à l'équivalent H103C du CCD de l'IN de RAV-1 (Figure 15, page 40). Cette mutation permettra de reproduire le « nouvel » assemblage dimérique à l'échelle de l'IN de VIH-1.

La seconde partie de mon manuscrit porte sur l'étude que j'ai menée sur l'IN de PERV-A/C et son cofacteur cellulaire humain Brd2. J'ai tout d'abord modélisé l'intasome de PERV A/C en complexe avec le raltégravir, dans le cadre d'une collaboration menée avec les équipes des Drs Ronfort (INRA Lyon) et Blanchard (ANSES Ploufragan), pour étudier l'effet inhibiteur de cette molécule sur PERV. Cette étude a donné lieu à un article que je signe en co-premier auteur (Demange *et al.*, 2015). Puis j'ai mené une étude biochimique, biophysique et structurale pour caractériser l'association de l'IN de PERV-A/C avec Brd2, *via* des fragments d'interaction minimaux impliquant les régions carboxy-terminales des deux protéines (article en préparation). La structure tridimensionnelle de ce complexe protéique n'est pas connue. Elle peut servir de guide pour la conception rationnelle d'inhibiteurs d'interaction protéine-protéine. À long terme et dans le cadre d'une xénogreffe, un traitement préventif composé d'un inhibiteur ciblant cette interaction et d'un inhibiteur ciblant l'activité de transfert de brins de l'IN

(raltégravir, elvitégravir, dolutégravir) pourrait être administré, pour éviter l'apparition d'une xénozoonose.

Matériels et méthodes

I. Techniques de biologie moléculaire

1. Constructions plasmidiques

Différentes constructions plasmidiques ont été nécessaires à la surproduction des protéines recombinantes d'intérêt. Elles contiennent toutes une origine de réplication bactérienne ColE1 indispensable à la réplication du plasmide, un gène de résistance à un antibiotique, le gène qui code le répresseur LacI et l'opérateur de l'opéron lactose sous le contrôle du promoteur T7 et le terminateur T7 (Studier *et al.*, 1986), sauf le plasmide pPROEX-Hta (*Thermo Fisher Scientific*) qui porte le promoteur *trc.* Ce dernier permet un fort taux d'expression basale du gène placé sous sa dépendance (Tegel *et al.*, 2011). La séquence qui code la protéine d'intérêt est insérée dans un site multiple de clonage (*Multiple Clonage Site* ou MCS), entre différents sites de restriction, sauf dans le plasmide pETG-10a créé avec le système Gateway (*Invitrogen*) avant mon arrivée au laboratoire.

Une séquence nucléotidique correspondant à une étiquette poly-histidine (six résidus histidyl) est placée en amont du gène codant la protéine d'intérêt. Elle permet de purifier la protéine recombinante par chromatographie d'affinité pour le nickel (Porath *et al.*, 1975). Le plasmide pETM-30 contient, en aval de cette séquence, une séquence qui code la Glutathion-S-Transférase ou GST. Cette enzyme peut servir d'étiquette de purification lors d'une chromatographie d'affinité pour le glutathion (Smith *et al.*, 1988, Terpe, 2003).

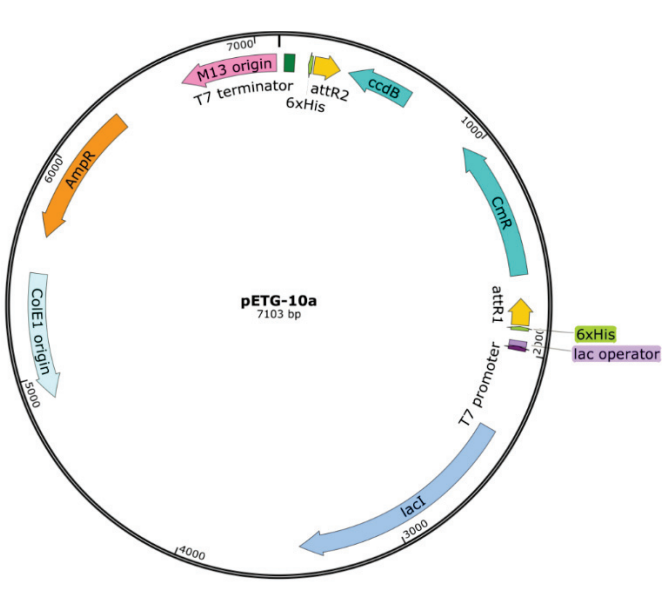
L'étiquette de purification peut être éliminée par l'action d'une protéase : la thrombine ou la protéase du virus de la gravure du tabac (*Tobacco Etch Virus* ou TEV). Une séquence nucléotidique correspondant au site de clivage de l'enzyme peut ainsi être ajoutée entre les séquences qui codent l'étiquette de purification et la protéine d'intérêt.

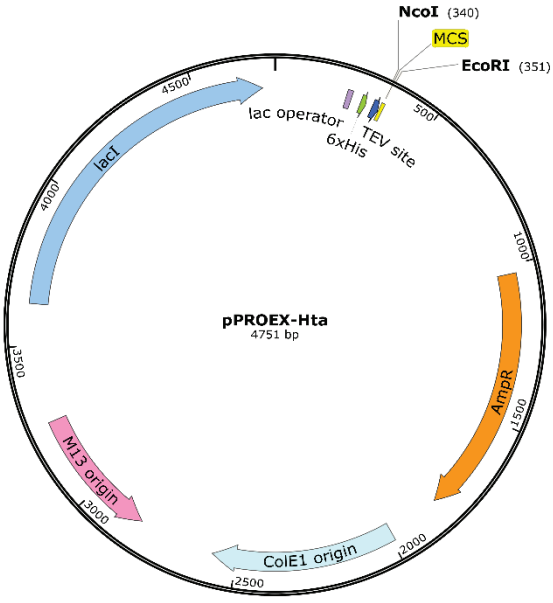
Les plasmides pETG-10a et pETM-30 proviennent de l'EMBL (*European Molecular Biology Laboratory*). Le plasmide pET28a, fourni par l'équipe du Dr Samarut (*Functional Genomics of Thyroid Signaling*, IGFL Lyon), servira à cloner les gènes codant la protéase du TEV, le CTD de l'IN de PERV-A/C et le CTD de Brd2. Cette construction « multiTEV » est décrite dans le chapitre suivant.

L'ensemble des plasmides utilisés pour surproduire les protéines recombinantes d'intérêts est répertorié ci-dessous (Tableau 1, Figure 31).

Plasmide d'expression	Gène de résistance à un antibiotique	Promoteur	Insert	Sites de restriction	Sites de clivage
pETG-10a	Ampicilline	T7	- CCD sauvages et mutés des IN de RAV-1 et de VIH-1		thrombine
pPROHEX-Hta	Ampicilline	*trc*	- IN de RAV-1 - CTD de Brd2	*NcoI* et *EcoRI*	Protéase du TEV
pETM-30	Kanamycine	T7	CTD de l'IN de PERV-A/C	*NcoI* et *SalI*	Protéase du TEV
pET28a	Kanamycine	T7	Construction « multiTEV »	*NcoI* et *NotI*	Protéase du TEV (plusieurs sites)

Tableau 1. Récapitulatif des différents éléments des plasmides d'expression nécessaires à la surproduction des protéines d'intérêt.





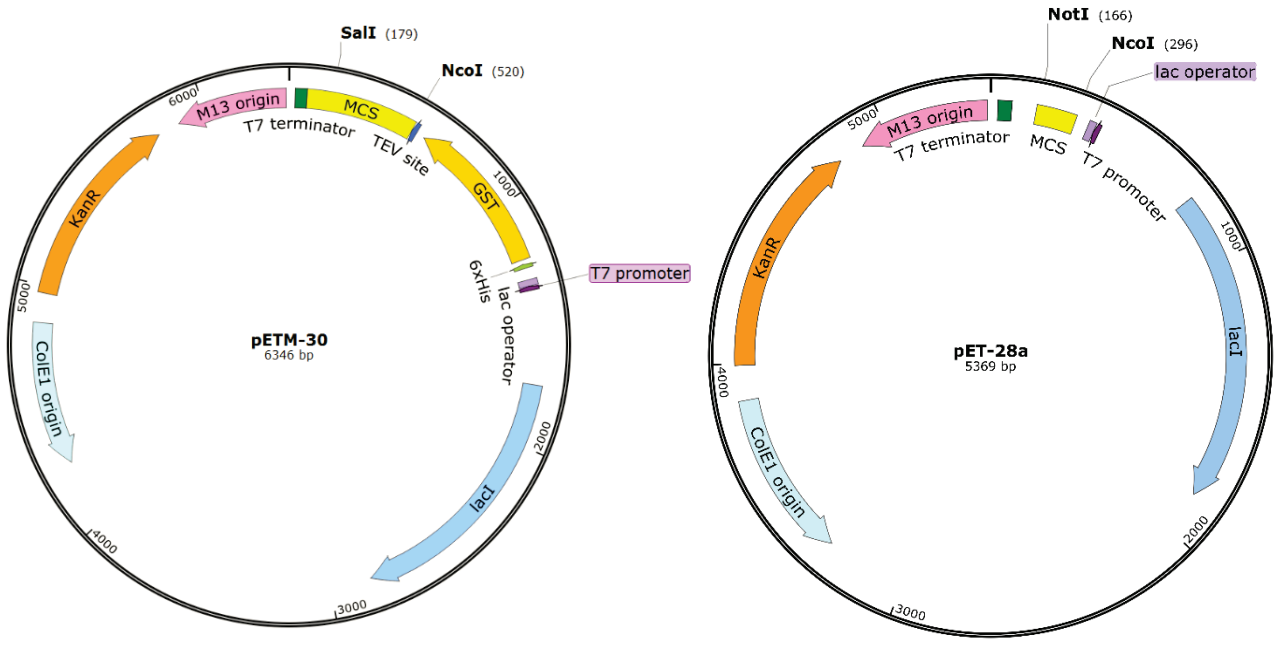


Figure 31. Cartes des plasmides d'expression construits pour surproduire les protéines d'intérêt.

2. Construction « multiTEV »

Cette construction a été clonée dans le plasmide pET28a. Elle est composée de trois gènes codant i) la protéase du TEV ; ii) le CTD de l'IN de PERV-A/C porteur d'une étiquette de purification Strep-tag®II du côté amino-terminal ; iii) le CTD de Brd2 avec une étiquette de purification poly-histidine à son extrémité amino-terminale. Ils sont séparés entre eux par des séquences qui codent des sites de clivage à la protéase du TEV (sites TEV) (Figure 32).



Figure 32. Représentation schématique de l'insert utilisé dans la construction « multiTEV ».

En théorie, une fois l'enzyme du TEV produite, elle reconnaît les sites TEV (ENLYFQG ou ENLYFQS) et libère les CTD des deux partenaires protéiques. Le clivage par la protéase du TEV est décrit dans la suite du manuscrit (*cf.* § II.7). Le complexe entre les CTD de l'IN de PERV-A/C et Brd2, nommé IN_{CTD}-$Brd2_{CTD}$, sera déjà préformé ou se formera après l'action de cette protéase.

Cette stratégie originale a été développée à partir des travaux de recherche menés par l'équipe du Dr Ruff (« Biologie structurale intégrative », IGBMC Illkirch) (Levy *et al.*, 2016).

Kathy Gallay, assistant-ingénieur dans l'équipe du Dr Ronfort (INRA Lyon) et moi-même avons réalisé cette construction.

1. Amorces utilisées

Différentes amorces sont utilisées pour la fabrication de ce plasmide recombinant. Elles sont synthétisées par le fournisseur *Sigma*. Certaines d'entre elles portent les sites TEV et les étiquettes de purification des deux CTD (Tableau 2).

Gène	Nom des amorces	Séquence des amorces
Protéase du TEV	TEV_F1	5′ GGAGATATACCATGGGCGGAGAAAGCTTGTTTAAGG 3′
	TEV_R1	5′ TGGCTCCAGCTAGCACCTTGAAAATAAAGATTTTCGAGCTCGCG ACGGCGACGACGATTCATGAGTTGAGTCGC 3′
CTD de l'IN de PERV-A/C	CTDPERV_F1	5′ CAAGGTGCTAGCTGGAGCCACCCGCAGTTCGAAAAAGCCTACTC AGGAGGAGACTTG 3′
	CTDPERV_R1	5′ ATGATGATGATGACCTTGAAAATAAAGATTTTCGTCGACACTTG AGGAGTTATTGACAGAGTAAGG 3′
CTD de Brd2	CTDBrd2_F1	5′ TTTCAAGGTCATCATCATCATCATCATCATGATTCAGAGGAGGA GGAAGA 3′
	CTDBRD2_R1	5′ CTCGAGTGCGGCCGCTTATTAAGCGCTAAGGCGTGA 3′

Tableau 2. Récapitulatif des amorces nécessaires à la construction « multiTEV ». Séquences nucléotidiques de l'étiquette Strep-Tag®II (vert), de l'étiquette poly-histidine (bleu) et des sites de clivage à la protéase du TEV (orange).

Il est important de noter que certaines amorces sont complémentaires entre elles au niveau de leurs extrémités. En effet, les amorces TEV_R1 et CTDPERV_F1 sont complémentaires au niveau de leurs extrémités 5'. Il en est de même pour les amorces CTDPERV_F2 et CTDBrd2_F1. Ces séquences complémentaires serviront à la formation d'un insert contenant les trois gènes d'intérêt, décrit dans le chapitre suivant.

2. Fabrication du plasmide recombinant « multiTEV »

1. Amplification des gènes d'intérêt

Les gènes codant la protéase du TEV, le CTD de l'IN de PERV-A/C et le CTD de Brd2 sont apportés respectivement par les plasmides pRK793b, pGEM-T Easy et pPROEX-Hta, fournis par l'équipe du Dr Ronfort (INRA Lyon). Ils sont amplifiés par une réaction en chaîne par polymérase (*Polymerase Chain Reaction* ou PCR) (Figure 33).

Les amorces sens (TEV_F1) et anti-sens (TEV_R1) sont nécessaires pour l'amplification du gène de la protéase du TEV. L'amorce TEV_F1 apporte le site de restriction *NcoI*, tandis que l'amorce TEV_R1 contient les séquences qui codent le premier site TEV et une partie de l'étiquette Strep-tag®II.

Les amorces CTDPERV_F1 et R1 sont utilisées pour amplifier le gène du CTD de l'IN de PERV-A/C. L'amorce sens CTDPERV_F1 porte les séquences codant une partie du premier site TEV et la totalité de l'étiquette Strep-tag®II. Les séquences correspondant au second site TEV et à une partie de l'étiquette poly-histidine sont contenues dans l'amorce anti-sens CTDPERV_R1.

Les amorces nucléotidiques CTDBrd2_F1 et R1 sont utilisées pour amplifier le gène codant le CTD de Brd2. L'amorce sens CTDBrd2_F1 apporte des séquences qui codent une partie du second site TEV et la totalité de l'étiquette poly-histidine. Enfin, l'amorce anti-sens CTDBrd2_R1 introduit le site de restriction *NotI*.

En pratique, pour chaque fragment, le mélange réactionnel contient 10 ng d'ADN matrice (plasmide recombinant), une unité d'ADN polymérase ADN-dépendante thermostable (Mullis *et al.*, 1986) (Phusion® Haute Fidélité, *New England Biolabs*), 10 µl de tampon 5X de cette enzyme (*New England Biolabs*), 1 µl d'un mélange de désoxyribonucléotides (dNTP, *New England Biolabs*) et 0,5 µM de chaque amorce nucléotidique (*Sigma*). Le mélange est porté à un volume total de 50 µl par ajout d'eau distillée stérile (*Otec*). La PCR est divisée en trois étapes principales. La première étape dite de dénaturation, à une température de 95°C, conduit à l'obtention d'ADN monocaténaires à partir de l'ADN matriciel bicaténaire. La diminution de la température (jusqu'à atteindre un palier compris entre 50°C et 70°C) entraîne l'hybridation des amorces nucléotidiques sur ces ADN monocaténaires. Enfin, l'élongation de l'ADN par l'ADN polymérase ADN-dépendante thermostable a lieu à 72°C. La répétition de ce cycle en trois étapes permet une amplification exponentielle de l'ADN.

Après amplification, les gènes d'intérêt portent des séquences communes apportées par les amorces nucléotidiques utilisées (Figure 33).

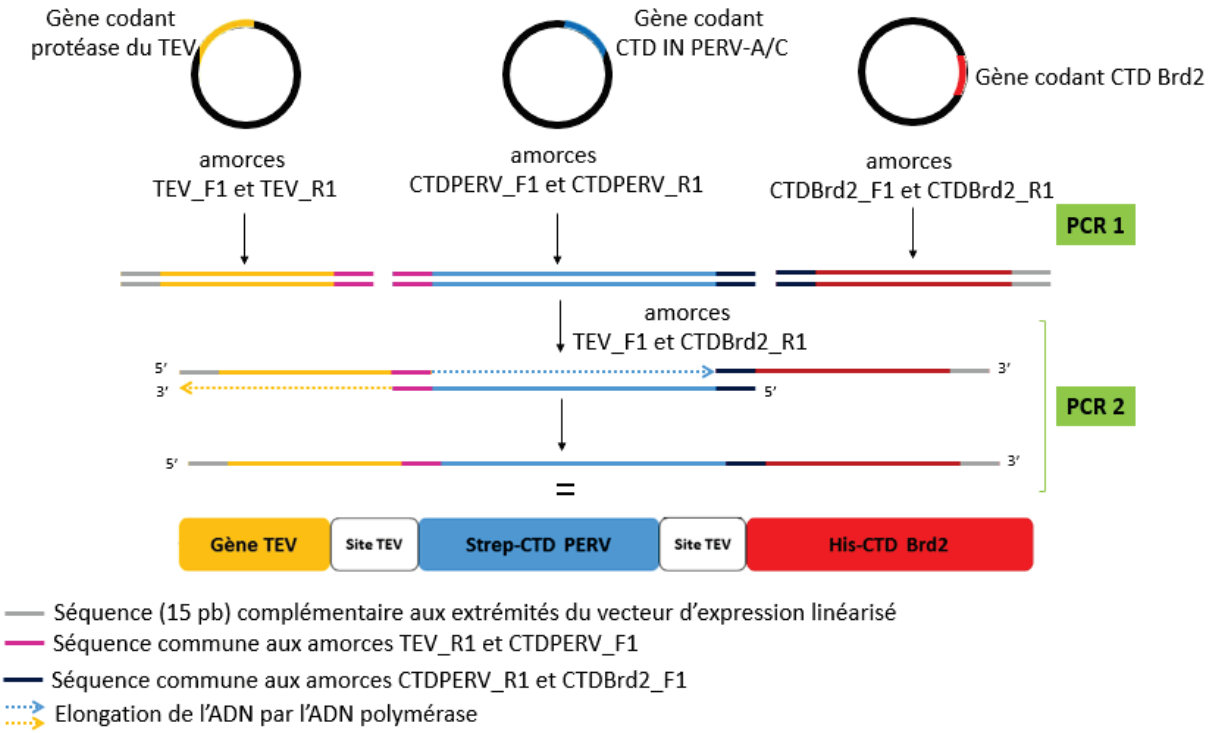


Figure 33. Schéma des cycles d'amplification et des amorces nécessaires à l'obtention de l'insert de la construction « multiTEV ».

2. Formation de l'insert

Une nouvelle PCR est nécessaire pour obtenir l'insert d'intérêt (Figure 33). Cette PCR utilise les amorces complémentaires des extrémités de l'insert désiré, à savoir l'amorce sens TEV_F1 et l'amorce anti-sens CTDBrd2_R1. Le protocole de PCR est identique à celui décrit dans le paragraphe précédent. Les trois gènes précédemment amplifiés jouent le rôle de matrice. Ils portent des séquences communes (apportées par les amorces nucléotidiques) qui vont permettre l'hybridation et l'élongation de l'ADN (Figure 33). Le produit de PCR obtenu est alors formé des trois gènes d'intérêt, placés dans l'ordre attendu et séparés entre eux par des sites TEV (Figure 32). Ce produit de PCR est directement cloné dans le vecteur pET28a préalablement digéré avec le système In-Fusion® (*Clontech*).

3. Clonage avec le système In-Fusion®

Le principe de ce système de clonage repose sur la complémentarité de 15 paires de bases entre les extrémités de l'insert (apportées par les amorces TEV_F1 et CTDBrd2_R1) et celles du vecteur linéarisé. Cette homologie permet la fusion de ces séquences nucléotidiques et l'obtention du plasmide recombinant d'intérêt (Figure 34). Ainsi, nous nous affranchissons de l'étape de ligation.

En pratique, le vecteur pET28a est digéré extemporanément par les enzymes de restriction *NcoI* et *NotI* (*New England Biolabs*) dans un volume final de 20 µl. Il contient 2 µl de tampon de digestion 3.1 (1X final, *New England Biolabs*), 2 µg de vecteur pET28a et une unité de chaque enzyme de restriction par µg d'ADN à digérer. Le mélange est placé une nuit à 37°C, avant de contrôler et de purifier le vecteur digéré sur gel d'agarose.

Le kit In-Fusion® HD Cloning est ensuite utilisé. Il s'agit d'un prémix contenant une ADN polymérase ADN-dépendante haute fidélité, appelée enzyme In-Fusion®. Un volume total de 10 µl comprenant un rapport molaire insert:vecteur de 2:1 est incubé en présence du prémix In-Fusion® (1X final), dans un bain-marie à 50°C durant 15 minutes. Une transformation en DH5α (*Lucigen*) est ensuite réalisée pour amplifier les produits réactionnels et isoler les clones contenant la construction « multiTEV » (Figure 34).

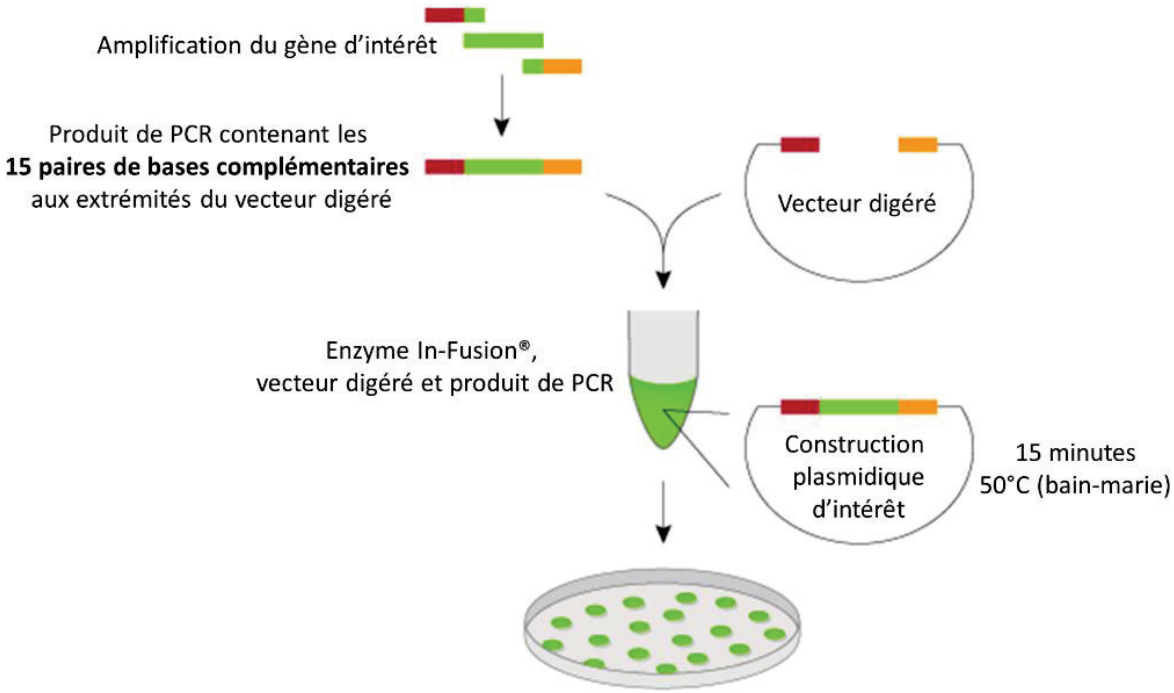


Figure 34. Principe général du clonage In-Fusion® (*Clontech*).

3. Transformation de bactéries compétentes

Le plasmide d'intérêt est introduit dans des bactéries *Escherichia coli* (*E. coli*) rendues compétentes par traitement chimique.

En pratique, 50 µl de bactéries compétentes sont mélangés à 10 ng d'ADN plasmidique d'intérêt. Dans le cas d'un clonage In-Fusion®, la quantité d'ADN plasmidique n'étant pas connue, les bactéries sont mises en présence du mélange de 10 µl contenant l'insert, le vecteur

et l'enzyme In-Fusion®. Le mélange est alors placé dans la glace pendant 30 minutes. Les bactéries sont ensuite incubées durant 45 secondes à 42°C, puis 2 minutes dans la glace afin de réaliser un choc thermique. Ce changement brutal de température fragilise les bactéries, entraînant ainsi l'incorporation du plasmide d'intérêt. Un volume de 200 µl de milieu de culture liquide Luria-Bertani (LB, *Sigma*) est ajouté aux cellules. Les bactéries transformées sont incubées à 37°C, pendant une heure, sous agitation (180 rpm). Les bactéries sont enfin incubées une nuit à 37°C, après étalement sur un milieu de culture solide LB-agar contenant les antibiotiques appropriés pour la sélection des transformants.

4. Extraction de l'ADN

Une extraction par lyse alcaline avec un kit commercial (NucleoSpin® Plasmid, *Macherey-Nagel*) permet de récupérer les ADN plasmidiques d'intérêt des bactéries transformées (Birnboim *et al.*, 1979). Ces ADN pourront être séquencés et/ou utilisés pour transformer une nouvelle souche bactérienne.

En pratique, des bactéries DH5α porteuses du plasmide d'intérêt sont cultivées dans un milieu de culture liquide LB, contenant l'antibiotique approprié à la sélection de ce plasmide. Le volume de milieu utilisé dépend du taux de réplication du plasmide d'intérêt (5 à 10 ml pour un faible taux de réplication, 1 à 5 ml pour un fort taux de réplication). Après une incubation une nuit à 37°C, le milieu est centrifugé et l'ADN est extrait selon le protocole du fournisseur.

5. Dosage de l'ADN

La quantité et la qualité de l'ADN sont contrôlées par spectroscopie d'absorption dans l'ultraviolet, avec un spectrophotomètre Nanodrop 2000 (*Thermo Fisher Scientific*). Les bases de l'ADN, à caractère aromatique, ont la capacité d'absorber la lumière ultraviolette (240-280 nm) avec un maximum à 260 nm.

Ainsi, la concentration en ADN est déterminée par la loi de Beer-Lambert :

$$A_{260} = \varepsilon \times L \times c$$

avec A_{260}, la valeur d'absorbance à 260 nm ; ε le coefficient d'extinction molaire en $M^{-1}.cm^{-1}$; l, la longueur du trajet optique en cm^{-1} ; et c, la concentration en $mol.L^{-1}$.

La pureté d'une solution d'ADN est estimée par la mesure du rapport d'absorbance à 260 et 280 nm A_{260}/A_{280}. Sa valeur est de l'ordre de 1,8 pour un ADN double brin. Un échantillon

d'acides nucléiques contaminé par des protéines aura un rapport A_{260}/A_{280} inférieur à 1,8. À l'inverse, la contamination de cet échantillon par de l'ARN se traduira par un rapport A_{260}/A_{280} supérieur à 1,8.

6. Séquençage de l'ADN

La composition en bases azotées de l'ADN est déterminée par la méthode de Sanger (Sanger *et al.*, 1977), dont le principe repose sur l'utilisation de didésoxyribonucléotides (ddNTP) marqués. Chaque ddNTP est porteur d'un fluorophore spécifique, capable d'émettre un signal fluorescent à une longueur d'onde donnée.

En pratique, une PCR est réalisée avec une solution composée de l'ADN matrice, d'une amorce spécifique, d'une ADN polymérase, de dNTP non marqués et de ddNTP marqués. L'incorporation des ddNTP, lors de l'amplification, conduit à l'interruption de l'élongation et à un marquage spécifique du brin d'ADN par le fluorophore du dernier ddNTP, responsable de l'arrêt de l'élongation. Les produits d'amplification sont dénaturés et séparés par électrophorèse capillaire. Une analyse automatique permet l'identification des fluorophores successifs et ainsi la reconstitution de la séquence ADN.

La société *GATC Biotech* a réalisé l'ensemble des séquençages d'ADN nécessaire à mon travail de thèse.

7. Souches bactériennes utilisées

Après avoir vérifié l'exactitude de la séquence de l'ADN qui a servi à transformer des bactéries, nous pouvons stocker ces dernières ou les utiliser pour surproduire les protéines d'intérêt. Les souches bactériennes utilisées durant mes travaux de recherche sont toutes des souches *E. coli* modifiées.

1. Stockage des bactéries transformées avec l'ADN d'intérêt

La souche bactérienne DH5α (*Lucigen*) est porteuse de mutations améliorant la stabilité des vecteurs recombinants (Taylor *et al.*, 1993). Elle permet de produire et de stocker les bactéries contenant les ADN plasmidiques d'intérêt. En pratique, les bactéries sont cultivées sur un milieu de culture solide LB-agar (*Sigma*), contenant l'antibiotique approprié à la sélection du plasmide d'intérêt. En parallèle, la souche DH5α sert à la réalisation de stocks

glycérol : il s'agit de bactéries porteuses du plasmide d'intérêt, conservées dans du glycérol à -80°C, et qui peuvent ainsi être stockées plusieurs années.

Pour fabriquer un stock glycérol, les bactéries sont cultivées à 37°C une nuit, à 180 rpm, dans un milieu de culture liquide LB contenant l'antibiotique approprié. Un volume final de 1 mL contenant les bactéries et le glycérol (20% (v/v), *Sigma*) est placé dans un cryotube de 2 mL (*Greiner Bio-One*). Les aliquots sont stockés à -80°C.

2. Surproduction des protéines recombinantes

La souche bactérienne BL21(DE3) (*Lucigen*) est utilisée pour l'expression des plasmides recombinants. Elle est déficiente en protéases Ion3 et ompT, responsables de la dégradation des protéines produites. Elle contient le lysogène DE3 placé sous le contrôle du promoteur de l'opéron lactose *lacUV5*. Ce lysogène code l'ARN polymérase ADN-dépendante du bactériophage T7. L'ajout de lactose ou d'un analogue non hydrolysable comme l'IsoPropylThioGalactoside (IPTG, *Euromedex*) déclenche une levée d'inhibition du répresseur LacI sur l'opérateur de l'opéron lactose, ce qui entraîne la synthèse de l'ARN polymérase T7 par l'ARN polymérase ADN-dépendante de la bactérie.

Dans le cas de la surproduction d'une protéine recombinante, une stratégie développée par Studier est utilisée en routine au laboratoire. Elle consiste à introduire dans une bactérie, un plasmide porteur du gène d'intérêt placé sous le contrôle du promoteur reconnu par l'ARN polymérase T7 (Studier, Moffatt, 1986). Ce gène remplace les trois gènes (*LacZ, LacY et LacA*) de l'opéron lactose. L'ajout de lactose ou d'IPTG conduit à la synthèse de l'ARN polymérase T7 qui assure alors la transcription du gène d'intérêt.

La souche bactérienne BL21(DE3)pLysS contient, en plus des éléments précédemment décrits, un plasmide pLysS porteur d'un gène de résistance au chloramphénicol. Ce plasmide code le lysozyme T7 qui inhibe l'activité basale de l'ARN polymérase T7. Cette inhibition empêche l'expression basale du gène d'intérêt avant induction, ce qui peut être crucial pour la surproduction d'une protéine potentiellement toxique pour la bactérie (Figure 35).

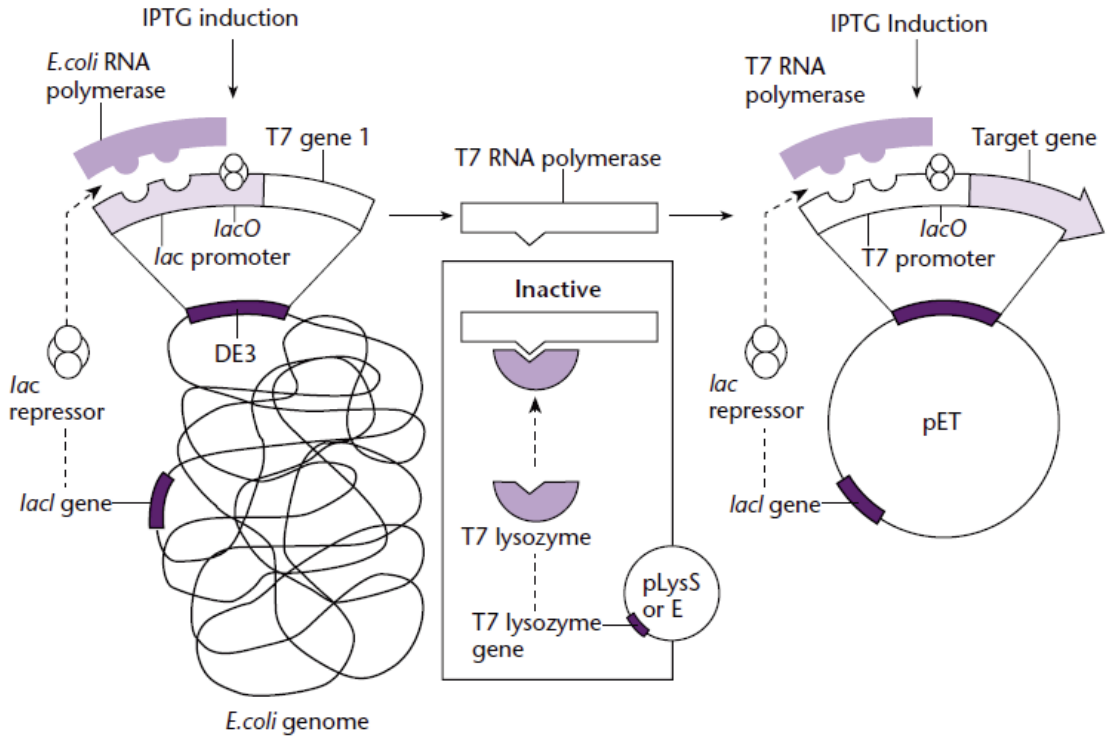


Figure 35. Stratégie de régulation de l'expression des gènes clonés dans un vecteur pET, selon le système de Studier (Primrose *et al.*, 2006).

II. Surproduction et purification des protéines d'intérêt

1. Culture bactérienne et induction

Après transformation des bactéries BL21(DE3) ou BL21(DE3)pLysS par le plasmide d'intérêt, celles-ci sont placées dans un milieu de culture liquide LB en présence de l'antibiotique correspondant. Cette préculture est incubée une nuit à 37°C, sous agitation (180 rpm). Pour certains essais de surproduction protéique, du *Terrific Broth* (TB, *Sigma*) est utilisé à la place du LB pour cultiver les bactéries.

Après incubation, la préculture est placée dans un plus grand volume de milieu (100 fois supérieur) contenant l'antibiotique approprié, à 37°C sous agitation (180 rpm). La croissance bactérienne est suivie avec un spectrophotomètre, par mesure de la turbidité du milieu à une longueur d'onde de 600 nm. La phase exponentielle de croissance est atteinte pour une absorbance comprise entre 0,5 et 0,9. La surproduction protéique est alors induite par ajout d'IPTG (0,3 à 1 mM final, *Euromedex*) dans le milieu de culture. La concentration finale d'IPTG, la température et le temps d'induction varient selon la protéine à surproduire. Après induction, les bactéries sont sédimentées par centrifugation à 6 000 × g, à 4°C pendant 10 minutes. Le surnageant est éliminé, le culot bactérien est stocké à -20°C en attendant l'étape de lyse bactérienne.

2. Lyse bactérienne

La lyse bactérienne permet d'extraire les protéines des cellules bactériennes. Cette extraction se fait dans un tampon de lyse approprié, dont la composition varie en fonction de la protéine d'intérêt. En pratique, 20 ml de tampon de lyse est utilisé pour resuspendre un culot correspondant à un litre de culture. Du lysozyme (0,2 mg/ml, *Sigma*), de la DNase I (20 µg/ml final, *Sigma*), du $MgCl_2$ nécessaire à l'activité de la DNase I (10 mM final, *Sigma*) et un cocktail d'inhibiteurs de protéases 1X (cOmplete ULTRA *EthyleneDiamine Tetraacetic Acid* EDTA-free, *Roche*) sont ajoutés au tampon de lyse. La DNase I dégrade l'ADN bactérien libéré au cours de la lyse, qui elle-même est facilitée par l'action du lysozyme. Le cocktail d'anti-protéases inhibe l'action des protéases bactériennes et évite ainsi la dégradation des protéines d'intérêt. Ce cocktail ne contient pas d'EDTA, ce qui évite d'éventuelles interférences avec la colonne d'affinité lors de la purification protéique.

Les bactéries resuspendues sont détruites par sonication, à l'aide de l'appareil Sonifier 450 (*Branson*). Cette méthode de lyse mécanique utilise une sonde qui émet des ondes ultrasonores. Une fois la sonde immergée dans la solution de bactéries resuspendues, les ultrasons provoquent une augmentation de la taille de bulles microscopiques, qui conduit à leur implosion. Ce phénomène a pour conséquence une augmentation de la pression et de la température locales, responsable de la destruction des cellules en solution. Plusieurs cycles d'une minute de sonication/une minute de refroidissement dans la glace sont réalisés pendant une durée totale de 8 minutes (puissance 40%).

Suite à la lyse bactérienne, une centrifugation de 45 minutes à 14 000 × g, à 4°C permet d'éliminer les débris cellulaires (fraction insoluble) dans le culot. Le surnageant correspondant à la fraction soluble contient, entre autres, la protéine d'intérêt surproduite. Il est filtré avec un filtre de porosité de 0,45 µm (*VWR International*). Les fractions soluble (lysat) et insoluble sont conservées à 4°C.

3. Purification de la protéine d'intérêt

L'étape de purification permet d'éliminer les contaminants protéiques et de purifier la protéine d'intérêt. Les systèmes de chromatographie décrits dans ce chapitre ont été couplés à un système de chromatographie liquide des protéines à basse pression (*Fast Protein Liquid Chromatography* ou FPLC, système *Aktä purifier, GE Healthcare*).

1. Chromatographie d'affinité pour le nickel

Le système de chromatographie d'affinité pour le métal (*Immobilized Metal Ion Affinity Chromatography* ou IMAC) et particulièrement pour le nickel, a été développé par l'équipe de Porath (Porath *et al.*, 1975).

Lors d'une chromatographie d'affinité pour le nickel, les résidus histidyl de l'étiquette de purification poly-histidine établissent des liaisons de coordination avec les ions nickel (Figure 36). Ces ions sont immobilisés sur un support solide (résine), par l'intermédiaire de bras espaceurs constitués d'acide nitrilotriacétique (*NitriloTriacetic Acid* ou NTA, Protino® Ni-NTA 5 mL, *Macherey Nagel*). Chaque ion Ni^{2+} établit six liaisons de coordination. Deux d'entre elles relient l'ion Ni^{2+} aux atomes d'azote du cycle imidazole de la chaîne latérale de deux résidus d'histidine. Les quatre liaisons restantes assurent l'attachement de l'ion Ni^{2+} au bras NTA. La protéine d'intérêt fixée est ensuite éluée par ajout d'un compétiteur en excès dans le tampon d'élution : l'imidazole libre. L'EDTA peut également être utilisé pour chélater les ions nickel de la résine et éluer la protéine.

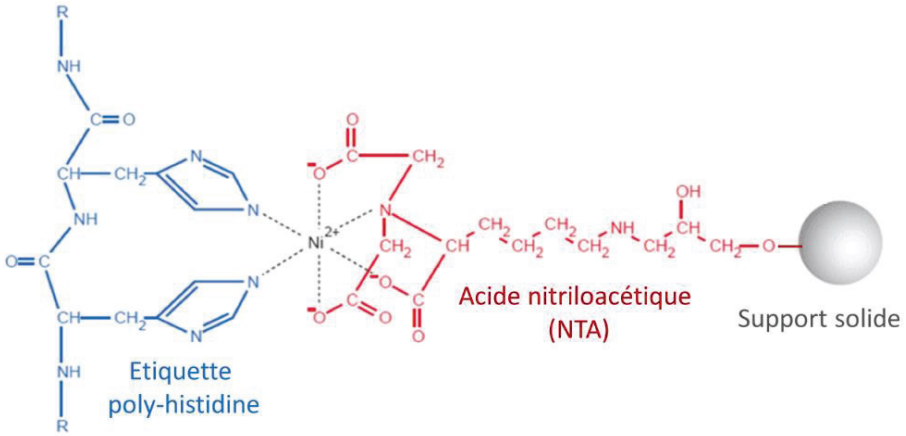


Figure 36. Liaison de l'étiquette poly-histidine d'une protéine recombinante à une colonne Ni-NTA, utilisée dans le système IMAC (d'après http://www.ou.edu/OpenEducation/ou-resources/biochemical-methods/lab-8/).

Les tampons employés étant différents selon la protéine à purifier, leurs compositions seront décrites dans le chapitre 'Résultats et Discussion'.

En pratique, la purification du complexe protéique est réalisée à partir du surnageant filtré par FPLC, sur le système Äkta Purifier (*GE Healthcare*). Le surnageant filtré (environ 20 mL) est chargé sur la colonne Ni-NTA préalablement équilibrée avec le tampon de lyse. La présence d'imidazole, en faible quantité dans ce tampon, permet de limiter les fixations aspécifiques. Cette étape permet d'éluer les protéines non fixées à la colonne. Celle-ci est ensuite lavée avec un tampon de lavage plus stringent que le tampon de lyse. Cette étape de lavage permet d'éluer

les protéines et acides nucléiques fixés de façon aspécifique à la colonne. La colonne est de nouveau lavée avec le tampon de lyse. L'élution est ensuite réalisée par un gradient croissant d'imidazole, par mélange progressif du tampon de lyse et du tampon d'élution. Le suivi de de la valeur A_{280} des fractions d'élution permet de détecter la sortie des protéines de la colonne de purification. Les fractions d'élution récupérées sont analysées par électrophorèse en gel discontinu de polyacrylamide, en conditions dénaturantes (*Sodium Dodecyl Sulfate – PolyAcrylamide Gel Electrophoresis* ou SDS-PAGE).

2. Chromatographie d'affinité pour la streptavidine

L'étiquette de purification Strep-tag® II est un peptide synthétique formé de huit résidus d'acides aminés (WSHPQFEK). Il présente une affinité pour une résine sur laquelle est fixée une streptavidine modifiée, la Strep-Tactin® (Strep-Tactin® Superflow® cartridge 5mL, *IBA*, Schmidt *et al.*, 2007). Les modifications appliquées sur la streptavidine (Strep-Tactin®) lui assurent une plus grande stabilité et une affinité plus élevée pour l'étiquette Strep-tag® II (100 fois plus que la streptavidine native). L'étiquette Strep-tag® II se lie à la streptavidine modifiée, au niveau du site de fixation de son substrat naturel, la biotine. Cette interaction réversible est rompue par ajout de biotine libre ou d'un analogue comme la desthiobiotine. L'élimination de la desthiobiotine par l'acide benzoïque 2-(4-hydroxy-azobenzène) ou HABA permet la régénération de la colonne Strep-Tactin® (Figure 37).

En pratique, le lysat est appliqué sur la colonne Strep-Tactin®. Cette colonne est lavée avec un tampon de lavage W (100 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA) pour éliminer les protéines liées de façon aspécifique à la colonne. La protéine d'intérêt fixée est ensuite éluée par compétition, avec le tampon d'élution E contenant la desthiobiotine (100 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 2,5 mM desthiobiotine). Après élution, la colonne de purification est régénérée en ajoutant le tampon de régénération R (100 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 1 mM HABA). Une étape de lavage avec le tampon W permet d'éluer le composé HABA et d'utiliser à nouveau la colonne.

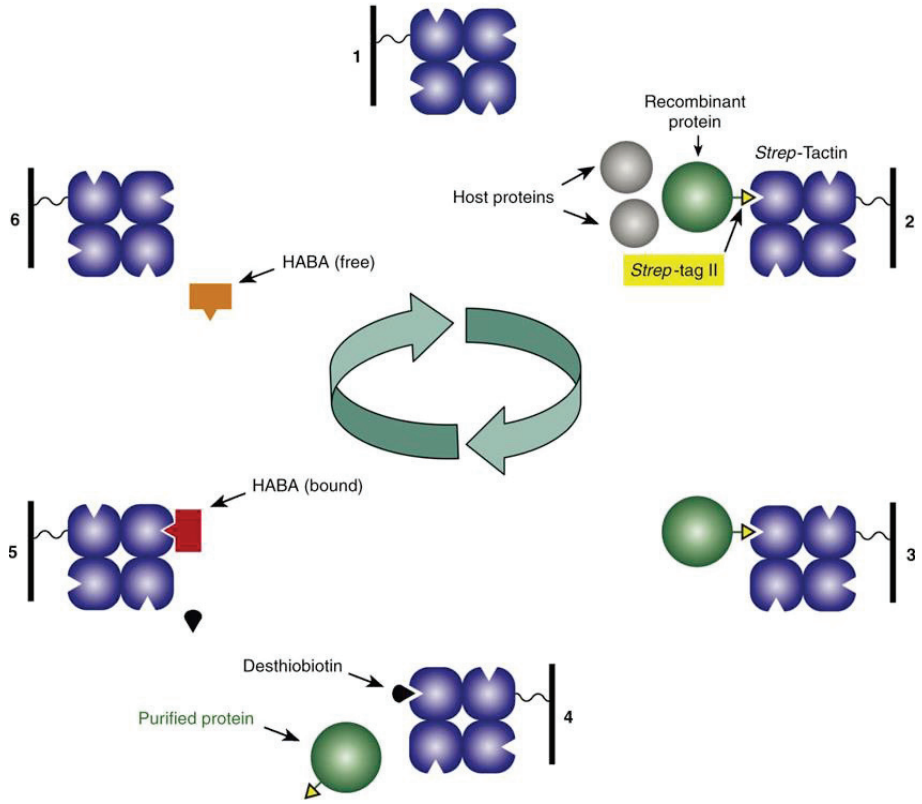


Figure 37. Cycle de purification d'une protéine avec étiquette Strep-tag® II avec la colonne Strep-Tactin® (Strep-Tactin® Superflow® cartridge, *IBA*, Schmidt *et al*., 2007).

3. Chromatographie d'exclusion de taille

La chromatographie d'exclusion de taille (*Size Exclusion Chromatography* ou SEC) est une technique permettant la séparation des macromolécules en fonction de leur masse moléculaire. Elle a lieu après la purification des macromolécules par chromatographie d'affinité. Une colonne de type Sephadex contenant des polyosides bactériens (dextran et agarose) est utilisée. Des granules de gel poreux constituent la phase solide de ce système de chromatographie. Les macromolécules ayant un diamètre supérieur aux pores du gel sont directement éluées. À l'inverse, les macromolécules de diamètre inférieur sont retenues en passant à travers ces pores et éluées plus tardivement. Il existe une relation linéaire entre le volume d'élution des macromolécules et le logarithme de leurs masses moléculaires. La calibration de la colonne avec des macromolécules de masses moléculaires connues permet d'estimer l'état oligomérique de la protéine analysée. Deux colonnes Superdex aux gammes de séparation différentes sont utilisées dans ces travaux. Les colonnes Superdex 200 HR 10/30 et 75 HR 10/30 (*GE Healthcare*) permettent respectivement une séparation dans la gamme 10-600 kDa et 3-70 kDa d'après le fournisseur.

En pratique, la colonne est lavée et préalablement équilibrée avec le tampon d'élution ou le tampon final de la protéine étudiée. Afin d'améliorer la résolution de séparation, l'échantillon protéique est injecté dans le plus petit volume manipulable possible. L'élution des protéines est suivie par mesure de la valeur A_{280}. Les valeurs de volume d'élution correspondant à chaque pic sont comparées à celles des standards de masses moléculaires connues.

4. Dosage des protéines

Comme l'ADN, la quantité et la qualité des protéines sont contrôlées par spectroscopie d'absorption ultraviolet-visible, avec le spectrophotomètre Nanodrop 2000 (*Thermo Fisher Scientific*). Les protéines absorbent la lumière ultraviolette (175 à 350 nm). Cette faculté d'absorption est apportée principalement par les groupements carbonyles des liaisons peptidiques (190 à 230 nm), les résidus de tryptophane et tyrosine (250 à 320 nm) et, dans une moindre mesure, les ponts disulfures ou cystines (250 à 300 nm). Les coefficients d'extinction molaire des résidus de tryptophane, tyrosine et des cystines correspondent respectivement à 5 500, 1 490 et 125 $M^{-1}.cm^{-1}$ à une longueur d'onde de 280 nm. Ainsi, le coefficient d'extinction molaire $\varepsilon_{\text{molaire}}$ d'une protéine (dénaturée ou correctement repliée) peut être déterminé selon l'équation suivante :

$$\varepsilon = 5\,500.n_w + 1\,490.n_y + 125.n_c$$

avec n_w, n_y et n_c, les nombres de résidus de tryptophane, tyrosine et de cystine, respectivement et ε, le coefficient d'extinction molaire en $M^{-1}.cm^{-1}$ d'une protéine (Pace *et al.*, 1995).

La loi de Beer-Lambert, décrite précédemment pour le dosage de l'ADN (*cf.* § I.6), permet de déterminer la concentration protéique d'un échantillon, par mesure de la valeur A_{280}. À cette longueur d'onde se trouve le maximum d'absorption des résidus W, Y et des cystines. Le rapport A_{260}/A_{280} permet d'évaluer la pureté d'une solution protéique. Un échantillon contaminé par des acides nucléiques se caractérise par un rapport A_{260}/A_{280} supérieur à 0,6.

5. Concentration des protéines

L'augmentation de la concentration des protéines dans un échantillon peut s'avérer nécessaire pour certaines applications. Elle est réalisée en utilisant un système d'ultrafiltration (concentrateur) qui comprend une membrane semi-perméable. Dans cette étude, le concentrateur est formé d'une membrane en PolyEtherSulfone ou PES (*Vivaspin*). Cette membrane autorise la séparation des macromolécules en fonction de leur masse moléculaire.

Un seuil de coupure adéquat sert à retenir la protéine d'intérêt et ainsi à augmenter la concentration de l'échantillon. Des cycles de centrifugation à 4 000 × g et à 6°C sont effectués jusqu'à obtenir la concentration protéique désirée. Ce système d'ultrafiltration sert parfois aux échanges de tampons, bien que le dessalage soit le plus souvent utilisé dans cette étude.

6. Echange de tampons par dessalage

Il est parfois nécessaire de modifier la composition d'une solution tampon contenant la protéine d'intérêt. Pour échanger des tampons, nous avons recours à un système de dialyse ou à une chromatographie d'exclusion de taille, dont le principe a été décrit précédemment (*cf.* § II.1.3). Cependant, dans ce cas, une colonne de type Sephadex HiPrep 26/10 (*GE Healthcare*) est utilisée afin de séparer les macromolécules de plus de 5 kDa des petites molécules inférieures à 1 kDa. En pratique, cette colonne est lavée à l'eau avant d'être équilibrée dans le tampon final désiré. L'échantillon est ensuite injecté, les macromolécules supérieures à 5 kDa sont éluées en premier et récupérées, tandis que les molécules inférieures à 1 kDa sortent de la colonne en dernier. L'élution des protéines est suivie par mesure de la valeur A_{280}.

7. Clivage par la protéase du TEV

Pour éviter que l'étiquette de purification n'interfère avec les étapes de cristallisation, celle-ci peut être éliminée par digestion enzymatique. L'enzyme agit sur son site de clivage, présent entre l'étiquette de purification et la protéine d'intérêt. Dans cette étude, seuls les plasmides contenant un site de clivage à la protéase du TEV, sont porteurs d'une séquence codant l'étiquette poly-histidine. La protéase du TEV est plus stable et a une spécificité de séquences plus élevée que la thrombine. Cette protéase à sérine reconnait préférentiellement les séquences ENLYFQG et ENLYFQS et les clive entre les résidus Q et G/S (Carrington *et al.*, 1988). Elle est surproduite en fusion avec une étiquette poly-histidine au laboratoire.

En pratique, la protéase du TEV est incubée avec la solution protéique d'intérêt, sous agitation douce. Les rapports massiques TEV/protéine, la température et le temps d'incubation sont propres à chaque protéine de fusion à cliver. L'efficacité de clivage est contrôlée par migration sur SDS-PAGE. Après digestion, l'étiquette poly-histidine et la protéase du TEV (porteuse de cette même étiquette) sont éliminées par chromatographie d'affinité pour le nickel.

III. Analyse qualitative et quantitative des protéines d'intérêt

1. Techniques biochimiques

1. Electrophorèse discontinue en conditions dénaturantes

La technique de SDS-PAGE permet de séparer les protéines sous l'effet d'un champ électrique. Cette technique, décrite en 1964 (Ornstein, 1964), a évolué avec l'ajout du détergent anionique SDS (Laemmli, 1970) qui dénature les protéines et les charge négativement. De cette façon, les protéines migrent de la cathode vers l'anode et sont séparées uniquement selon leurs poids moléculaires. Le gel de polyacrylamide est dit discontinu car il est constitué de deux parties : un gel de concentration (partie supérieure) et un gel de séparation (partie inférieure). Le gel de concentration permet l'obtention de zones fines, chargées en protéines, qui migrent ensuite dans le gel de séparation. Ces deux gels diffèrent par leur pH et leur concentration en acrylamide/bis-acrylamide 30% (w/w) (*Euromedex*). Les taux d'acrylamide et de bis-acrylamide déterminent la densité du gel et donc la gamme de séparation des protéines. Plus leur concentration est élevée, plus les pores du gel sont petits. La réaction de polymérisation de l'acrylamide nécessite la présence de radicaux libres. Ces derniers sont obtenus par l'ajout de persulfate d'ammonium (*Ammonium PerSulfate* ou APS, *Sigma*) et de N,N,N',N'-tétraméthyl-éthylènediamine (TEMED, *Euromedex*). Le TEMED catalyse la décomposition des ions persulfate et permet ainsi la libération des radicaux libres. Dans cette étude, plusieurs pourcentages de gels sont utilisés selon les protéines à séparer (Tableau 3).

	Gel de concentration		Gel de séparation
	4%		12%
Acrylamide/bis-acrylamide 30%	660 µL		4 ml
1 M Tris HCl pH 6,8	625 µL	3 M Tris HCl pH 8,8	1,25 ml
Eau distillée	3,48 mL		4,24 ml
SDS 10% (w/v)	200 µL		400 µl
APS 10% (w/v)	50 µL		100 µl
TEMED	5 µL		10 µl

Tableau 3. Composition des gels de concentration et de séparation pour deux gels discontinus de polyacrylamide.

Avant d'être déposés sur le gel, les échantillons sont chauffés 5 minutes à 95°C dans du tampon Laemmli 1X final (65,8 mM Tris-HCl pH 6,8, glycérol 26,3% (w/v), SDS 2,1% (w/v), 0,01% bleu de bromophénol (w/v), *BioRad*) . Certains échantillons ne sont pas chauffés avant d'être déposés sur le gel : les protéines qu'ils contiennent se dégradent sous l'effet de la chaleur. Une solution de marqueurs de masses moléculaires connues (Protein Ladder Plus, *Euromedex*) est également déposée sur le gel. Elle permet d'estimer le poids moléculaire de la protéine d'intérêt. Un champ électrique est appliqué dans un tampon Tris Glycine-SDS pH 8,5 (25 mM Tris Base, 192 mM glycine, SDS 0,1% (w/v), *Euromedex*), pendant 10 minutes à 90 V, puis à 140 V jusqu'à ce que le front de migration atteigne la limite du gel.

Coloration au bleu de Coomassie

Après migration, le gel est rincé dans de l'eau distillée, puis plongé dans une solution commerciale contenant du bleu de Coomassie colloïdal G-250 (PageBlue™ Protein Staining Solution, *ThermoFisher Scientific*) durant une heure. Le bleu de Coomassie a la capacité de lier les résidus d'acides aminés basiques et aromatiques (seuil de détection de 5 ng). Deux lavages de 10 minutes dans l'eau distillée sont nécessaires pour décolorer le gel. Cette coloration permet de vérifier la présence et la pureté de la protéine d'intérêt et d'estimer sa quantité.

2. *Western-blot*

Cette méthode, également connue sous le nom d'immuno-empreinte, permet de détecter et d'identifier des protéines dans un échantillon biologique, en utilisant des anticorps spécifiques (Mahmood *et al.*, 2012). Après séparation des protéines par SDS-PAGE, une membrane de nitrocellulose (Amersham Protran 0.2 NC, *GE Healthcare*) est placée face au gel. Sous l'effet d'un courant électrique à ampérage constant (400 mA), les protéines sont transférées durant 45 minutes sur cette membrane dans du tampon 10 mM CAPS pH 11 (acide 3-(cyclohexylamino)-1-propanesulfonique, *Sigma*)/méthanol 10% (v/v) (*Carlo Erba*). Les protéines migrent du gel vers la membrane en conservant leurs positions relatives. Des liaisons hydrophobes et ioniques sont impliquées dans la fixation des protéines sur la membrane.

Une étape de saturation est ensuite effectuée pour minimiser les interactions aspécifiques entre la membrane et les anticorps. Elle consiste à plonger la membrane de nitrocellulose dans du tampon de saturation PBS 1X (*Euromedex*)/Tween20 0,1% (v/v) (*Sigma*)/lait écrémé 5% (v/v) (*Régilait*), pendant une heure à 20°C. Ainsi, les sites libres de la membrane sont saturés en protéines aspécifiques contenues dans le lait.

Pour détecter les protéines portant une étiquette poly-histidine, un anticorps monoclonal de souris (1 µg/ml, *Sigma*) dirigé contre cette étiquette est mis en présence de la membrane, pendant une heure à 20°C dans du tampon de saturation. La membrane est rincée trois fois avec du tampon PBS 1X/Tween20 0,1% (v/v) pour éliminer les anticorps primaires non liés à la membrane. Puis, la membrane est incubée avec un anticorps secondaire de chèvre anti-souris couplé à la phosphatase alcaline (1 µg/ml final, *Sigma*) dans du tampon PBS1X/Tween20 0,1% (v/v), pendant une heure à 20°C, puis rincée plusieurs fois avec du tampon PBS1X/Tween20 0,1% (v/v) puis avec du PBS 1X avant révélation.

Pour une protéine avec étiquette Strep-tag® II, l'anticorps primaire est remplacé par de la Strep-Tactin® couplée directement à la phosphatase alcaline et mise en contact dans les conditions décrites ci-dessus pour l'anticorps primaire anti-poly-histidine. Trois rinçages avec du tampon PBS 1X/Tween20 0,1% (v/v) sont réalisés pour retirer l'excès de streptavidine marquée avant révélation.

Dans tous les cas, la révélation se fait par détection colorimétrique avec le substrat chromogénique commercial BCIP-NBT (5-Bromo-4-Chloro-3-Indolyl Phosphate (BCIP)-NitroBlue Tetrazolium (NBT), *Sigma-Fast*). La phosphatase alcaline (apportée par la Strep-Tactin® ou l'anticorps secondaire selon l'étiquette) convertit un colorant jaunâtre soluble en un produit violet insoluble révélant ainsi la protéine d'intérêt. La membrane est ensuite rincée avec de l'eau distillée pour arrêter la coloration.

3. Pontage covalent

Une interaction protéine-protéine peut être figée de façon permanente par pontage covalent, puis observée après migration sur SDS-PAGE. Deux agents chimiques de pontage sont utilisés dans ce travail : le Bis(SulfoSuccinimidyl) Suberate (BS^3) et le glutaraldéhyde (GTA).

1. Pontage au BS^3

Les extrémités de la molécule BS^3 séparées par un bras espaceur de 11 Å, portent une fonction ester de N-HydroxySulfosuccinimide (NHS). Le groupement NHS réagit avec les amines primaires des chaînes latérales de deux résidus lysyl à proximité l'un de l'autre et forme une liaison amide stable, avec libération du groupement NHS. En pratique, 2 µg de protéine sont incubés avec 0,5 µl de BS^3 80 mM (1X final, *Thermo Fisher Scientific*), 2 µl de 1 M tampon HEPES pH 8, 2,4 µl de 5 M NaCl et de l'eau pour atteindre un volume de 20 µl. Après une incubation de 30 minutes à température ambiante, la réaction est stoppée par des amines

primaires libres par l'ajout de 2,4 µl de 250 mM Tris HCl pH 7,5, 30 mM glycine et 0,7 µl d'eau. La solution est incubée pendant 15 minutes à température ambiante. Du tampon Laemmli (1X final) est ajouté aux échantillons qui sont chauffés 5 minutes à 95°C, avant d'être séparés par SDS-PAGE.

2. Pontage au GTA

Le GTA est moins spécifique et plus court que le BS^3. En effet, il porte deux fonctions aldéhydes qui réagissent avec les fonctions amines des protéines (chaînes latérales et squelette peptidique) et forment des bases de Schiff (fonctions imines). Un volume de 1 µl de GTA 1% (w/v) (*Pierce*) est ajouté à un mélange contenant 3,8 µg de protéine et 4 µl de tampon de stockage de la protéine. Une incubation de 20 minutes à température ambiante est réalisée, puis 1 µl de 200 mM glycine est ajouté pour arrêter la réaction. Après l'addition du tampon Laemmli (1X final) et chauffage 5 minutes à 95°C, les produits de réaction sont séparés sur SDS-PAGE.

2. Techniques biophysiques

1. Spectrométrie de masse

L'identification des protéines par spectrométrie de masse est réalisée à partir d'une solution protéique ou d'un SDS-PAGE. L'analyse est faite par la plateforme de spectrométrie de masse au Centre Commun de Micro-analyse des Protéines CCMP de l'UMS 3444 à l'IBCP, avec un spectromètre de masse *Matrix Assisted Laser Desorbtion Ionisation-Time Of Flight* (MALDI-TOF) VOYAGER De Pro (*Applied Biosystems*). Cette technique repose sur l'ionisation des protéines de l'échantillon dont la digestion enzymatique à la trypsine conduit à l'obtention de peptides chargés. Ces derniers sont séparés selon leur rapport masse/charge (m/z) par un analyseur. Les masses moléculaires de ces peptides sont ensuite comparées à celles de peptides connus et répertoriés dans des banques de données, ce qui autorise l'identification des protéines initialement présentes.

2. Diffusion dynamique de la lumière

La diffusion dynamique de la lumière (*Dynamic Light Scattering* ou DLS) permet d'estimer le rayon hydrodynamique R_h de particules en solution ainsi que leur polydispersité. Ces valeurs sont obtenues à partir d'un dispositif expérimental qui illumine une solution de particules avec une lumière laser et mesure la variation d'intensité diffusée dans le temps à un angle donné.

Dans des conditions standard, l'intensité varie aléatoirement autour d'une moyenne. Cette fluctuation découle d'un phénomène d'interférences directement relié au mouvement brownien des particules en solution (Brown *et al.*, 1866) : il est rapide pour les petites particules et lent pour les grosses particules. Le calcul de la fonction d'auto-corrélation de l'intensité diffusée permet d'obtenir le coefficient de diffusion transversale D_T des particules en solution. À partir de ce coefficient, il est possible de calculer le rayon hydrodynamique de ces particules par l'équation de Stokes-Einstein :

$$R_h = K_B.T / (6\pi.\eta.D_T)$$

avec R_h, le rayon hydrodynamique ; K_B, la constante de Boltzmann ; T, la température ; η, la viscosité et D_T, le coefficient de diffusion transversale.

Le calcul de la fonction d'auto-corrélation temporelle permet également d'estimer le coefficient de polydispersité PdI de la solution. Une solution protéique faiblement polydisperse présente un PdI inférieur à 20%. En pratique, l'échantillon liquide (100 µl à concentration variable) est contenu dans une cuve stérile en plastique (*Brand*), placée dans la cellule à 25°C d'un appareil de mesure Zetasizer Nano S ZEN1600 (*Malvern Instruments*).

Une courbe de calibration est établie pour des protéines globulaires standard de masses moléculaires connues. Elle représente la valeur du R_h apparent en fonction de la masse moléculaire. Il est ainsi possible de déduire la masse moléculaire et l'état oligomérique des protéines en solution. Cependant, les valeurs de R_h et de masse moléculaire obtenues par DLS pour des protéines non sphériques sont à utiliser avec précaution.

3. SEC-MALS

La chromatographie d'exclusion de taille (SEC, *cf.* § II.3.3) peut être couplée à un détecteur de diffusion de la lumière multi-angulaire MALS (*Multi-Angle Light Scattering*). Ce système MALS, couplé à un réfractomètre OptiLab T-rEX system (*Wyatt Technology*), effectue des mesures de diffusion simultanées suivant 16 angles différents (2θ) compris entre 20° et 150°. Environ 1 mg de protéine pure a été injecté sur une colonne Superdex 200HR10/30 couplée à ce système. Une double extrapolation à concentration et angle nuls permet de déterminer la masse moléculaire d'une macromolécule en solution, selon l'équation de Zimm (Zimm, 1948) :

$$\frac{Kc}{R_{2\theta}} = \left(\frac{1}{M} + 2A_2c + O(c^2)\right)\left(1 + \frac{(4\pi n \sin\theta)^2 R_g{}^2}{\lambda^2 3} + O(\theta^4)\right)$$

avec Rθ, le rapport de Rayleigh mesuré pour l'angle de diffusion 2θ ; K, une constante optique ; c, la concentration protéique ; M, la masse moléculaire ; A_2, le deuxième coefficient du viriel ; R_g, le rayon de giration de la particule ; n, l'indice de réfraction du solvant et l, la longueur d'onde du laser.

Cette technique permet d'estimer la taille de particules en solution et d'étudier leur état oligomérique.

4. Dichroïsme circulaire

Le dichroïsme circulaire (*Circular Dichroism* ou CD) est une méthode spectroscopique utilisée pour analyser la composition en structures secondaires (hélice α, brins β, structure apériodique ou *random coil*) de protéines en solution. La plateforme de biophysique de l'IBCP dispose d'un dichrographe Chirascan (*Applied Photophysics Ltd*). Il mesure la différence d'absorbance d'une solution entre une onde lumineuse polarisée circulairement à droite (dextrogyre) et une onde lumineuse de même amplitude polarisée circulairement à gauche (lévogyre), à une longueur d'onde donnée. Les chromophores absorbent de façon inégale ces deux types d'onde lumineuse. Ce phénomène engendre la déviation du plan de polarisation mesurée par CD.

Le spectre CD représente l'ellipticité molaire par résidu [θ] en fonction d'une gamme de longueurs d'onde (180-260 nm). À ces longueurs d'onde, le chromophore principal est le groupement amide porté par le squelette peptidique. Dans l'absolu, ce groupement plan est optiquement inactif, mais la présence de liaisons peptidiques environnantes le rende optiquement actif. Le spectre CD dépend donc de la géométrie relative des liaisons peptidiques, c'est-à-dire des structures secondaires. Ainsi, chaque structure secondaire est reliée à un spectre CD caractéristique (Figure 38).

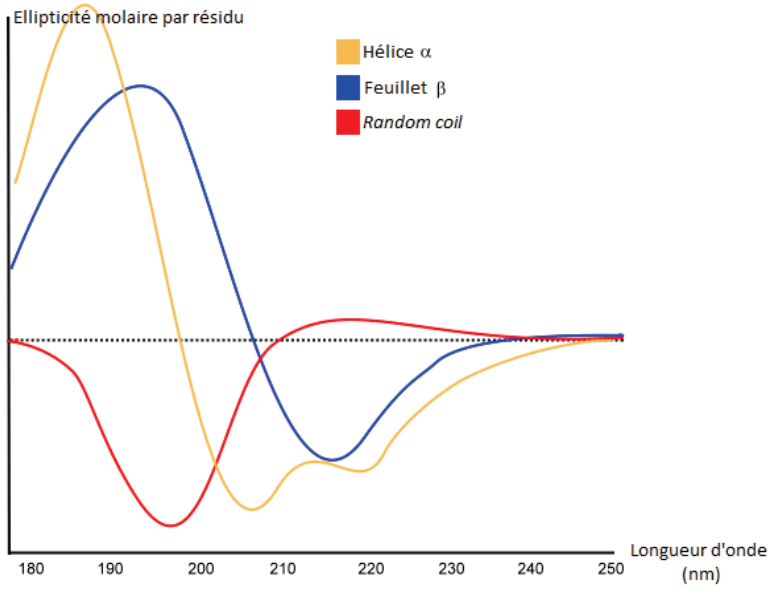


Figure 38. Spectres CD de référence caractéristiques des différentes structures secondaires des protéines (d'après http://www.proteinchemist.com/cd/cdspec.html).

Un volume de 150 µl de solution protéique pure est placé dans une cuve en quartz de 0,1 cm de trajet optique. Les mesures sont réalisées à 22°C, à des longueurs d'onde comprises entre 180 et 260 nm, avec un incrément de 0,2 nm et un temps d'intégration de 1 seconde. Le tampon protéique ne devant pas interférer avec les mesures, une soustraction du signal généré par le tampon seul est réalisée sur les signaux obtenus à partir de l'échantillon protéique.

Les spectres CD ainsi enregistrés sont traités par le logiciel Chirascan qui corrige la ligne de base, lisse le spectre et convertit l'ellipticité totale de l'échantillon ψ en ellipticité molaire par résidu $[\theta]$. Cette conversion est rendue possible par la relation suivante :

$$[\theta] = 100\psi / l.c.n$$

avec $[\theta]$, l'ellipticité molaire par résidu ; ψ, l'ellipticité totale de l'échantillon ; l, le trajet optique de la cuve ; c, la concentration protéique et n, le nombre de résidus d'acides aminés.

Des méthodes de déconvolution spectrale, utilisant différents algorithmes mathématiques, permettent d'estimer le pourcentage de chacune des structures secondaires au sein de la protéine. Cette estimation a été faite avec le serveur DichroWeb et l'algorithme mathématique SELCON3 avec le set 7 (190-240 nm) (Whitmore *et al.*, 2004).

IV. Etude structurale des protéines d'intérêt

1. Cristallographie par diffraction aux rayons X

1. Cristallogenèse

La cristallogenèse permet le passage d'une molécule d'un état soluble à un état solide ordonné (cristal). Un cristal est un assemblage tridimensionnel et tripériodique d'un motif composé d'un ensemble de molécules se répétant à l'identique d'une maille à l'autre. Le principe de cristallisation repose sur le diagramme de solubilité en fonction de la concentration en protéine et en agent précipitant (Figure 39). Ce diagramme regroupe les étapes nécessaires à la formation d'un cristal de protéine. L'augmentation progressive de la concentration d'une solution protéique conduit à une solution saturée en protéine (sursaturation). Une fois la zone de sursaturation atteinte, des interactions protéine-protéine peuvent être établies. Elles entraînent la précipitation de la protéine ou la formation de germes cristallins (nucléation) qui, eux seuls, conduisent à l'apparition de cristaux. L'étape de croissance cristalline entraîne alors la diminution de la concentration protéique en solution jusqu'à atteindre un équilibre, correspondant à la zone métastable. Lors de cette étape, les germes cristallins augmentent en taille.

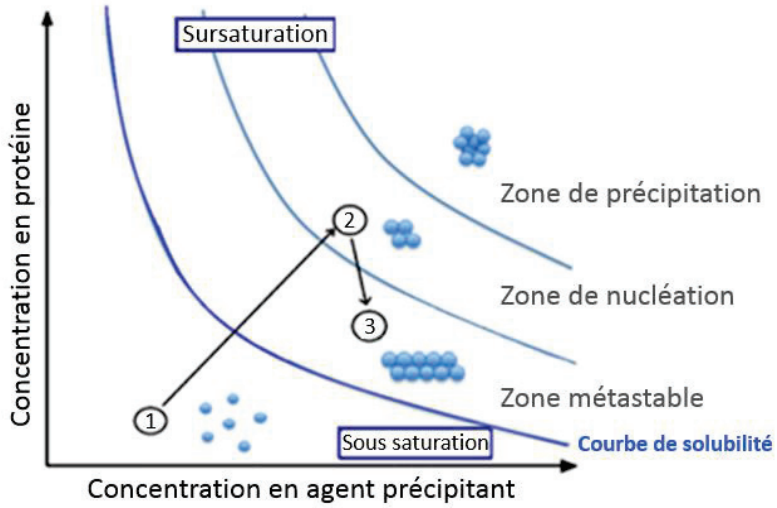


Figure 39. Diagramme de solubilité en fonction de la concentration en protéine et en agent précipitant.

1. Méthode de diffusion de vapeur

La diffusion de vapeur est la méthode la plus couramment utilisée pour la cristallogenèse. Elle permet le passage d'un état de sous-saturation à un état de sursaturation de la solution protéique d'intérêt. Cette transition a lieu grâce à un milieu hermétiquement clos formé d'un

réservoir contenant la solution de cristallisation et d'une goutte composée par un mélange agent précipitant/protéine, avec un rapport volumique différent selon nos essais de cristallisation. La différence osmotique entre l'agent précipitant dans la goutte et dans le réservoir, provoque un transfert de vapeur d'eau de la goutte (moins concentrée en agent précipitant) vers le réservoir (plus concentré en agent précipitant). Cette diffusion s'arrête lorsqu'un état d'équilibre est atteint entre les deux parties. Au cours de ce processus lent, le volume de la goutte diminue et, parallèlement, les concentrations en protéine et en agent cristallisant augmentent progressivement. Ce phénomène conduit la protéine à un état de sursaturation, pouvant aboutir à la formation de cristaux.

2. Technique de la goutte assise

Elle utilise des plaques de 96 puits (*Greiner Bio-One*) composées d'un réservoir contenant 70 µl d'agent précipitant et d'une plateforme sur laquelle est déposée une goutte de 0,3 ou 0,4 µl (Figure 40). Cette goutte est formée par le rapport volumique agent précipitant/protéine 1/1 (0,2 µl de chaque solution), 1/2 ou 2/1 (0,2 µl de solution en excès et 0,1 µl de l'autre solution). Le dépôt des nanogouttes est automatisé à l'aide d'un robot de cristallisation *Mosquito* (TTP Labtech, *Molecular Dimensions Limited*). Les plaques sont ensuite placées dans un robot de visualisation automatique RockImager 182 (*Formulatrix*), thermostaté à 19°C. Ces deux robots sont accessibles au sein de la plateforme *Protein Science Facility* de l'IBCP.

3. Technique de la goutte suspendue

Cette technique utilise généralement des plaques de 24 puits (*Greiner Bio-One*) avec un grand réservoir (500 µl) et une goutte déposée manuellement, formée par le mélange de solution protéique et d'agent précipitant (Figure 40). Pour un rapport volumique agent précipitant/protéine 1/1, 1,5 µl de chaque solution est utilisé, alors que 2 µl d'une solution et 1 µl de l'autre solution sont prélevés pour un rapport volumique 1/2 ou 2/1.

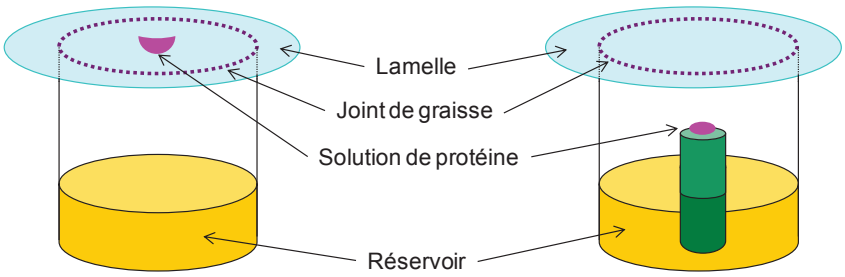


Figure 40. Croissance cristalline par la méthode de diffusion de vapeur, selon les techniques de la goutte suspendue (gauche) et de la goutte assise (droite).

Les conditions permettant la cristallisation de la protéine d'intérêt sont inconnues a priori (concentration et nature de l'agent précipitant, concentration de la protéine, pH, présence d'additifs, etc.), c'est la raison pour laquelle un maximum de paramètres doit être testé. Différents kits commerciaux de criblage des conditions de cristallisation sont disponibles au laboratoire. Ces kits permettent de faire varier le pH (4 à 9,5), la concentration et la nature de l'agent précipitant comme les sels (NaCl, sulfate d'ammonium, etc.) ou le PolyEthylène Glycol (PEG).

Dans le cas où des cristaux de taille suffisante apparaissent dans une goutte, ils sont pêchés avec une boucle en nylon et cryo-protégés par trempage dans une solution d'éthylène glycol l0 % (v/v) avant congélation en azote liquide. La cryo-protection tend à éviter la formation de glace autour et au sein du cristal, empêchant ainsi sa dégradation lors de sa congélation dans l'azote liquide.

Le cristal pêché et congelé peut être irradié par un faisceau de rayons X à la source de rayonnement synchrotron de l'*European Synchrotron Radiation Facility* (ESRF de Grenoble), pour tenter d'enregistrer des jeux de données à haute résolution et de résoudre la structure tridimensionnelle de la protéine d'intérêt.

2. Diffraction aux rayons X

Les cristaux obtenus sont illuminés avec un faisceau de rayons X, généralement monochromatique. Le faisceau incident interagit de manière élastique avec les électrons de chaque atome du cristal, qui réémettent une onde sphérique. Ces ondes réémises provoquent la formation d'interférences constructives et destructives, à l'échelle de l'objet périodique qu'est le cristal.

Ce phénomène, dit de diffraction, correspond à la réémission des rayons X dans des directions privilégiées de l'espace repérées suivant des indices *hkl* (Janin *et al.*, 1997). Le facteur de structure $F(hkl)$ est un nombre complexe représentant l'amplitude et la phase φ_{hkl} de l'onde diffusée :

$$F(hkl) = |F(hkl)|.e^{i\varphi_{hkl}}$$

Ce facteur est égal à la somme des facteurs de diffusion atomique des atomes x_j, y_j, z_j, contenus dans la maille du cristal, en considérant les décalages de phase des ondes diffusées par les différents atomes :

$$F(hkl) = \sum_{j=1}^{j=n} f_j . e^{2\pi i(hx_j + ky_j + lz_j)}$$

avec f_j, le facteur de diffusion de l'atome j et n le nombre d'atomes dans la maille.

La connaissance de *F*(*hkl*) permet de remonter à la fonction de densité électronique ρ(x,y,z), qui correspond à la répartition des électrons dans le cristal par unité de volume. En effet, la densité électronique est égale à la transformée de Fourier inverse du facteur de structure *F*(*hkl*) en un point quelconque de coordonnées (x,y,z) du cristal :

$$\rho(x,y,z) = \frac{1}{V} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F(hkl) \cdot e^{-2\pi i(hx+ky+lz)}$$

avec V, le volume du cristal de protéine.

Expérimentalement, seules les intensités I(*hkl*) des ondes diffractées sont mesurées par le détecteur. Elles correspondent au carré des amplitudes mesurées selon l'équation :

$$I(hkl) = |F(hkl)|^2$$

L'information de phase φ_{hkl}, inaccessible à l'expérience, est perdue. Des méthodes de phasage doivent alors être utilisées, comme celle du remplacement moléculaire qui permet d'estimer les phases, à partir d'une structure connue au repliement 3D proche de celui de la macromolécule étudiée.

2. Diffusion des rayons X aux petits angles

Les mesures de SAXS permettent une analyse structurale de macromolécules en solution. Elle utilise un faisceau incident monochromatique de rayons X qui illumine une solution protéique.

L'absence de phénomène significatif d'interférences constructives, contrairement à celui observé pour un assemblage cristallin, fait que l'intensité totale diffusée est beaucoup plus faible qu'en cristallographie aux rayons X. Les intensités ne sont alors mesurables que jusqu'à un faible angle de diffusion 2θ. Ainsi, les données sont obtenues à basse et moyenne résolution et donnent le rayon de giration R_g et le modèle *ab initio* de l'enveloppe moléculaire (enveloppe SAXS) de la protéine analysée.

Cette technique nécessite une solution protéique monodisperse (préalablement analysée par DLS) et permet de s'affranchir de l'étape de cristallogenèse. En pratique, chaque mesure de courbe de diffusion effectuée sur un échantillon est soustraite à la mesure de la solution tampon correspondante. Pour une solution monodisperse, le R_g et la valeur de l'intensité diffusée à angle nul I(0) peuvent être déterminés dans le domaine de Guinier de la courbe (région où $qR_g<1,3$ pour une particule globulaire avec $q = 4\pi\sin\theta / \lambda$, Putnam *et al.*, 2007) (Figure 41), selon l'équation suivante :

$$\ln I(q) = \ln I(0) - \left(\frac{q^2 R_g^{\ 2}}{3} \right)$$

La masse moléculaire de la macromolécule en solution est reliée à la valeur de I(0) selon l'équation :

$$I(0) = V_{sol} \times \frac{cM_w}{N} \times \left[\left(\frac{m_p N}{M_w} \right) - \rho_s V_p \right]^2$$

avec V_{sol} et c, le volume et la concentration de la solution protéique ; M_w, la masse moléculaire de la particule ; m_p, le nombre d'électrons dans la macromolécule ; N, le nombre d'Avogadro ($6,022.10^{23}$ mol^{-1}) ; Vp, le volume spécifique partiel de la particule et ρ_s, la densité de longueur de diffusion du solvant.

Les intensités mesurées I(q) permettent de calculer la fonction de distribution des distances ou P(r) par une transformée de Fourier inverse avec le logiciel GNOM (Figure 41). Cette fonction décrit l'ensemble des distances interatomiques de la macromolécule en solution et la distance interatomique maximale D_{max}. Ces données expérimentales 1D seront utilisées comme contraintes par le logiciel DAMMIN, pour le calcul *ab initio* de l'enveloppe SAXS (Figure 41).

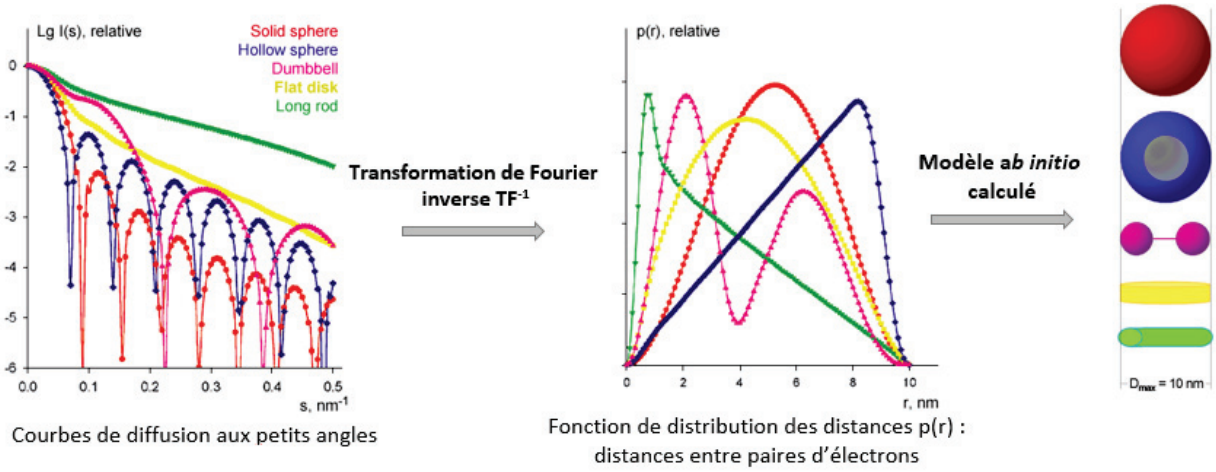


Figure 41. Principe général du traitement de données obtenues par SAXS (d'après Svergun *et al.*, 2003).

Les analyses SAXS ont été réalisées avec la source synchrotron à l'ESRF de Grenoble, sur la ligne BM29. Après purification, concentration et analyse de la monodispersité par DLS, plusieurs solutions protéiques ont été analysées par SAXS, dans une cellule sous vide thermostatée à 20°C. Différents paramètres expérimentaux (concentration protéique, pH, composition du tampon) ont été modifiés afin, notamment, d'explorer l'état oligomérique des protéines étudiées.

3. Modélisation par homologies de séquence

La modélisation 3D par homologie est une prédiction *in silico* de la structure tridimensionnelle d'une protéine. Elle peut être réalisée en fonction des homologies partagées entre la séquence d'une protéine de structure connue (*template*) et la séquence de la protéine à modéliser. Pour modéliser la structure 3D de l'IN de PERV-A/C à partir de sa séquence, nous avons utilisé le serveur SWISS-MODEL (Biasini *et al.*, 2014). Ce système automatisé compare la séquence en acides aminés de l'enzyme avec les séquences protéiques de structures connues issues de la PDB (structures résolues expérimentalement). L'outil d'alignement BLAST (*Basic Local Alignment Search Tool*, Altschul *et al.*, 1990), développé par le NCBI (*National Center for Biotechnology Information*), permet de sélectionner les séquences homologues à la protéine étudiée. L'alignement de ces séquences est réalisé par l'outil PSI-BLAST également développé par le NCBI (Altschul *et al.*, 1997). Cet alignement permet de sélectionner la séquence *template*, qui servira de patron à la construction du modèle. Il est ensuite utilisé pour générer un modèle de la protéine d'intérêt avec le logiciel MODELLER (Sali *et al.*, 1995). Le modèle établi est visualisé avec le logiciel de visualisation de structures 3D PyMOL (Delano, 2008).

V. Mesures d'interactions

1. *Histone array*

Le criblage d'interactions entre des queues d'histones et une protéine d'intérêt est réalisé avec une puce à histones ou *histone array*. Il s'agit d'une lame de verre sur laquelle des queues amino-terminales d'histones H2A, H2B, H3 et H4 sont immobilisées (MODified™ Histone Peptide Array, *Active Motif*). Ces peptides (19mers) peuvent porter différentes modifications post-traductionnelles telles que l'acétylation, la méthylation et la phosphorylation. Chaque *histone array* contient 384 combinaisons de différentes modifications d'histones dupliquées.

En pratique, les étapes du protocole sont similaires à celles décrites pour le *western-blot* (*cf.* § III.1.2, Figure 42). La puce est plongée dans 3 ml de tampon de blocage commercial (10 mM Tris-HCl pH 7,4, Tween20 0,05% (v/v), 150 mM NaCl, lait écrémé 5% (v/v), *Active Motif*) pendant 3 heures, à température ambiante. Après saturation de la puce, trois lavages de 5 minutes avec du tampon *Wash* (10 mM Tris-HCl pH 7,4, Tween20 0,05% (v/v), 150 mM NaCl, *Active Motif*) permettent d'éliminer l'excès de protéines du lait. La puce est ensuite incubée dans 3 ml de solution de *binding* (20 mM HEPES pH7,5, 20 mM DiThioThréitol (DTT), DiMéthySulfOxyde (DMSO) 15% (v/v), 8% PEG 6 000 (v/v), 10 mM $MgCl_2$) avec la protéine d'intérêt (45 nM) pendant deux heures, dans une armoire thermostatée à 30°C, sous agitation douce.

Trois lavages de 5 minutes en tampon *Wash* sont réalisés pour éliminer les protéines non fixées. La lame de verre est ensuite incubée avec 3 ml de solution de blocage, contenant un anticorps de souris dirigé contre la protéine d'intérêt (1/1 000), pendant une nuit à 4°C, sous agitation douce. Trois lavages de 5 minutes en tampon *Wash* sont réalisés avant d'incuber la lame avec 3 ml de solution de blocage contenant un anticorps secondaire de lapin (1/3 000) dirigé contre l'anticorps primaire, couplé à la peroxydase de raifort. Après une heure d'incubation à température ambiante, la puce est lavée trois fois 5 minutes dans du tampon *Wash*, avant d'ajouter 1 ml de chaque solution de révélation A et B (ECL *Western Blotting Substrate*, *Promega*).

Les produits de la réaction d'électrochimiluminescence sont détectés avec le visualiseur ImageQuant™ LAS 4000 (*GE Healthcare*). Ils correspondent à la protéine d'intérêt en interaction avec les queues d'histones. Ces dernières sont identifiées par comparaison de l'*histone array* avec le plan de dépôts fourni par le fabricant.

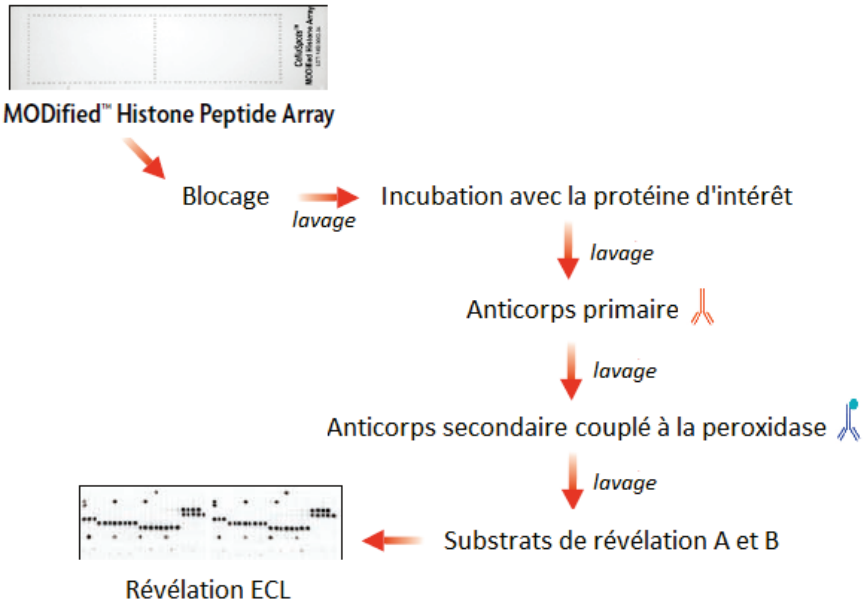


Figure 42. Principe du criblage d'interactions entre une protéine et des queues d'histones par *histone array*.

2. Criblage d'inhibiteurs par intégration concertée *in vitro*

Cette expérience d'intégration permet de suivre l'insertion des deux extrémités virales (ADN donneur) dans un ADN receveur, catalysée par un tétramère d'IN (Faure *et al.*, 2005). Elle sert également à tester des composés potentiellement inhibiteurs sur l'activité d'intégration. Cette technique a été mise au point sur plusieurs modèles rétroviraux telles les IN de VIH-1 (Faure *et al.*, 2005, Lesbats *et al.*, 2011, Lesbats *et al.*, 2008), de PFV, de l'*Avian Sarcoma Virus* et de MLV (Benleulmi *et al.*, 2015).

L'ADN donneur de 246 pb mimant l'ADN viral a été gracieusement fourni par l'équipe du Dr Parissi (« Microbiologie Fondamentale et Pathogénicité », MFP Bordeaux). Il a été obtenu par clivage du plasmide pUC19-*SupF*. Il contient l'ARNt du *SupF* (200 pb), flanqué par deux extrémités pré-clivées (CA-3'-OH) de 23 pb, qui miment les régions U5 (en 3') et U3 (en 5') des LTRs, marquées au phosphate radioactif γP^{32} sur le premier nucléotide en 5' (Faure *et al.*, 2005). Cet ADN substrat porteur des séquences LTR a été généré par PCR en utilisant les amorces RAV-1_U3_SupF_5' et RAV-1_U5_SupF_3' (Tableau 4).

Nom des amorces	Séquence des amorces
RAV-1_U3_SupF_5'	5′ AATGTAGTCTTATGCAATACTCTAATGCAATATTAACGTTGCCCGGATC 3′
RAV-1_U5_SupF_3'	5′ AATGAAGCCTTCTGCTTCATGCATGTGCTCGTGCGGCGCGTCATTTG 3′

Tableau 4. Séquences des amorces utilisées pour fabriquer l'ADN substrat dans la réaction d'intégration concertée *in vitro*.

Le plasmide p481 qui correspond au plasmide pBSK$^+$ (3 kb) porteur du gène de résistance à la zéocine (PBSK-zeo, *Stratagene*), a été utilisé comme ADN receveur (Moreau *et al.*, 2004).

En pratique, l'IN RAV-1 est diluée dans un tampon de dilution d'IN (1 M NaCl, 20 mM HEPES pH 7, 10 mM DTT) puis pré-incubée avec 10 ng d'ADN donneur et 50 ng d'ADN plasmidique receveur, dans un volume total de 5 µl, à 4°C pendant 30 minutes. La solution est ensuite incubée avec un mélange contenant 20 mM HEPES pH 7, 5 mM DTT, 10 mM $MgCl_2$, 50 mM $ZnCl_2$, 10% DMSO (v/v), 25% PEG 6 000 (v/v), dans un volume total de 10 µl, pendant 3 heures à 4°C. La réaction est arrêtée par l'ajout d'un mélange 0,8% SDS, 2 mM EDTA, dans un volume total de 12 µl. La protéinase K est également ajoutée au mélange à une concentration finale de 4 mM (*Promega*). Après une heure d'incubation à 55°C, l'ADN est isolé par une extraction phénol/chloroforme/alcool isoamylique 25/24/1 ((v/v/v), *Sigma*). Un volume de 10 µl de surnageant est ajouté à 2 µl de tampon de charge (95% bleu formamide (v/v), 2 M urée, 0,05% bleu de bromophénol (v/v), *Sigma*). Les produits de réaction sont analysés par électrophorèse sur gel d'agarose 1% dans un tampon Tris-Borate-EDTA (TBE) pH 7,6 révélé par autoradiographie.

Résultats et discussion

I. CCD isolés de RAV-1 et VIH-1

Ces CCD ont été surproduits en système bactérien et purifiés afin de tester l'activité inhibitrice des composés chimiques décrits dans le paragraphe (*cf.* § III.5.1).

1. CCD sauvage et H103C de l'IN de RAV-1

1. Surproduction et purification

En début de thèse, j'ai repris les protocoles de surproduction et de purification qui avaient été mis au point dans l'équipe (Ballandras *et al.*, 2011) et que j'avais utilisé avec succès lors de mon stage de Master.

La souche bactérienne BL21(DE3)pLysS transformée par le plasmide pETG-10a a servi à la production des CCD. Pour les deux CCD (sauvage et muté H103C), la culture a été réalisée dans un milieu LB à 37°C, en présence d'ampicilline et de chloramphénicol (50 µg/ml). Lors de la phase exponentielle de croissance bactérienne, une concentration finale de 1 mM d'IPTG a servi à l'induction de l'expression du gène d'intérêt. La culture a été incubée pendant 4 heures à 37°C pour le CCD sauvage, tandis qu'une incubation d'une nuit à 25°C a été réalisée pour surproduire le CCD H103C.

La lyse bactérienne est réalisée par sonication dans un tampon de lyse (20 mM Tris HCl pH 8, 500 mM NaCl, 25 mM imidazole, 5 mM β-mercaptoéthanol). Dans le cadre d'une purification du mutant CCD H103C, le tampon de lyse diffère par l'absence de l'agent réducteur β-mercaptoéthanol, ceci pour éviter la rupture des ponts disulfures précisément formés par la mutation. Les CCD surproduits porteurs d'une étiquette poly-histidine ont ensuite été purifiés par chromatographie d'affinité pour le nickel.

Hormis le fait qu'aucun agent réducteur n'a été ajouté dans les solutions tampons du mutant CCD H103C, un même protocole de purification a été appliqué pour purifier les CCD d'intérêt. Le surnageant obtenu après la lyse bactérienne a été injecté sur la colonne préalablement équilibrée avec le tampon de lyse. Lors de cette injection, la fraction non retenue est récupérée (Figure 43A, pic A). La colonne est lavée une première fois avec le tampon de lyse, puis avec le tampon de lavage à plus forte force ionique (20 mM Tris HCl pH 8, 1 M NaCl, 25 mM imidazole, 5 mM β-mercaptoéthanol). Ainsi, les protéines et les acides nucléiques, fixés de façon aspécifique à la résine, sont décrochés (Figure 43A, pic B). Les CCD d'intérêt ont été

élués par un gradient croissant d'imidazole formé par le mélange des tampons de lyse et d'élution (20 mM Tris HCl pH 8, 500 mM NaCl, 1 M imidazole, 5 mM β-mercaptoéthanol). Ils se décrochent à une valeur de gradient d'environ 25% tampon d'élution/ tampon de lyse (v/v), soit à une concentration d'imidazole proche de 280 mM (Figure 43A, pic C). Les fractions d'élution ont été récupérées et analysées par SDS-PAGE (Figure 43B).

L'analyse par SDS-PAGE met en évidence la présence du CCD H103C de l'IN de RAV-1, avec peu de contaminants dans les fractions d'élution (Figure 43B, piste C). En effet, une bande majoritaire, proche du marqueur de masse moléculaire de 17 kDa, coïncide avec la masse moléculaire du monomère de CCD de l'IN de RAV-1 calculée à partir de sa séquence (17,5 kDa). Seules les fractions contenant la protéine d'intérêt avec le plus haut degré de pureté ont été combinées. Un profil de purification similaire a été obtenu pour le CCD sauvage de l'IN de RAV-1 (données non montrées).

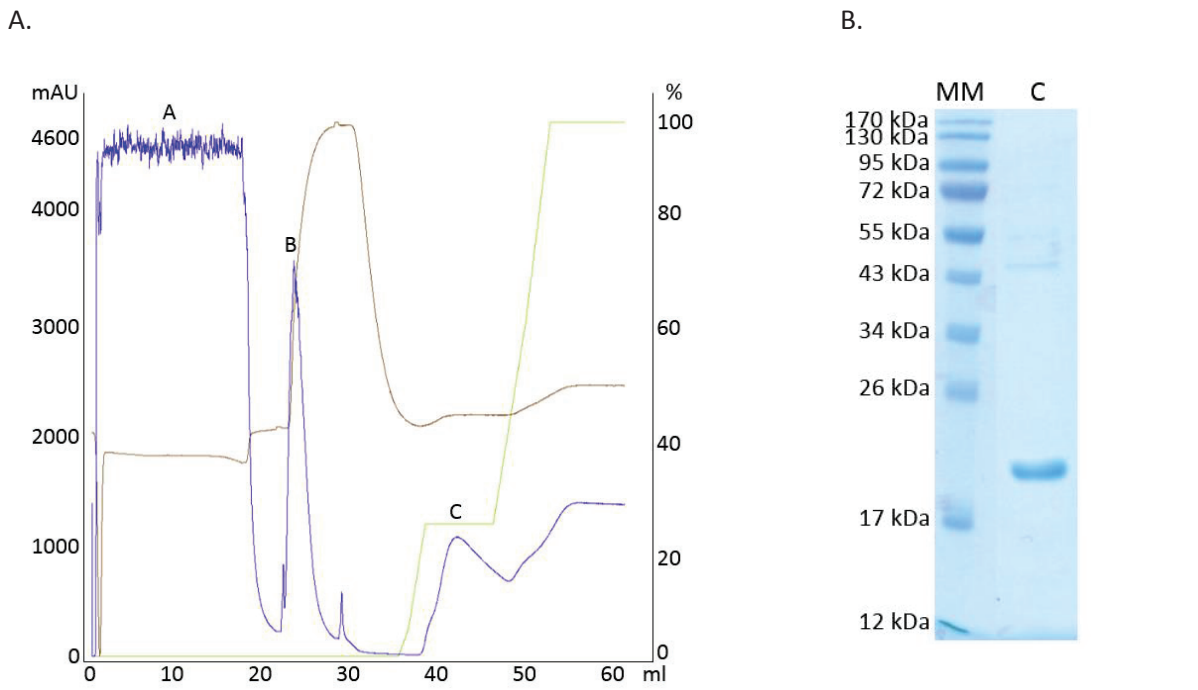


Figure 43. A. Chromatogramme de purification du CCD H103C de l'IN de RAV-1 par chromatographie d'affinité pour le nickel. Courbe bleue : suivi de l'absorbance à 280 nm (axe des ordonnés gauche) en fonction du volume d'élution (axe des abscisses). Courbe marron : suivi de la conductance du tampon. Courbe verte : suivi du gradient d'imidazole (% tampon d'élution/tampon de lyse, axe des ordonnées droit). Fraction A : fraction non retenue. Fraction B : fraction de lavage. Fraction C : fraction d'élution.
B. Analyse par SDS-PAGE 12% du CCD H103C de l'IN RAV-1 en fin de purification, avec révélation au bleu de Coomassie. Piste MM : marqueurs de masses moléculaires connues. Piste C : CCD H103C de l'IN de RAV-1 purifié (pic C).

Un clivage de l'étiquette poly-histidine par la thrombine (1 unité de thrombine pour 5 mg de protéine) a ensuite été réalisé à 4°C pendant une nuit. Une nouvelle chromatographie d'affinité pour le nickel a été réalisée afin d'éliminer les étiquettes de purification clivées.

Une analyse par SDS-PAGE des fractions protéiques récupérées a révélé un phénomène d'agrégation protéique. La protéine d'intérêt ne parvient pas à traverser les mailles du gel de polyacrylamide et s'accumule au fond des puits de dépôts. La solution protéique a été placée sur un filtre de 0,02 µm (Anapore^{TM}, *Whatman*). Après un cycle de centrifugation à 4 000 × g pendant 20 minutes à 4°C, la totalité de la protéine est retrouvée dans le retentât. Une analyse par DLS a confirmé ce phénomène d'agrégation. Pour tenter d'obtenir une protéine soluble, j'ai reproduit le protocole de surproduction et de purification en m'affranchissant de l'étape de clivage de l'étiquette de purification, mais aucune amélioration n'a été observée. Cette protéine, soluble après purification, n'est donc pas stable dans le tampon d'élution.

Des problèmes similaires d'agrégation ont été observés dans le cadre de la surproduction et purification du CCD sauvage de l'IN de RAV-1. J'ai donc décidé de modifier les protocoles de surproduction et de purification, en espérant m'affranchir de cette agrégation.

2. Essais de solubilisation

Pour tenter d'obtenir une protéine soluble, des changements ont été apportés aux protocoles de surproduction et de purification précédemment mis au point. Une liste non exhaustive de ces modifications est détaillée dans cette partie.

Avant de modifier les paramètres expérimentaux, je me suis assurée de l'absence de mutation dans la séquence du plasmide d'expression utilisé, par séquençage de l'ADN.

Afin d'assurer une pousse bactérienne optimale, une transformation bactérienne a été systématiquement réalisée avant chaque surproduction protéique. L'ajout de glycérol 10% (v/v) dans le milieu de culture, la diminution de la concentration d'IPTG (de 1 mM à 0,5 mM puis 0,3 mM) et/ou de la durée d'induction (2 heures au lieu de 3 heures) n'ont pas permis d'améliorer la solubilité protéique. La culture bactérienne a également été placée à 4°C, pendant 3 heures après avoir atteint une valeur de DO de 0,6. Ce stress permet de stimuler l'activité des chaperonnes et ainsi d'améliorer le repliement de la protéine. Des tests de surproduction par fuite d'expression ont été conduits en laissant la culture non induite (sans IPTG) pousser à 30°C et à 37°C, sous agitation (180 rpm) et sans chloramphénicol afin de lever le blocage de l'expression basale. Cette méthode aurait pu permettre une surproduction progressive de la protéine d'intérêt, une amélioration de son état de repliement et donc de son état de solubilité ; mais le phénomène d'agrégation persistait.

Un milieu d'auto-induction contenant différents composés chimiques (lactose, glucose, glycérol, Na_2HPO_4, KH_2PO_4, NH_4Cl, Na_2SO_4, $MgSO_4$) a également été testé mais il n'a pas permis de palier le problème d'agrégation (Studier, 2005).

La vitesse de centrifugation du lysat bactérien a été augmentée pour tenter de séparer les protéines solubles des protéines agrégées (de 14 000 × g à 20 000 × g). Des problèmes de qualité d'eau (présence d'un excès massif de fer et de cuivre) ont été observés dans l'IBCP pendant plusieurs mois. Nous avons donc eu recours à de l'eau commerciale (*Carl Roth*) pour réaliser les tampons de purification. Du glycérol (5% et 10% (v/v)) et/ou du CHAPS (3-[(3-cholamidopropyl)diméthylammonio]-1-propanesulfonate, 5 et 10 mM final, *Sigma*) ont été ajoutés à ces tampons pour tenter d'éviter l'agrégation protéique (Matuo *et al.*, 1985), sans succès.

Le Tris HCl a été substitué à de l'HEPES pour tenter d'améliorer la solubilité de la protéine (Graslund *et al.*, 2008). L'imidazole présent dans les tampons de purification a été remplacé par de l'histidine (2 mM dans les tampons de lyse et de lavage, 100 mM dans le tampon d'élution), mais le gradient d'histidine n'a pas permis d'obtenir une protéine soluble. Il en a été de même avec un gradient d'arginine (50 mM dans les tampons de lyse et de lavage, 100 mM dans le tampon d'élution) (Arakawa *et al.*, 2003). La colonne de purification NiNTA a été renouvelée et le stock de thrombine utilisé pour cliver l'étiquette a été changé. L'ensemble de ces modifications ne m'a pas permis de résoudre le problème d'agrégation du CCD de l'IN de RAV-1 (sauvage ou muté).

En parallèle, des essais de surproduction et de purification des mutants solubles F185K et cystéine Y99C/F185K du CCD de VIH-1 ont été entrepris afin de tester les composés chimiques potentiellement inhibiteurs.

2. CCD mutés de l'IN de VIH-1

1. Surproduction et purification

La mutation F185K est connue pour améliorer la solubilité de cette IN rétrovirale (Jenkins *et al.*, 1996), tandis que la mutation Y99C correspond à l'équivalent H103C du CCD de l'IN de RAV-1 (Figure 15, page 40) et permettra de reproduire le « nouvel » assemblage dimérique à l'échelle de l'IN de VIH-1. Les expériences menées sur ces CCD ont été réalisées avec Catherine Luengo (équipe du Dr Ronfort, INRA Lyon).

Les premiers protocoles que nous avons testés pour purifier les mutants F185K et Y99C/F185K du CCD de VIH-1, correspondent à ceux utilisés dans l'équipe du Dr Ronfort pour purifier l'IN de VIH-1 entière (adaptés de l'article (Cellier *et al.*, 2013).

En pratique, un clone (issu de la transformation de bactéries BL21(DE3)pLysS avec le plasmide pETG-10a) a été inoculé dans un milieu LB supplémenté d'ampicilline et de chloramphénicol à 50 µg/ml. Une induction à l'IPTG (1 mM final) a été réalisée lorsque la DO a atteint une valeur de 0,6. Le milieu a ensuite été incubé à 37°C pendant trois heures, à 180 rpm, avant d'être centrifugé 20 minutes à 4 000 × g.

Chaque culot bactérien a été resuspendu dans 8 ml de tampon de lyse A1 (20 mM Hepes pH 7,5, 2 M KCl, 5 mM β-mercaptoéthanol) contenant du lysozyme (1,5 mg/ml final). Après 30 minutes d'incubation à 4°C, du $MgCl_2$ (10 mM final) et de la DNase (20 µg/ml final) ont été ajoutés au mélange incubé à nouveau à 4°C, pendant 30 minutes. Après sonication, une centrifugation a été réalisée (14 000 × g, 4°C, 20 minutes). Le surnageant récupéré a été dilué de moitié dans un tampon A2 (tampon A1 sans KCl). Ainsi, la concentration finale de KCl dans cette solution est de 1 M. Le surnageant dilué a été filtré et injecté sur une colonne NiNTA pré-équilibrée en tampon B (20 mM Hepes pH 7,5, 1 M KCl, 5 mM β-mercaptoéthanol). Un gradient croissant tampon C/tampon B (20 mM Hepes pH 7,5, 1 M KCl, 1 M imidazole, 5 mM β-mercaptoéthanol) a été effectué, mais aucun pic d'élution n'a été observé. La forte concentration en KCl dans les tampons A1 et B aurait pu empêcher la fixation du CCD sur la colonne, mais l'absence de ce domaine dans les fractions non retenues a permis d'écarter cette hypothèse.

2. Essais de solubilisation

Un protocole identique utilisant des tampons moins stringents (1 M KCl dans le tampon A1, 500 mM KCl dans les tampons B et C) a été appliqué pour améliorer les conditions de purification. Un pic d'élution a été observé à une concentration d'imidazole d'environ 80 mM, lors de la purification de chaque CCD muté (Figure 44A). Les fractions d'élution ont été combinées, la solution protéique résultante a été dosée et analysée par SDS-PAGE (Figure 44B). Une bande, plus basse que le marqueur de masse moléculaire de 26 kDa, a pu être observée. Une analyse par spectrométrie de masse a permis d'identifier la protéine d'intérêt de masse moléculaire proche de 20 kDa.

Environ 2 mg de protéine pure a été obtenue à partir d'un litre de culture bactérienne. Les profils de purification obtenus étant les mêmes pour les deux CCD mutés de l'IN de VIH-1, seuls ceux correspondant au CCD F185K sont montrés.

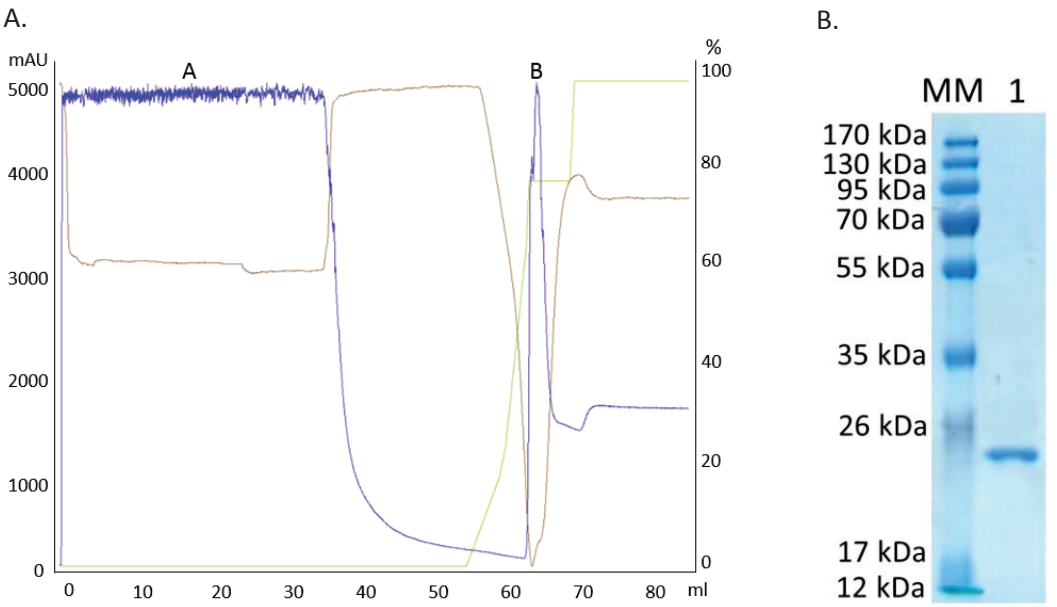


Figure 44. A. Chromatogramme de purification du CCD F185K par chromatographie d'affinité pour le nickel. Courbe bleue : suivi de l'absorbance à 280 nm en sortie de colonne (axe des ordonnés gauche) en fonction du volume d'élution (axe des abscisses). Courbe marron : suivi de la conductance du tampon. Courbe verte : suivi du gradient d'imidazole (% tampon d'élution/tampon de lyse, axe des ordonnées droit). Fraction A : fraction non retenue. Fraction B : fraction d'élution.
B. Analyse par SDS-PAGE 12% du CCD F185K de l'IN de VIH-1 après purification, avec révélation au bleu de Coomassie. Piste MM : marqueurs de masses moléculaires connues. Piste B : CCD F185K de l'IN de VIH-1 purifié (pic B).

Les CCD F185K et Y99C/F185K ont été concentrés à 1,5 mg/ml avant de conduire une analyse par DLS. Cette analyse décrit plusieurs pics (PdI supérieurs à 20%), à des diamètres apparents hauts correspondant à des masses moléculaires estimées plus élevées que celles attendues. Par exemple, dans le cas du CCD F185K, les diamètres de 16,4 nm et 93,8 nm correspondent à des particules supérieures à 440 kDa. Ces données traduisent donc la présence d'agrégats dans nos échantillons protéiques.

Une étape de dessalage a été ajoutée au protocole de purification des CCD mutés de l'IN de VIH-1. Cet échange de tampon a été réalisé à partir des fractions d'élution rassemblées, avec un tampon final D (tampon C sans imidazole). Il a été utilisé pour éliminer l'imidazole du tampon de stockage et ainsi tenter d'augmenter la stabilité des CCD purifiés. Cependant, d'après les analyses DLS menées après dessalage, le phénomène d'agrégation n'a pas pu être résolu.

Le protocole de surproduction utilisé pour l'IN entière de VIH-1 avec la mutation F185K a également été testé pour ces CCD mutants (Jenkins *et al.*, 1996), mais les quantités protéiques étaient insuffisantes pour tenter une purification.

II. Etude de l'IN de RAV-1 sauvage entière

Les essais de surproduction et de purification n'ont pas permis d'obtenir les CCD des IN de RAV-1 et de VIH-1 solubles. J'ai alors focalisé mes efforts sur la protéine IN RAV-1 entière sauvage, pour tester les molécules chimiques potentiellement inhibitrices.

1. Surproduction et purification

La souche bactérienne BL21(DE3) a été transformée par le plasmide pPROEX-Hta, qui contient le gène codant l'IN de RAV-1. Un clone bactérien a permis d'inoculer une préculture contenant du milieu LB supplémenté en ampicilline (100 µg/ml final). Après une incubation à 37°C pendant une nuit, la surproduction protéique a été réalisée dans un milieu LB avec de l'ampicilline (100 µg/ml final), du glucose 1% (v/v) et du Tris HCl (25 mM, pH 7). Une induction de l'expression a été réalisée à une DO de 0,6 par 1 mM d'IPTG, pendant 4 heures à 37°C. Après l'étape d'induction, les bactéries ont été lysées par sonication dans un tampon de lyse (50 mM BisTris pH 7, 0,5 M NaCl, 1 M urée, 25 mM imidazole). Une analyse par SDS-PAGE a permis de vérifier la surproduction et la solubilité de l'IN de RAV-1 (Figure 45).

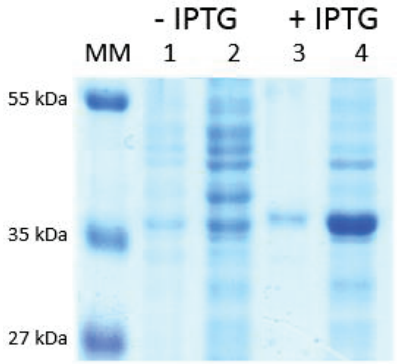


Figure 45. Analyse par SDS-PAGE 12% de la surproduction et de la solubilité de l'IN de RAV-1, après révélation au bleu de Coomassie. Piste MM : marqueurs de masses moléculaires connues. Pistes 1 : fractions insoluble (culot) et soluble (surnageant) avant induction à l'IPTG. Piste 3 et 4 : fractions insoluble et soluble avec IPTG.

Les résultats obtenus montrent que l'IN de RAV-1 est bien surproduite après l'ajout d'IPTG dans le milieu de culture (Figure 45, piste 4). Cette protéine est majoritairement présente dans la fraction soluble après la lyse bactérienne (Figure 45, pistes 2 et 4). L'absence

de fuite d'expression significative du gène d'intérêt se traduit par une faible quantité d'intégrase surproduite dans la fraction soluble, en absence d'IPTG (Figure 45, piste 2).

Une chromatographie d'affinité pour le nickel a ensuite été conduite afin de purifier l'IN de RAV-1 portant une étiquette poly-histidine (Figure 46A). Après injection du surnageant sur la colonne de purification préalablement équilibrée avec le tampon de lyse (Figure 46A, pic A), un tampon de lavage (identique au tampon de lyse avec 1 M NaCl) a permis de décrocher des contaminants fixés de façon aspécifique sur la colonne (Figure 46A, pic B). L'élution de la protéine a été réalisée grâce à un gradient d'EDTA par mélange du tampon de lyse avec le tampon d'élution (50 mM BisTris pH 7, 0,5 M NaCl, 1 M urée, 100 mM EDTA). La protéine est décrochée lorsque le gradient atteint la valeur de 25% de tampon d'élution (v/v) correspondant à une concentration de 25 mM d'EDTA (Figure 46A, pic C).

Les fractions non retenue (Figure 46B, pic A, piste 1), de lavage (pic B, piste 2) et d'élution (pic C, pistes 3 à 9) ont été analysées par SDS-PAGE. Les fractions d'élution ont été rassemblées et un échange de tampon a été réalisé par dessalage avec le tampon de stockage (50 mM BisTris pH 7, 100 mM $AmSO_4$, 100 mM $MgSO_4$). Environ 20 mg de protéine par litre de culture ont ainsi été produits, purifiés et stockés à -80°C. L'intégrité de la protéine en fin de purification a été évaluée par SDS-PAGE (Figure 46B).

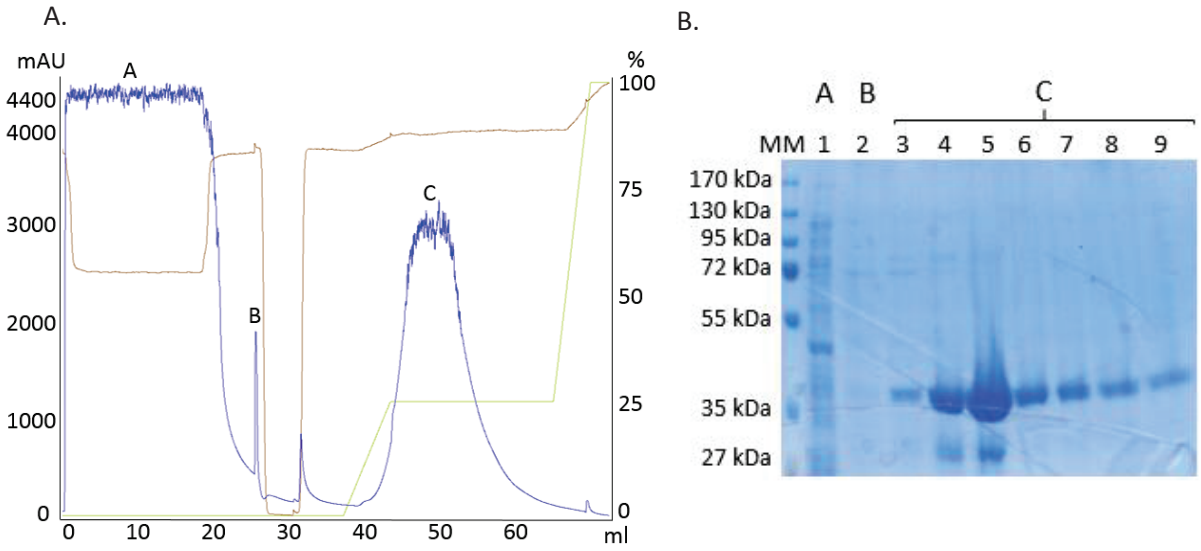


Figure 46. A. Chromatogramme de purification de l'IN de RAV-1 par chromatographie d'affinité pour le nickel. Courbe bleue : suivi de l'absorbance à 280 nm en sortie de colonne (axe des ordonnés gauche) en fonction du volume d'élution (axe des abscisses). Courbe marron : suivi de la conductance du tampon. Courbe verte : suivi du gradient d'EDTA (% tampon d'élution/tampon de lyse, axe des ordonnées droit). Fraction A : fraction non retenue. Fraction B : fraction de lavage. Fraction C : fraction d'élution.
B. Analyse par SDS-PAGE 12% de l'IN de RAV-1 après purification, avec révélation au bleu de Coomassie. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : fraction non retenue (pic A). Piste 2 : fraction de lavage (pic B). Piste 3 à 9 : fractions d'élution (pic C).

Une bande de plus bas poids moléculaire, observée sur l'analyse SDS-PAGE post-purification (Figure 46B, pistes 4 et 5), a été isolée pour une analyse par spectrométrie de masse. Il s'agit de l'IN de RAV-1, clivée par une protéase non inhibée par le cocktail d'anti-protéases utilisé lors de la lyse des bactéries (*cf.* § II.2, Matériels et méthodes). Ce clivage a lieu à l'extrémité carboxy-terminale de la protéine et l'étiquette poly-histidine est conservée.

2. Clivage de l'étiquette de purification

Un rapport protéase du TEV/intégrase de 1/20 (w/w) et une incubation durant une nuit à 4°C ont été appliqués pour tenter de cliver l'étiquette poly-histidine. Les fractions incubées ou non avec la protéase du TEV ont été analysées par SDS-PAGE (Figure 47B).

L'examen du gel indique la présence de l'IN de RAV-1 avec un haut niveau de pureté, avant action de la protéase (Figure 47B, piste 1). La bande correspondant à la protéine incubée avec la protéase du TEV (Figure 47B, piste 2) est plus basse que celle de la protéine seule (Figure 47, piste 1), ce qui signifie que le clivage de l'étiquette a bien eu lieu. Le produit de la protéolyse de l'IN de RAV-1, observé sur l'analyse SDS-PAGE ci-dessous (Figure 47B), est toujours visible après clivage de l'étiquette (Figure 47B, piste 2).

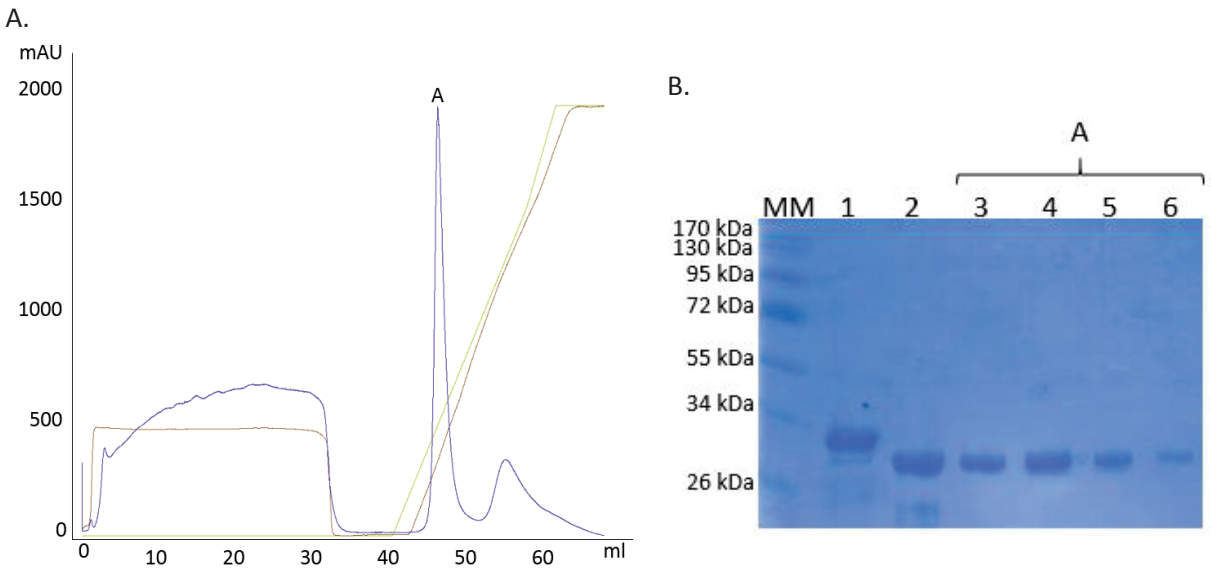


Figure 47. A. Chromatogramme de purification de l'IN de RAV-1 par chromatographie d'affinité pour le nickel, après clivage de l'étiquette de purification. Courbe bleue : suivi de l'absorbance à 280 nm en sortie de colonne (axe des ordonnés) en fonction du volume d'élution (axe des abscisses). Courbe marron : suivi de la conductance du tampon. Courbe verte : suivi du gradient d'EDTA (tampon de stockage/tampon d'élution). Fraction A : IN RAV-1 sans étiquette.

B. Analyse par SDS-PAGE 12% de l'IN de RAV-1 avant et après clivage de l'étiquette, avec révélation au bleu de Coomassie. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : fraction sans protéase du TEV (témoin non clivé). Piste 2 : fraction incubée avec la protéase du TEV (témoin clivé). Piste 3 à 6 : fractions d'élution correspondant au pic A du chromatogramme.

Pour éliminer l'étiquette de purification ainsi que la protéase du TEV en fusion avec une étiquette poly-histidine, la solution protéique a été réinjectée sur une colonne de nickel. L'ajout d'une faible concentration d'EDTA (10 mM) a servi à décrocher la protéine d'intérêt dépourvue de son étiquette (Figure 47A, pic A ; Figure 47B, pistes 3 à 6), qui a probablement une affinité pour la résine. Le produit de dégradation de la protéine d'intérêt n'étant pas élué dans le pic A, il est possible que la région manquante soit à l'origine de l'affinité de l'IN de RAV-1 entière (sans étiquette) pour la colonne NiNTA. Un dessalage a été réalisé avec le tampon de stockage pour permettre d'éliminer l'EDTA résiduel, avant de stocker la protéine clivée à -80°C.

3. Analyse par diffusion dynamique de la lumière (DLS)

Après décongélation sur glace, la solution d'IN de RAV-1 a été analysée par DLS afin de vérifier la monodispersité de l'échantillon (Figure 48). À une concentration de 2 mg/ml, un seul pic est observé sur les profils de distribution de volume. Avec une valeur de PdI égale à 20%, ce résultat indique que la solution protéique est monodisperse (pic unique) et stable dans le temps. Le diamètre apparent de la protéine en solution est de 9,3 nm ; il correspond à une masse moléculaire comprise entre 102 kDa et 150 kDa. D'après la masse moléculaire théorique de l'IN de RAV-1 sans étiquette (32 kDa), elle formerait un tétramère en solution à 2 mg/ml.

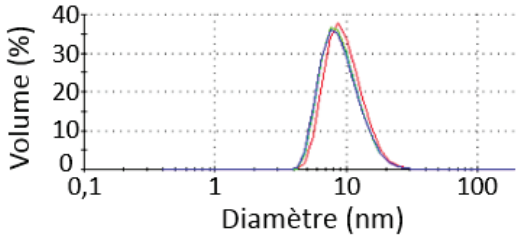


Figure 48. Profil de distribution du volume (%) en fonction du diamètre hydrodynamique (nm) de l'IN de RAV-1 concentrée à 2 mg/ml, obtenus par DLS à 25°C (superposition de trois mesures successives).

4. Essais de cristallisation

La structure de l'IN de RSV a été résolue par cristallographie aux rayons X sans le NTD, celui-ci n'étant pas observé dans les cartes de densité électronique (code PDB 4FW2, Shi *et al.*, 2013). Pour tenter de résoudre la structure de l'IN entière de RAV-1 (sans étiquette), différentes conditions de cristallisation ont été testées par la méthode de diffusion de vapeur avec la technique de la goutte assise. Les kits commerciaux PEG/Ion, PEG/Ion2, CrystalScreen (*Hampton Research*), $AmSO_4$ Suite et JCSG+ (*Qiagen*) ont permis de cribler près de 1 000 conditions. Après trois à cinq jours, des cristaux ont pu être observés dans trois conditions

expérimentales décrites ci-dessous (Tableau 5). Environ trente cristaux ont été soumis aux rayons X à l'ESRF de Grenoble et, malheureusement, il s'agissait de cristaux de sels.

La présence de sulfate d'ammonium et de sulfate de magnésium dans le tampon de la protéine a pu aboutir à la formation de ces cristaux de sels. J'ai tenté de les éliminer du tampon, mais la protéine s'agrégeait. La diminution de leurs concentrations respectives a également conduit à un phénomène d'agrégation. À ce jour, aucun cristal de la protéine IN RAV-1 entière n'a pu être obtenu.

Kit commercial	Condition de cristallisation testée	Concentration d'IN de RAV-1	Tampon de stockage	Rapport volumique agent précipitant/protéine
PEG/Ion2 (*Hampton Research*)	0,1 M $CaCl_2$ 0,1 M acétate de sodium pH 4,6 PEG 400 15% (w/v)	5 mg/ml	50 mM BisTris pH 7, 100 mM $AmSO_4$, 100 mM $MgSO_4$	1/1, 2/1
PEG/Ion2 (*Hampton Research*)	0,2 M $CaCl_2$ 0,1 M Tris pH 8,5 PEG 4 000 25% (w/v)			1/1

Tableau 5. Conditions de cristallisation testées à 19°C en présence de l'IN de RAV-1, pour lesquelles des cristaux sont apparus.

5. Criblage d'inhibiteurs anti-IN RAV-1

1. Conception des molécules

La « nouvelle » interface dimérique, observée sur le CCD de l'IN de RAV-1 (Ballandras *et al.*, 2011), présente une crevasse stabilisée par une molécule de MES issue de la solution de cristallisation (Figure 49A). Des études préliminaires de cette cavité ont mis en évidence la présence de nombreux résidus polaires, capables d'établir des liaisons hydrogène et dipôle-dipôle avec un ligand.

Une première série de molécules, analogues au MES, a été synthétisée par l'équipe du Dr Maurin (ENS Lyon) et testée par l'équipe du Dr Ronfort (INRA Lyon) avant mon arrivée au laboratoire. Les tests réalisés *in vitro* sur l'activité catalytique de l'IN de RAV-1 et *in cellulo* sur la réplication du virus RAV-1, n'ont montré aucun effet inhibiteur de ces composés.

Des premières expériences de *docking*, menées par l'équipe du Dr Maurin, ont montré que des molécules dérivées du raltégravir pouvaient se loger, avec une forte affinité, dans la

crevasse fixant le MES (Figure 49B). Le Dr Merceron a testé ces composés mais aucune interaction n'a été détectée. Des tests de trempage des molécules, réalisés avec des cristaux de CCD de l'IN de RAV-1 présentant la « nouvelle » interface, n'ont pas permis d'observer de complexe par cristallographie aux rayons X et aucun cristal n'a été obtenu lors des essais de co-cristallisation.

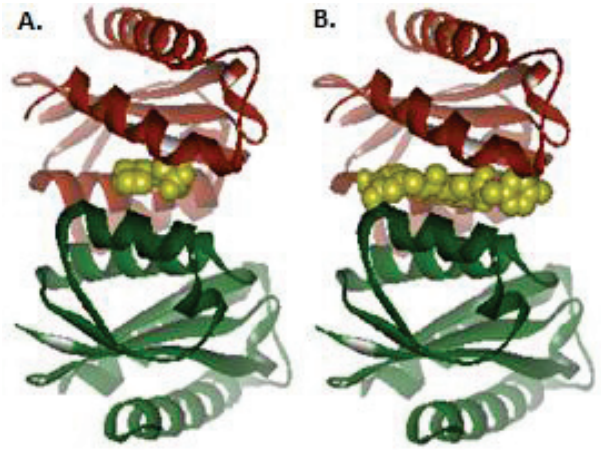


Figure 49. A. « Nouvel » assemblage dimérique du CCD de l'IN de RAV-1 en interaction avec la molécule de MES ou B. de raltégravir (en jaune) logée dans une des deux cavités de l'interface.

Une troisième série de molécules a été sélectionnée par *docking* (fixation dans la cavité comparable à celle du raltégravir, Figure 49B) et synthétisée par l'équipe du Dr Maurin, pour tenter de bloquer la réaction d'intégration. Elles portent un noyau central rigide type coumarine -OH ou -COOH (composés 1 à 5), ou indole (composés 6 à 11), fonctionnalisé avec des liaisons peptidiques (Tableau 6). Elles présentent une meilleure affinité pour la « nouvelle » interface dimérique (énergie de liaison calculée par *docking* : -7 à -11 kcal/mol) que les deux premières séries de molécules (environ -5 kcal/mol).

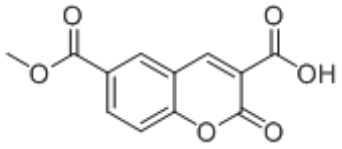
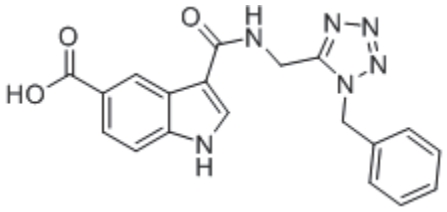
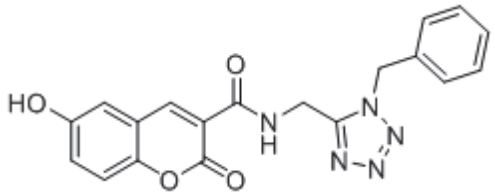
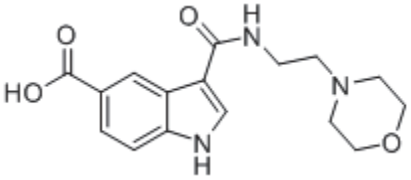
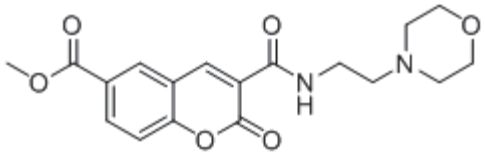
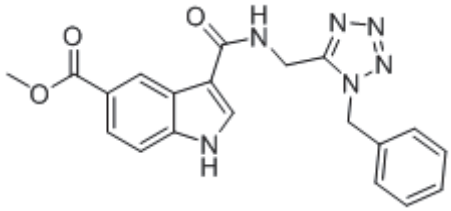
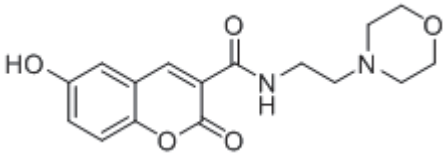
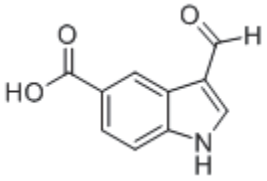
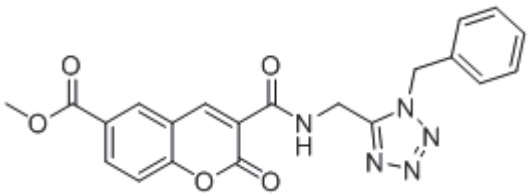
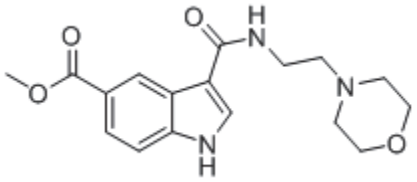
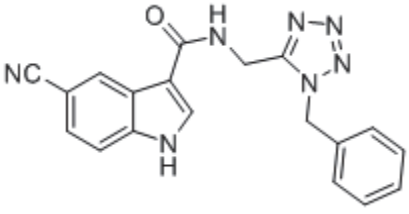
Molécules à noyau coumarine	Molécules à noyau indole
1.	6.
	
2.	7.
	
3.	8.
	
4.	9.
	
5.	10.
	
	11.
	

Tableau 6. Molécules appartenant à la troisième série de composés potentiellement inhibiteurs de l'activité d'intégration rétrovirale.

2. Solubilisation des composés

Les nouveaux composés synthétisés m'ont été fournis à l'état solide pendant mon stage doctoral. Différentes stratégies ont été envisagées pour les solubiliser. Malgré leur nature hydrophobe, j'ai d'abord testé leur solubilité dans l'eau, pour éviter l'ajout de solvants pouvant interférer dans la réaction enzymatique d'intégration. Au vu de la turbidité du milieu, même à faible concentration (de l'ordre du µM), les composés sont insolubles. Un chauffage pendant 6 heures à 90°C n'a pas permis de les solubiliser dans l'eau.

La solubilisation dans l'eau n'ayant pas fonctionné, le solvant polaire organique DMSO a été ajouté à des concentrations croissantes. Il a permis de solubiliser les composés chimiques à une concentration de 1 M, dans une solution de 100% (v/v) de DMSO. Les composés solubilisés ont ensuite été dilués dans l'eau à 1 mM pour tester leur pouvoir inhibiteur sur l'activité d'intégration de l'IN de RAV-1. La concentration finale de DMSO est négligeable (0,0001% (v/v)). Elle n'interfère pas dans la réaction d'intégration concertée *in vitro* et est suffisante pour assurer la solubilité des composés. La concentration de chaque composé dans le mélange réactionnel est de 100 µM.

3. Intégration concertée *in vitro*

1. Optimisation de la concentration d'IN de RAV-1

L'IN de RAV-1 a été incubée avec les ADN donneur et accepteur décrits dans le paragraphe V.2 du chapitre 'Matériels et méthodes'. Les réactions d'intégration concertée ont été réalisées selon le protocole de l'équipe du Dr Parissi (MFP Bordeaux, Benleulmi *et al.*, 2015), à différentes concentrations d'IN de RAV-1 (0,2 ; 0,4 et 0,9 µM final).

Sur la piste 1 (Figure 50B), la bande la plus haute correspond aux produits d'intégration concertée (*Full-Site Integration* ou FSI) et aux produits d'intégration partielle (*Half-Site Integration* ou HSI) (Figure 50A). Les formes circulaires FSI ne sont pas discernables sur gel mais elles peuvent être clonées et quantifiées dans des bactéries préalablement transformées. La bande nommée *linear FSI* correspond au produit d'intégration concertée, sous forme linéaire (Figure 50A). Les produits d'auto-intégration (donneur-donneur) sont représentés par les bandes *AutoIntegration* ou AI (Figure 50A). La bande S de plus bas poids moléculaire correspond au substrat ADN donneur (Figure 50A).

L'autoradiographie du gel d'agarose met en évidence un effet de la concentration d'IN de RAV-1 sur l'activité d'intégration. Cette activité diminue avec l'augmentation de la concentration d'intégrase. En effet, les produits d'intégration sont moins visibles à 0,9 µM (Figure 50B, piste 3) et à 0,4 µM final (Figure 50B, piste 2) d'IN de RAV-1 qu'à 0,2 µM (Figure 50B, piste 1). Les bandes correspondant à l'ADN donneur libre (S) sont similaires dans les trois conditions de concentration enzymatique testées. Bien que l'IN de RAV-1 présente une faible préférence pour les régions fortement transcrites (Mitchell *et al.*, 2004), elle pourrait intégrer l'ADN viral en des sites non spécifiques. Ainsi, l'excès d'IN pourrait induire de multiples coupures dans l'ADN receveur et empêcher les évènements d'intégration.

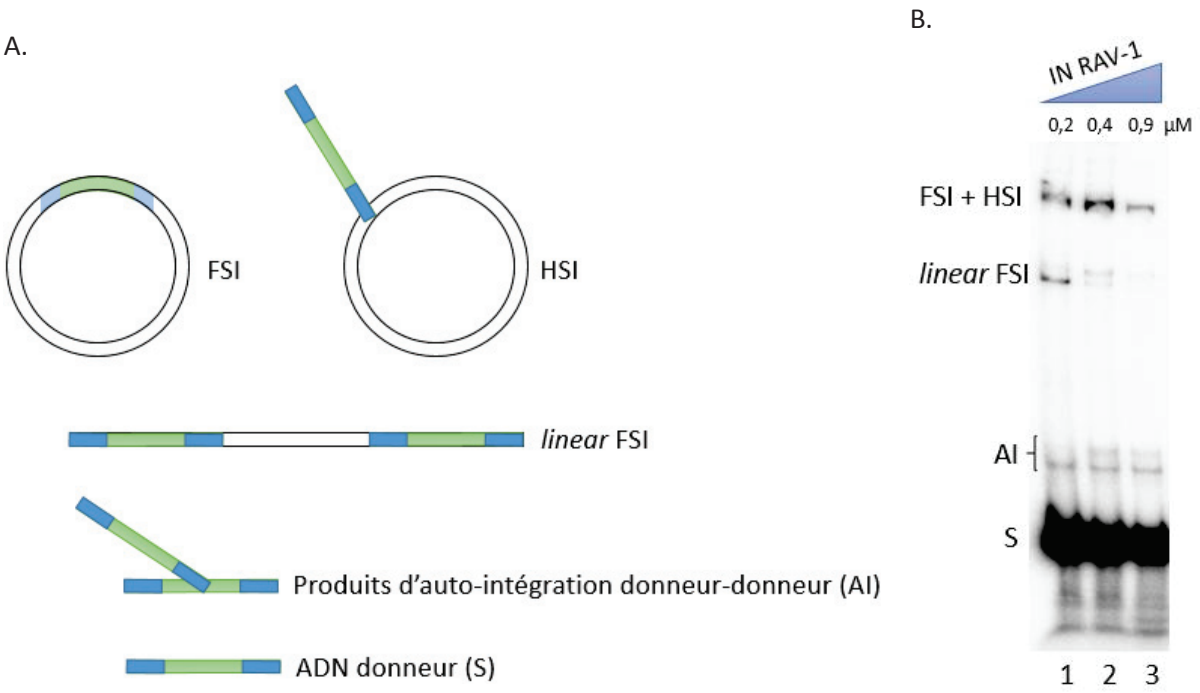


Figure 50. A. Représentation schématique de l'ADN donneur (S), des produits d'auto-intégration (AI), d'intégration partielle (HSI) et concertée (FSI circulaire et linéaire).
B. Test d'intégration concertée *in vitro* en présence de l'IN de RAV-1 recombinante (concentrations finales croissantes), incubée avec un ADN donneur et un plasmide circulaire p481 (ADN cible).

Ces tests ont permis de sélectionner la concentration d'IN de RAV-1 optimale (0,2 µM final) nécessaire à une intégration concertée efficace dans ce système. Cette concentration a ensuite été utilisée pour tester l'effet des composés potentiellement inhibiteurs sur l'activité catalytique de cette intégrase.

2. Tests d'intégration concertée en présence des composés potentiellement inhibiteurs

Un premier test a été réalisé avec un ADN viral (donneur) non spécifique de l'IN de RAV-1 (mêmes conditions que celles décrites dans le paragraphe V.2 du chapitre 'Matériels et

méthodes', sans marquage, révélation au bromure d'éthidium). Il a permis de visualiser les formes FSI+HSI, *linear* FSI et l'ADN cible surenroulé, en présence des composés potentiellement inhibiteurs (données non montrées). À la suite de cette expérience, les composés 1 et 7 (Tableau 6, page 106) semblaient avoir un effet inhibiteur sur la réaction d'intégration concertée (baisse de la quantité de produits d'intégration).

Pour montrer que ces composés agissent sur l'IN et non sur l'ADN substrat, nous avons réalisé le même test avec un ADN substrat non marqué, spécifique de l'IN de RAV-1, en présence et en l'absence d'IN (Figure 51). Une concentration de 200 µM de composés a été utilisée pour ce test. Le composé 2, ne présentant aucun effet sur la réaction d'intégration, a servi de contrôle négatif.

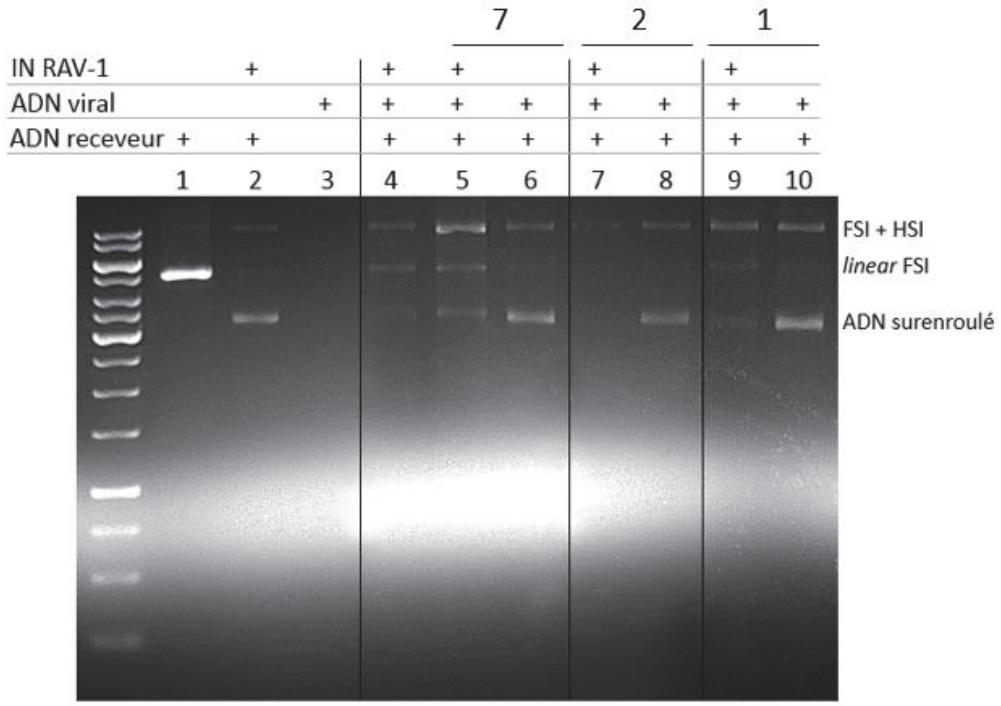


Figure 51. Test d'intégration concertée *in vitro* sans marquage radioactif, avec l'IN de RAV-1 recombinante et les composés 1, 2 et 7 potentiellement inhibiteurs (Tableau 6, page 106).

En présence des composés et en l'absence d'enzyme (pistes 6, 8 et 10), les profils obtenus sont comparables à celui du mélange IN-ADN receveur (piste 2). Dans ces pistes, malgré la présence de plasmide relâché (*linear* FSI), les composés seuls ne linéarisent pas l'ADN receveur. En présence de l'IN nous observons les produits d'intégration partielle (HSI) et concertée (FSI et *linear* FSI) (piste 4). Cette piste est semblable à celle correspondant au mélange réactionnel incubé avec le composé 7 (piste 5), bien que la quantité d'ADN récupérée après extraction soit plus importante dans la piste 5. Le composé 2 présent dans les pistes 7 et 8 ne semble pas affecter l'activité enzymatique de l'IN (contrôle négatif). À l'inverse, le

composé 1 entraîne une légère diminution de la quantité du produit d'intégration concertée *linear* FSI (comparaison des pistes 4 et 9) mais il n'a pas d'effet sur la quantité de produit d'intégration partielle HSI. Ainsi, le composé 1 aurait un effet inhibiteur sur l'intégration concertée. Ces résultats suggèrent la réalisation de tests d'intégration, à envisager avec une plus haute concentration en composés.

L'ADN viral de 250 pb n'étant pas visible sur le gel (quantité trop faible pour une révélation au bromure d'éthidium, piste 3), nous envisageons de reproduire l'expérience en utilisant un ADN donneur marqué radioactivement. L'augmentation de la concentration des composés nous permettra une observation plus nette de leur effet sur l'activité de l'IN.

III. Modélisation par homologie de l'intasome de PERV-A/C

Mon travail doctoral sur l'IN du *Gammaretrovirus* PERV-A/C a commencé par la modélisation de son intasome en complexe avec le raltégravir. Ce travail a été effectué en collaboration avec les équipes des Drs Ronfort (INRA Lyon) et Blanchard (ANSES Ploufagran), qui ont caractérisé l'effet inhibiteur de cette molécule sur PERV *in vitro* et *in cellulo* (Demange *et al.*, 2015), *cf.* chapitre 'Publication'). Aucune structure expérimentale d'IN de PERV n'étant connue à ce jour, j'ai utilisé une stratégie de modélisation 3D par homologie de séquences décrite ci-dessous, pour obtenir des informations structurales sur l'interaction entre l'intasome de PERV-A/C et le raltégravir.

Les trois domaines de l'IN de PERV-A/C ont été modélisés séparément avec l'aide du serveur SWISS-MODEL (Biasini *et al.*, 2014). Les structures des domaines NTD et CTD de l'IN du *Gammaretrovirus* MLV (codes PDB respectifs 4NZG et 2M9U (Kvaratskhelia *et al.*, 2014) ont servi d'empreintes pour modéliser les domaines équivalents dans l'IN de PERV-A/C. Notre modèle final contient ainsi un NED (retrouvé chez les *Gammaretrovirus*), comportant une hélice α et trois brins β, situé en amont du NTD (Figure 52). Le CCD de l'IN de PERV-A/C a été construit à partir de celui de l'IN de PFV (code PDB 2X6N, Rety *et al.*, 2010). Le CTD présente un repliement semblable au domaine SH3, mais également une queue flexible à son extrémité carboxy-terminale (motif EBM), essentielle à l'interaction des *Gammaretrovirus* avec les cofacteurs cellulaires de la famille BET (*cf.* § II.3.2.2, Introduction, Aiyer *et al.*, 2014, Crowe *et al.*, 2016).

La structure cristalline de l'intasome de PFV en complexe avec le raltégravir (code PDB 3OYA, Hare *et al.*, 2010) a permis de positionner les trois domaines de l'IN de PERV A/C les uns par rapport aux autres. Cette structure de complexe est quasiment identique à celle de l'intasome de PFV seul (code PDB 3L2R, Hare *et al.*, 2010) et est caractérisée par un domaine CTD positionné entre les domaines NED-NTD et CCD. Elle se superpose à la structure cristalline de l'IN de PFV avec une valeur de RMSD (*Root Mean Square Deviation*) de 2,1 Å sur 324 paires de Cα (Figure 52).

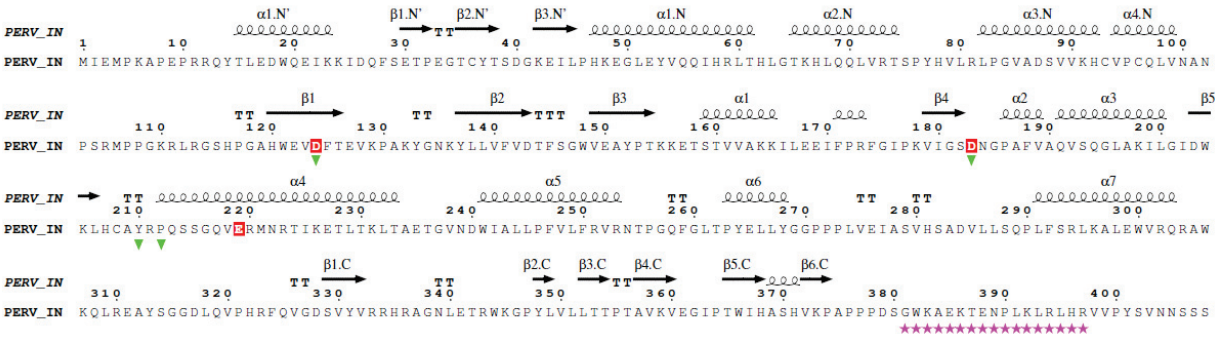


Figure 52. Haut : Séquence de l'IN de PERV-A/C avec les éléments de structures secondaires du modèle et les extensions .N', .N et .C qui symbolisent respectivement les éléments retrouvés dans les domaines NED, NTD et CTD (au-dessus de l'alignement). En rouge : résidus de la triade catalytique DDE. Triangles verts : résidus en contact avec le raltégravir. Etoiles roses : séquence de l'EBM.
Bas : Superposition des structures 3D des intasomes de PFV (en gris, code PDB 3OYA) et de PERV-A/C (modèle en vert, Demange *et al.*, 2015) en complexe avec le raltégravir, avec le logiciel PyMOL (Delano, 2008). L'ADN viral est en gris.

L'analyse de notre modèle intasome-raltégravir a permis d'identifier les résidus de l'IN de PERV-A/C en contact avec ce composé inhibiteur. Il s'agit des résidus Asp124, Asp183, Tyr210 et Pro212 (Figure 52). Les résidus d'aspartate appartiennent à la triade catalytique, les deux autres résidus sont strictement conservés dans l'IN de PFV et fortement conservés dans l'IN de VIH-1 (Figure 15, page 40). La conservation de ces résidus impliqués dans l'interaction avec le raltégravir suggère des constantes d'inhibition proches. Cette hypothèse a été validée par les expériences réalisées *in vitro* et *in cellulo* sur l'activité de l'IN et la réplication de PERV-A/C en présence de raltégravir, par l'équipe du Dr Ronfort, avec des valeurs d'IC_{50} de l'ordre du nanomolaire (Demange *et al.*, 2015).

À l'inverse, les résidus Tyr210 et Pro212 de l'IN de PERV A/C sont substitués respectivement par des résidus de sérine et de glycine dans les séquences de l'IN d'*Alpharetrovirus*. Ces substitutions sont responsables de la résistance significative de l'IN de RSV contre l'elvitégravir (Koh *et al.*, 2011).

IV. Etude moléculaire et structurale de l'IN de PERV-A/C en complexe avec le cofacteur cellulaire humain Brd2

Par la suite, je me suis focalisée sur l'étude des CTD de l'IN de PERV-A/C (312-408) et du cofacteur cellulaire humain Brd2 (666-836). En effet, il a été démontré une interaction entre ces deux domaines, par des expériences de double hybride chez la levure et co-immunoprécipitation, par les équipes des Drs Ronfort et Blanchard (Demange *et al.*, 2015). Mon premier objectif était de compléter la caractérisation de ce complexe par des données biochimiques et structurales, décrites dans la suite du manuscrit et prochainement publiées.

Mon second objectif était de résoudre la structure du complexe formé par ces deux CTD, par cristallographie aux rayons X. Comme décrit pour l'IN de MLV et Brd4 (Crowe *et al.*, 2016), le motif terminal EBM du CTD de l'IN de PERV A/C (zone flexible, 381-397) pourrait se structurer en interagissant avec le domaine ET du CTD de Brd2 (676-738). Une prédiction du désordre intrinsèque effectuée avec le serveur GlobPlot (Linding *et al.*, 2003) sur la séquence du CTD de Brd2 (666-836, Figure 53) montre un fort désordre de l'extrémité carboxy-terminale riche en serine (812-836, Figure 53). Les Drs Ronfort et Blanchard ont montré que cette partie n'interagissait pas avec le CTD de l'IN de PERV A/C (article en préparation). Nous avons donc travaillé sur le fragment 666-811 du CTD de Brd2. Dans ce fragment, la région 739-811 en aval du domaine ET n'a pas d'homologue structural connu (Figure 53).

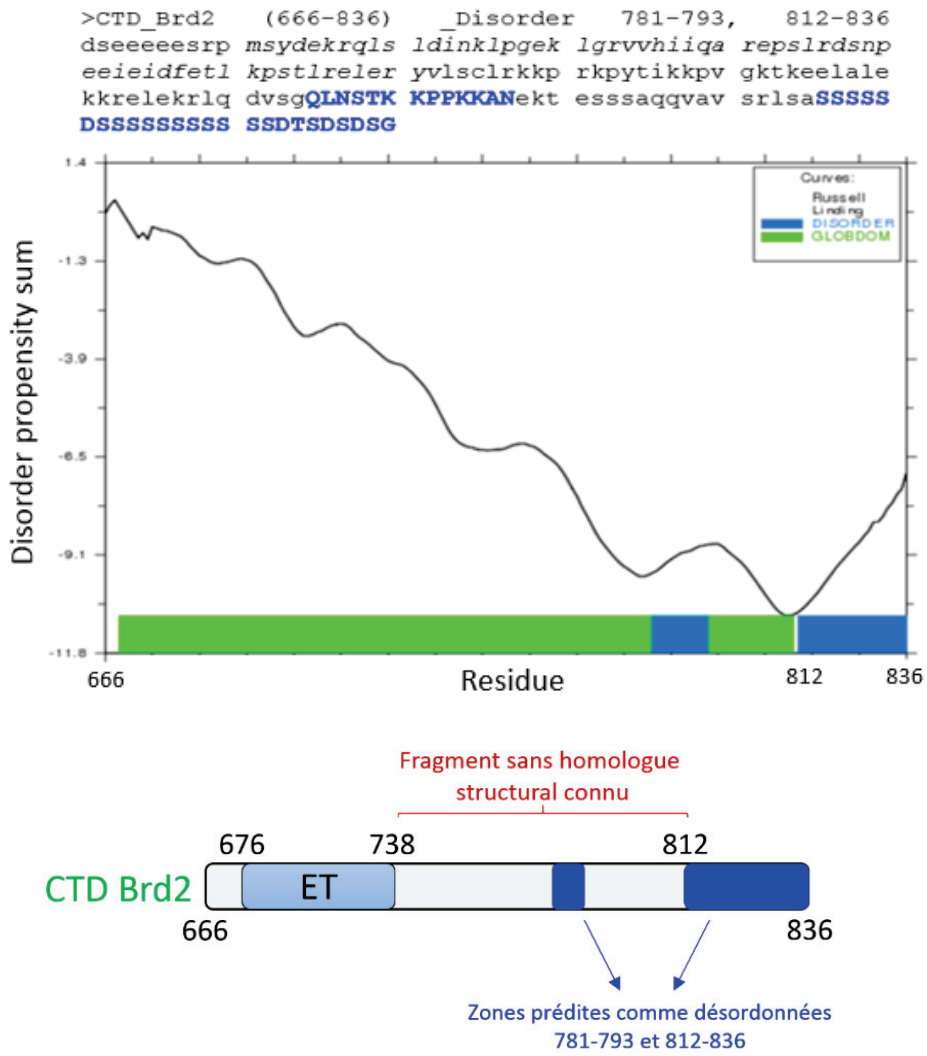


Figure 53. Haut. Prédiction du désordre intrinsèque du CTD de l'isoforme 2 de Brd2 humain (666-836) avec le serveur Globplot (Linding *et al.*, 2003). Les résidus de la région ET (676-738) sont en italique. Les zones prédites comme désordonnées sont en bleu (781-793 et 812-836). Bas. Schéma de l'organisation du CTD de Brd2 (isoforme 2).

J'ai commencé mes recherches biochimiques et structurales sur le complexe, en travaillant d'abord sur le CTD isolé de l'IN de PERV A/C (312-408).

1. Etude du CTD isolé de l'IN PERV A/C

1. Surproduction et purification

La souche bactérienne BL21(DE3) a été transformée par le plasmide pETM-30, qui porte le gène codant le CTD de l'IN de PERV-A/C. Un des clones résultants a servi à inoculer la préculture (milieu LB supplémenté en kanamycine 50 µg/ml). Après une nuit d'incubation à 37°C, la préculture a été versée dans un volume de milieu LB plus grand (culture) contenant de la kanamycine (50 µg/ml), placé à 37°C, sous agitation (180 rpm). À une valeur de DO de 0,6, l'expression du gène d'intérêt a été induite par 1 mM d'IPTG, à 37°C, à 180 rpm pendant trois

heures. Une lyse bactérienne par sonication a ensuite été initiée, dans un tampon de lyse (20 mM Tris HCl pH 8, 500 mM NaCl, 25 mM imidazole) contenant du lysozyme (0,2 mg/ml) et une pastille de cocktail d'anti-protéases (*cf.* § II.2, Matériels et méthodes).

Le protocole nécessaire à la purification du CTD de l'IN de PERV-A/C, est identique à celui du CCD H103C de l'IN de RAV-1 (*cf.* § I.1.1, Résultats et discussion). Sur le chromatogramme de purification, les pics A et B correspondent respectivement aux fractions retenues et de lavage (contaminants fixés de façon aspécifique à la résine) (Figure 54A). Le pic C d'élution apparaît pour une concentration en imidazole proche de 180 mM (gradient tampon d'élution/tampon de lyse de 16%). L'augmentation de l'absorbance observée sur la partie droite du chromatogramme résulte de l'augmentation de la concentration en imidazole qui absorbe également à 280 nm (Figure 54A). Les fractions d'élution ont été rassemblées, la solution protéique résultante a été dosée et analysée par SDS-PAGE (Figure 54B).

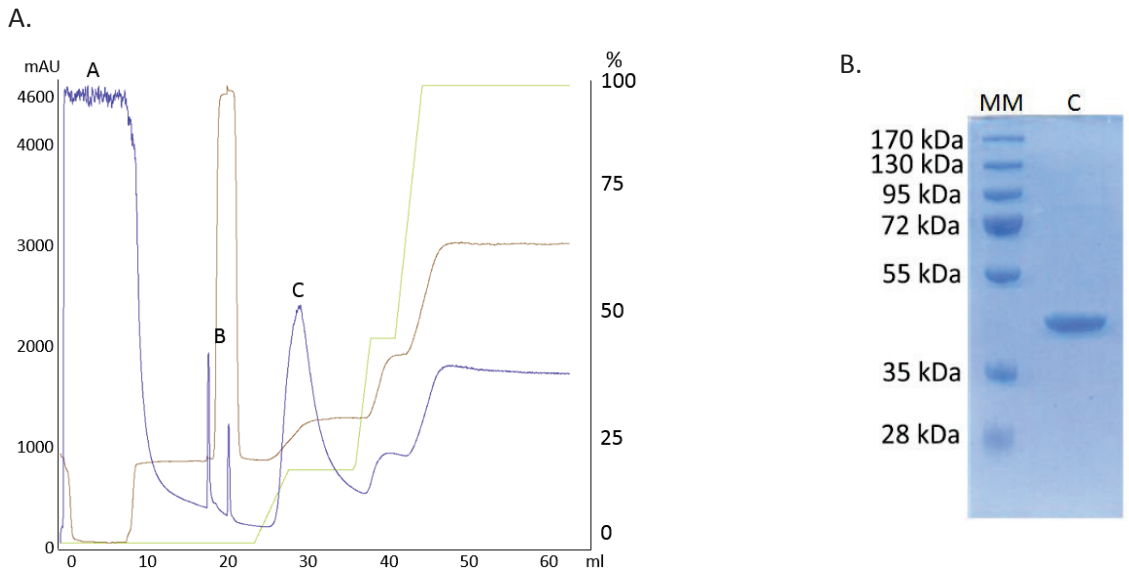


Figure 54. A. Chromatogramme de purification du CTD de l'IN de PERV-A/C par chromatographie d'affinité pour le nickel. Courbe bleue : suivi de l'absorbance à 280 nm en sortie de colonne (axe des ordonnés gauche) en fonction du volume d'élution (axe des abscisses). Courbe marron : suivi de la conductance du tampon. Courbe verte : suivi du gradient d'imidazole (% tampon d'élution/tampon de lyse, axe des ordonnées droit). Fraction A : fraction non retenue. Fraction B : fraction de lavage. Fraction C : fraction d'élution.
B. Analyse par SDS-PAGE 12% du CTD de l'IN de PERV-A/C en fin de purification, avec révélation au bleu de Coomassie. Piste MM : marqueurs de masses moléculaires connues. Piste C : CTD de l'IN de PERV-A/C purifié.

La bande majoritaire, entre les marqueurs de masse moléculaire de 36 kDa et 55 kDa, a été identifiée par analyse en spectrométrie de masse. Elle correspond au CTD de l'IN de PERV-A/C fusionné à l'étiquette poly-histidine et à la GST (environ 40 kDa).

2. Clivage de l'étiquette de purification

Après dessalage des fractions d'élution collectées, le clivage de l'étiquette polyhistidine-GST par la protéase du TEV a été mis au point. Pour ce faire, différents rapports protéase du TEV/CTD d'intérêt (w/w) ont été testés à 4°C et à 25°C pendant une nuit (Figure 55).

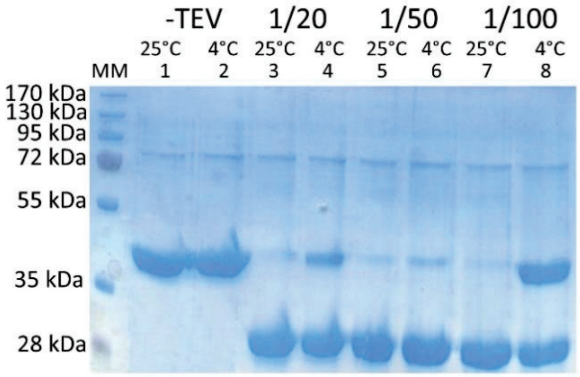


Figure 55. Analyse par SDS-PAGE des tests de clivage de l'étiquette de purification du CTD de l'IN de PERV-A/C, par la protéase du TEV. Piste MM : marqueurs de masses moléculaires connues. Pistes 1 et 2 : échantillon protéique sans protéase, à 25°C et 4°C. Pistes 3 et 4 : rapport massique protéase du TEV/protéine 1/20 à 25°C et 4°C. Pistes 5 et 6 : rapport massique protéase du TEV/protéine 1/50 à 25°C et 4°C. Pistes 7 et 8 : rapport massique protéase du TEV/protéine 1/100 à 25°C et 4°C.

En l'absence de la protéase du TEV, une bande majoritaire est observée. Elle correspond au CTD de l'IN de PERV-A/C avec son étiquette poly-histidine (Figure 55, pistes 1 et 2).

Les fractions contenant la protéase du TEV présentent une bande majoritaire située entre les marqueurs de masses moléculaires de 28 kDa et 35 kDa (Figure 55, pistes 3 à 8). Il pourrait s'agir de la protéine sans étiquette poly-histidine-GST, après digestion par la protéase du TEV. L'identification par analyse par spectrométrie de masse a confirmé la présence du CTD clivé (bande basse).

Il apparaît que le clivage de l'étiquette de purification est partiel à une température d'incubation de 4°C (Figure 55, pistes 4, 6 et 8). Il semble maximal à 25°C, pour un rapport protéase du TEV/protéine de 1/100 (Figure 55, piste 7), mais la protéine d'intérêt a tendance à s'agréger à 25°C (données non montrées). Après optimisation du clivage de l'étiquette de purification et pour assurer un clivage maximal, un rapport massique protéase du TEV/protéine 1/50 est utilisé dans les fractions précédemment rassemblées, avec une incubation à 4°C sur la nuit (Figure 55, piste 6). La solution protéique est ensuite réinjectée sur une nouvelle colonne de chromatographie d'affinité pour le nickel. Le CTD de l'IN de PERV-A/C dépourvu d'étiquette de purification ne se fixe pas à la colonne. Après concentration de cette solution

protéique dans le tampon de lyse dépourvu d'imidazole (tampon final), des essais de cristallisation ont été initiés.

3. Essais de cristallisation

Environ 500 conditions de cristallisation ont été testées pour le CTD de l'IN de PERV-A/C (tampon final), avec les kits commerciaux PEGIon, PEGII, CrystalScreen (*Hampton Research*), $AmSO_4$Suite et JCSG+ (*Qiagen*). Les essais ont été réalisés par la méthode de diffusion de vapeur en gouttes assise et suspendue, à 19°C, à des concentrations protéiques de 3 mg/ml à 20 mg/ml. Ils n'ont pas permis d'obtenir des cristaux de protéine.

4. Criblage d'interactions avec des histones par *histone array*

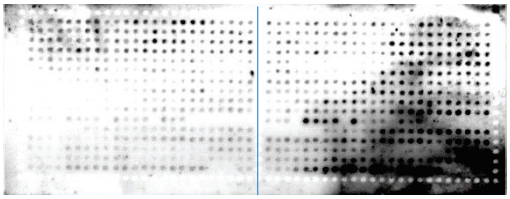
Depuis 2014, notre équipe est impliquée dans une collaboration avec l'équipe du Dr Parissi (MFP Bordeaux). Par des approches *in vitro* et *in cellulo*, celle-ci a observé une interaction entre le CTD de l'IN de VIH-1 et la queue d'histone H4 modifiée (une méthylation sur le résidu lysyl 20 ou H4K20me1) en l'absence du partenaire cellulaire LEDGF/p75. Cette marque épigénétique est associée aux régions transcriptionnellement actives, connues comme étant des sites préférentiels d'intégration de l'IN de VIH-1 (Schroder *et al.*, 2002, Vranckx *et al.*, 2016).

Des expériences de *docking in silico*, menées au sein de notre équipe par le Dr Robert, suggèrent l'implication d'une crevasse du CTD de l'IN de VIH-1 dans la fixation de la queue H4K20me1. Cette crevasse étant observée dans tous les CTD des IN rétrovirales, nous avons profité de l'optimisation de la surproduction du CTD de l'IN de PERV-A/C pour tester cette interaction pour la première fois sur un *Gammaretrovirus*. Fin 2016, j'ai donc réalisé des tests d'interactions entre ce CTD et les queues d'histones modifiées, avec l'équipe du Dr Parissi à Bordeaux.

Après révélation, des spots clairement définis sont apparus sur la puce à histones (Figure 56A). Ils reflètent l'interaction entre le CTD de l'IN de PERV-A/C et des histones H2B et H3 acétylées ou méthylées. Les histones acétylées participent à l'activation de la transcription, tandis que les histones méthylées contribuent à sa répression (Kouzarides, 2007). Le logiciel *Array Analyze* (*Active Motif* ®) a permis d'identifier les histones H2BK5ac, H2BK12ac, H3K27ac (acétylées sur les résidus lysyl 5, 12 et 27), -me2, -me3 (2 ou 3 méthylations sur les résidus lysyl 27) et -me1R26me2s (une méthylation sur le résidu K27, deux méthylations symétriques (s) sur le résidu R26) qui interagissent avec le CTD d'intérêt (Figure 56B). L'histone H4K20me1, retrouvé dans les tests menés sur le CTD de l'IN de VIH-

1, n'apparaît pas comme spécifique du CTD de l'IN de PERV-A/C. En effet, cette histone n'est pas associée aux sites d'initiation de la transcription, connus comme étant des sites préférentiels d'intégration des *Gammaretrovirus*.

A.



B.

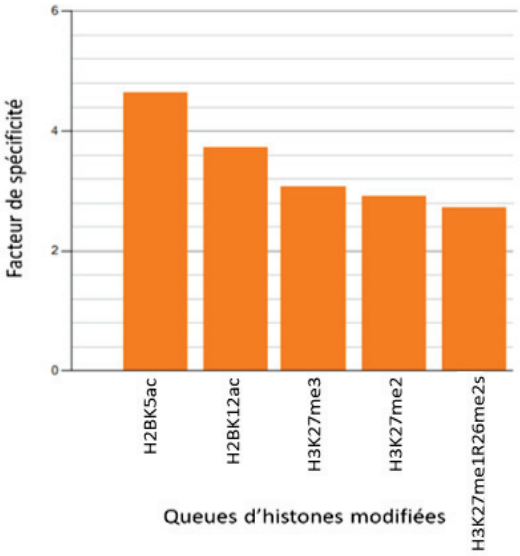


Figure 56. A. Puce à histones en présence du CTD de l'IN de PERV-A/C, après révélation ECL (duplicat).

B. Profils de spécificité représentant le facteur de spécificité (rapport entre la moyenne d'intensités des spots marqués et celle des spots non marqués et marqués) en fonction des queues d'histones modifiées.

La comparaison des profils a montré une spécificité du CTD de l'IN de PERV-A/C plus importante pour l'histone H2BK5ac que pour l'histone H2BK12ac (Figure 56B). Les histones H2BK5ac et K12ac participent à l'activation de la transcription (Kouzarides, 2007). De plus, l'histone H2BK5ac est décrite comme interagissant spécifiquement avec les sites d'initiation de la transcription (*Histone Infobase*, http://www.actrec.gov.in/histome/), ciblés préférentiellement par l'IN des *Gammaretrovirus* pour une intégration efficace (Moalic *et al.*, 2006).

Les histones H3K27me2, H3K27me3, me1R26me2s, participant à la répression transcriptionnelle (Kouzarides, 2007) et H3K27ac interagissent également avec le CTD de l'IN de PERV-A/C. De façon intéressante, la présence de la mutation S28P sur l'histone H3 conduit à la perte totale d'interaction.

L'ensemble de ces résultats révèle une spécificité d'interaction du CTD de l'IN de PERV-A/C pour les histones H2B et H3, qu'elles soient acétylées ou méthylées. Ces observations permettent d'envisager un modèle hypothétique, selon lequel les histones H2B acétylées participent au ciblage des sites préférentiels d'intégration de l'IN de PERV-A/C et activent la

transcription de l'ADN dans ces régions. L'interaction avec les histones H3 méthylées constituerait une boucle de contrôle négatif lors de l'intégration, en permettant de freiner l'activité transcriptionnelle.

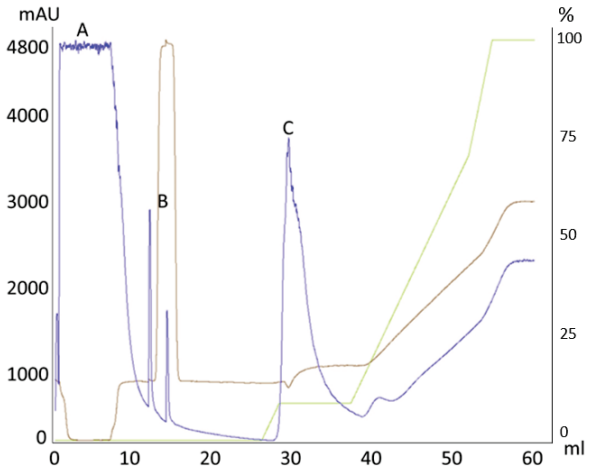
2. Etude du CTD isolé de Brd2

1. Surproduction et purification

Le plasmide pPROEX-Hta, porteur du gène codant le CTD de Brd2, a servi à la transformation des bactéries BL21(DE3). Un protocole similaire à celui utilisé pour le CTD de l'IN de PERV-A/C a été mis en œuvre pour surproduire ce domaine. Seul le milieu de culture TB supplémenté avec de l'ampicilline (100 µg/ml) a remplacé le milieu LB contenant de la kanamycine. La lyse bactérienne a été réalisée par sonication (8 cycles de 30 secondes, sur glace), dans un tampon de lyse (20 mM Tris HCl pH 8, 500 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoéthanol) contenant du lysozyme (0,2 mg/ml), du $MgCl_2$ (10 mM) et de la DNase (20 µg/ml). Le protocole de purification du CTD de Brd2 est identique à celui du CCD H103C de l'IN de RAV-1 (*cf.* § I.1.1, Résultats et discussion), avec des tampons de purification différents (tampon de lavage : tampon de lyse avec 1 M NaCl, tampon d'élution : tampon de lyse avec 1 M imidazole, tampon final : tampon de lyse sans imidazole).

Le chromatogramme de purification (Figure 57A) présente un pic A (fraction non retenue), deux pics B (fractions de lavage) et un pic d'élution C obtenu avec un gradient tampon de lyse/tampon d'élution d'environ 13%, soit à 150 mM d'imidazole. Les fractions d'élution ont été rassemblées et analysées par SDS-PAGE (Figure 57B).

A.



B.

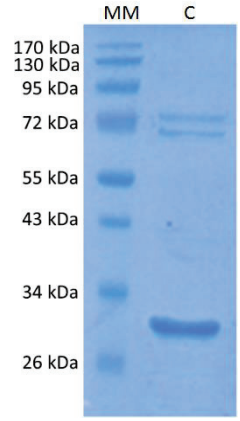


Figure 57. A. Chromatogramme de purification du CTD de Brd2 par chromatographie d'affinité pour le nickel. Courbe bleue : suivi de l'absorbance à 280 nm en sortie de colonne (axe des ordonnés gauche) en fonction du volume d'élution (axe des abscisses). Courbe marron : suivi de la conductance du tampon. Courbe verte : suivi du gradient d'imidazole (% tampon d'élution/tampon de lyse, axe des ordonnées droit). Pic A : fraction non retenue. Pics B : fractions de lavage. Pic C : fraction d'élution.
B. Analyse par SDS-PAGE 12% du CTD de Brd2 en fin de purification, avec révélation au bleu de Coomassie. Piste MM : marqueurs de masses moléculaires connues. Piste C : CTD de Brd2 purifié (pic C).

Une analyse par spectrométrie de masse a permis d'identifier les trois bandes visibles sur le gel d'électrophorèse (Figure 57B). Il s'agit du CTD de Brd2 (entre les marqueurs de 26 kDa et 34 kDa) et de protéines intrinsèques d'*E. coli* (autour du marqueur de 72 kDa). Malgré la présence de ces contaminants, nous avons réalisé un dessalage en tampon final (tampon de lyse sans imidazole) avant de cliver l'étiquette de purification.

2. Clivage de l'étiquette poly-histidine

Différents rapports massiques protéase du TEV/protéine d'intérêt ont été testés à 20°C et à 4°C, pendant une nuit, pour tenter de cliver l'étiquette poly-histidine du CTD de Brd2 (Figure 58A). Le clivage de l'étiquette de purification est total à 20°C avec les trois rapports massiques protéase du TEV/CTD de Brd2 testés (Figure 58A, pistes 3, 4, 5 et 7). Nous avons donc sélectionné la condition de clivage à 20°C, avec un rapport massique protéase du TEV/CTD de Brd2 1/20, pour ôter l'étiquette.

Après injection de la solution protéique traitée à la protéase du TEV sur une colonne Ni-NTA, la fraction non retenue a été analysée par SDS-PAGE (Figure 58B).

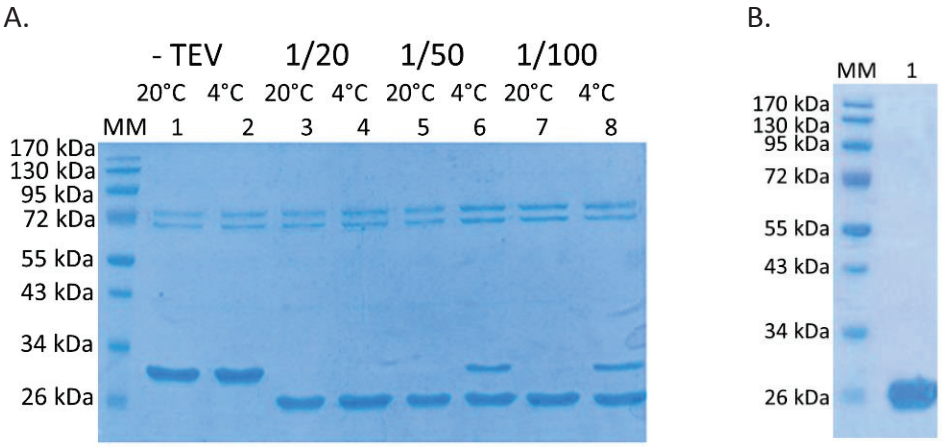


Figure 58. A. Analyse par SDS-PAGE des tests de clivage de l'étiquette de purification du CTD de Brd2, par la protéase du TEV. Piste MM : marqueurs de masses moléculaires connues. Pistes 1 et 2 : échantillon protéique sans protéase, à 20°C et 4°C. Pistes 3 et 4 : rapport protéase du TEV/protéine 1/20 (w/w) à 20°C et 4°C. Pistes 5 et 6 : rapport protéase du TEV/protéine 1/50 (w/w) à 20°C et 4°C. Pistes 7 et 8 : rapport protéase du TEV/protéine 1/100 à 20°C et 4°C (w/w).
B. Analyse par SDS-PAGE de la fraction non retenue sur colonne Ni-NTA, après clivage de l'étiquette poly-histidine. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : fraction non retenue, incubée une nuit à 20°C avec la protéase du TEV (rapport massique 1/20).

Cette seconde étape de purification a permis d'éliminer les contaminants protéiques précédemment observés et de récupérer la protéine pure sans étiquette.

3. Formation du complexe protéique avec les CTD isolés

Une fois les CTD de Brd2 et de l'IN de PERV-A/C surproduits, purifiés et dépourvus de leur étiquette de purification, ils ont été concentrés à 2 mg/ml pour une analyse DLS. Elle a permis de contrôler l'absence d'agrégats dans ces échantillons (données non montrées). Les deux partenaires protéiques ont ensuite été mis en contact pour tenter de former le complexe protéique d'intérêt.

Le mélange des CTD dans leurs tampons finaux respectifs a engendré un phénomène d'agrégation qui n'a pas permis de caractériser le complexe protéique. Nous avons alors tenté d'utiliser un tampon commun aux deux CTD, composé de 20 mM Tris HCl pH 8 et 500 mM NaCl. D'autres tampons ont été testés pour éviter la formation d'agrégats, mais aucun d'entre eux n'a conduit à un mélange protéique non agrégé (données non montrées).

Une nouvelle méthode, nommée stratégie « multiTEV », a été mise en place pour obtenir le complexe protéique d'intérêt soluble. Elle est basée sur l'assemblage du complexe protéique directement dans les bactéries d'expression. Elle a été développée par le Dr Ruff à l'IGBMC Illkirch sur un complexe impliquant les domaines de certains récepteurs et un co-activateur transcriptionnel (Levy *et al.*, 2016).

4. Stratégie « multiTEV »

1. Surproduction et purification

La souche bactérienne BL21(DE3) a été transformée par le plasmide pET28a, porteur de la construction « multiTEV ». Un clone résultant a été introduit dans du milieu LB supplémenté en kanamycine (50 µg/ml). Le protocole de surproduction du CTD de l'IN de PERV-A/C a été utilisé pour surproduire le complexe d'intérêt. Pour isoler ce dernier, une première étape de purification du CTD de l'IN de PERV-A/C, fusionné à une étiquette poly-histidine, a été effectuée. Ainsi, la fraction soluble obtenue après lyse bactérienne a été injectée sur une colonne Ni-NTA. L'absorbance à 280 nm et la conductance, qui traduisent les changements de concentrations salines, ont été suivies tout au long de la purification (Figure 59A).

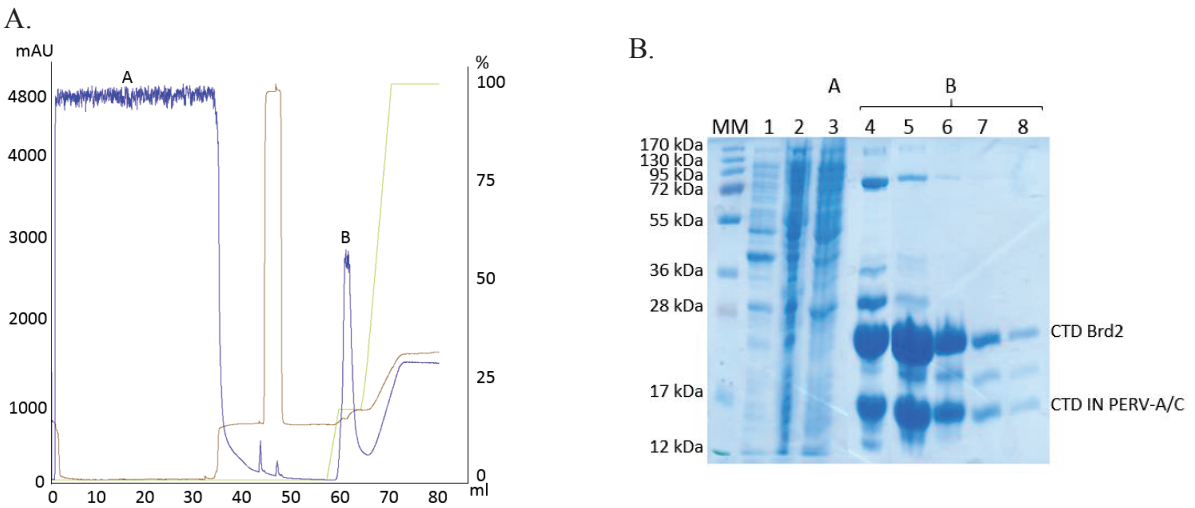


Figure 59. A. Chromatogramme de purification du complexe IN_{CTD}-$Brd2_{CTD}$ par chromatographie d'affinité pour le nickel. Courbe bleue : suivi de l'absorbance à 280 nm en sortie de colonne (axe des ordonnés gauche) en fonction du volume d'élution (axe des abscisses). Courbe marron : suivi de la conductance du tampon. Courbe verte : suivi du gradient d'imidazole (% tampon d'élution/tampon de lyse, axe des ordonnées droit). Pic A : fraction non retenue. Pic B : fraction d'élution.
B. Analyse par SDS-PAGE 12% du complexe IN_{CTD}-$Brd2_{CTD}$ en fin de purification, avec révélation au bleu de Coomassie. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : fraction insoluble. Piste 2 : fraction soluble. Piste 3 : fraction non retenue. Pistes 4 à 8 : fractions d'élution (pic B).

Le pic A traduit le passage de la fraction non retenue à travers la colonne de purification (Figure 59A). Les protéines bactériennes, n'ayant pas d'affinité pour la colonne, ont ainsi été éliminées dans le tampon de lyse (500 mM NaCl). Le tampon de lavage (1 M NaCl), utilisé pour laver la colonne, a permis d'éliminer les protéines aspécifiques. Dans ce cas précis, aucun contaminant n'a été décroché de la colonne. L'étape d'élution a ensuite été initiée par un gradient d'imidazole (tampon d'élution avec 1 M imidazole). Les protéines fixées à la colonne ont été éluées avec un gradient tampon d'élution/tampon de lyse de 15%, soit 170 mM d'imidazole (Figure 59A, pic B). Après élution, l'absorbance à 280 nm se stabilise mais sans

retour à une valeur basale. Ce phénomène s'explique par l'augmentation de la concentration en imidazole qui absorbe également à cette longueur d'onde. Les fractions collectées ont été analysées par SDS-PAGE (Figure 59B).

Aucune bande correspondant au précurseur protéique de 60 kDa n'est visible sur le gel. De plus, deux bandes majoritaires proches des marqueurs de masses moléculaires de 17 kDa et 28 kDa, ont été révélées (Figure 59B, pistes 2 et 4 à 8). Une analyse par spectrométrie de masse a confirmé qu'il s'agissait des CTD de l'IN de PERV-A/C (12 kDa) et de Brd2 (18 kDa). Du fait de son étiquette Strep-tag®II, le CTD de l'IN de PERV-A/C n'aurait pas pu être récupéré s'il n'appartenait pas au complexe entre CTD. Les conditions dénaturantes appliquées sur le gel ont ensuite entraîné la dissociation du complexe protéique. L'ensemble de ces observations indiquent que la protéase du TEV est fonctionnelle. Ainsi, la construction « multiTEV » permet de produire et de purifier le complexe natif formé par les CTD de l'IN de PERV-A/C et de Brd2.

Il est important de noter que le profil de migration des protéines d'intérêt tend à surestimer leurs masses moléculaires respectives. En effet, la bande identifiée comme étant le CTD de Brd2 de 18 kDa est plus proche du marqueur de 28 kDa que de celui de 17 kDa. Il est en est de même pour le CTD de l'IN de PERV-A/C de 12 kDa qui migre au-dessus du marqueur de même masse moléculaire. Ces profils de migration sur SDS-PAGE ne sont pas dus à la présence de modifications post-traductionnelles, puisque les surproductions protéiques ont lieu en système bactérien. Les points isoélectriques des CTD de l'IN de PERV-A/C et de Brd2 étant basiques, plus de molécules de SDS se fixeraient dessus pour les charger négativement, ce qui expliquerait ce « retard » de migration.

L'analyse des fractions insoluble (Figure 59B, piste 1) et soluble (Figure 59B, piste 2) nous indique que la lyse bactérienne a été efficace, avec la présence d'une quantité plus importante des deux CTD partenaires dans le surnageant que dans le culot. Les protéines sans étiquette poly-histidine, ou n'ayant pas d'affinité pour la colonne, ont été directement éluées (Figure 59B, piste 3). Les fractions d'élution contiennent les CTD d'intérêt (Figure 59B, pistes 4 à 8).

Les fractions d'élution ont été rassemblées puis injectées sur une colonne Strep-Tactin®, préalablement rincée avec le tampon W. Cette seconde étape de purification est basée sur la présence de l'étiquette Strep-tag®II, portée par le CTD de l'IN de PERV-A/C engagé dans le

complexe, l'objectif étant de purifier le complexe protéique seul en éliminant le CTD de Brd2 libre. L'absorbance à 280 nm et la conductance ont été suivies en temps réel (Figure 60A).

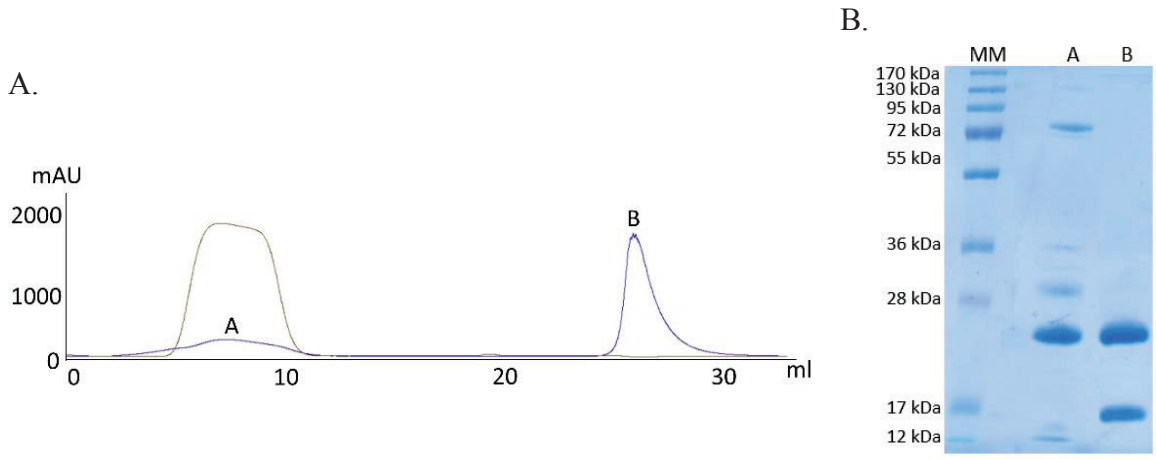


Figure 60. A. Chromatogramme de purification du complexe IN_{CTD}-$Brd2_{CTD}$ par chromatographie d'affinité pour la streptavidine. Courbe bleue : suivi de l'absorbance à 280 nm en sortie de colonne (axe des ordonnés) en fonction du volume d'élution (axe des abscisses). Courbe marron : suivi de la conductance du tampon. Pic A : fraction non retenue. Pic B : fraction d'élution. B. Analyse par SDS-PAGE 12% du complexe IN_{CTD}-$Brd2_{CTD}$ en fin de purification, avec révélation au bleu de Coomassie. Piste MM : marqueurs de masses moléculaires connues. Piste A : fraction non retenue (pic A). Piste B : fraction d'élution (pic B).

Le pic A représente le relargage des protéines n'ayant pas d'affinité pour la colonne Strep-Tactin® (fraction non retenue, Figure 60A). Le tampon de l'éluat injecté contient plus de sels que le tampon W (0,5 M NaCl contre 150 mM NaCl), ce qui se traduit par une augmentation de la conductance. La colonne est à nouveau rincée avec le tampon W ce qui provoque une diminution de la conductance et un retour à sa valeur initiale. Afin d'éluer les protéines fixées sur la colonne, celle-ci est lavée avec le tampon E contenant le compétiteur, la desthiobiotine. L'élution se traduit par une nette augmentation de l'absorbance à 280 nm au niveau du pic B (Figure 60A).

Les fractions non retenues et d'élution ont été analysées par SDS-PAGE. Sur la piste 1 (fraction non retenue, Figure 60B), les bandes à environ 75 kDa et 28 kDa correspondent à des contaminants. Le CTD de Brd2 seul a également été éliminé (bande la plus intense, masse moléculaire comprise entre 17 kDa et 28 kDa). Les fractions d'élution rassemblées contiennent uniquement les deux partenaires protéiques d'intérêt (Figure 60B, piste 2). À partir d'un litre de culture bactérienne et à l'issue de ces deux étapes de purification, environ 5 mg de complexe protéique pur ont été obtenus.

2. Identification du complexe protéique par *western-blot*

Les CTD de Brd2 et de l'IN de PERV-A/C ont été respectivement identifiés avec un anticorps et de la Strep-Tactin® reconnaissant respectivement l'étiquette de purification poly-histidine et Strep-tag®II (Figure 61).

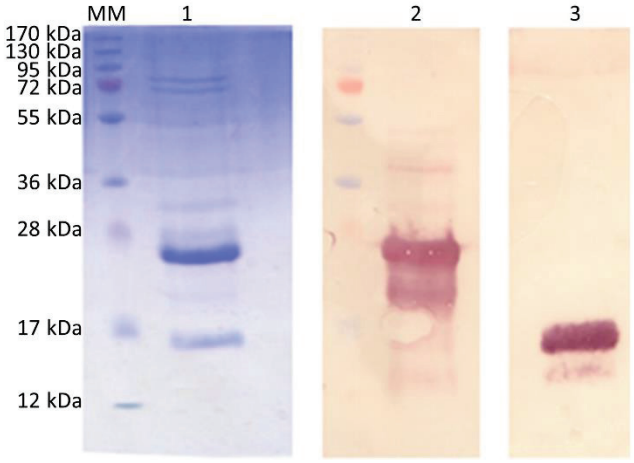


Figure 61. Analyse par *western-blot* de la fraction d'élution. MM : marqueurs de masses moléculaires connues. Piste 1 : fractions d'élution, coloration au bleu de Coomassie. Piste 2 : détection de l'étiquette poly-histidine. Piste 3 : détection de l'étiquette Strep-tag®II.

Des bandes sont observées aux tailles attendues, ces profils d'immunoempreinte confirment donc la présence des deux protéines d'intérêt dans la solution protéique purifiée. Cette solution a ensuite été concentrée à 6 mg/ml, avant d'être analysée par chromatographie d'exclusion de taille.

3. Caractérisation du complexe IN_{CTD}-$Brd2_{CTD}$ en solution

1. Etat oligomérique en solution

1. Chromatographie d'exclusion de taille (SEC)

Pour déterminer l'état oligomérique du complexe d'intérêt, l'échantillon a été injecté sur une colonne Superdex 75 HR 10/30 (*GE Healthcare*) (Figure 62).

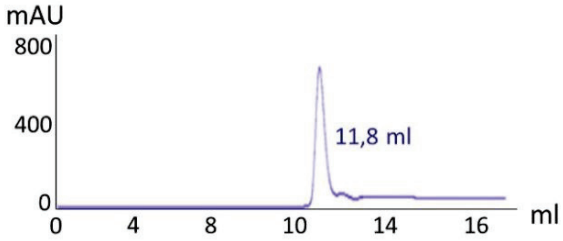


Figure 62. Profil d'élution du complexe IN_{CTD}-$Brd2_{CTD}$, formé par les CTD de Brd2 et de l'IN de PERV-A/C, obtenu par chromatographie d'exclusion de taille (Superdex 75 HR 10/30). Courbe bleue : suivi de l'absorbance à 280 nm (axe des ordonnées) en fonction du volume d'élution (axe des abscisses).

Le profil d'élution présente un pic unique, ce qui suggère que le complexe IN_{CTD}-$Brd2_{CTD}$ est présent sous une seule forme oligomérique. Sa masse moléculaire a été estimée à partir du volume d'élution et après calibration de la colonne avec des protéines globulaires standard.

Le volume d'élution correspondant au maximum d'intensité du pic est d'environ 11,8 ml. Ce volume est proche de celui obtenu pour le chymotrypsinogène A (12,4 ml), dont la masse moléculaire est de 25 kDa. La droite d'étalonnage (logarithme des masses moléculaires des protéines globulaires standard en fonction du volume d'élution) a permis d'estimer une masse moléculaire proche de 30 kDa. À 6 mg/ml, le complexe semble former un hétérodimère en solution : il est composé d'un monomère de CTD de Brd2 (18 kDa) et d'un monomère de CTD de l'IN de PERV-A/C (12 kDa). Cette stœchiométrie 1:1 est en adéquation avec la structure RMN du complexe formé par les domaines de protéines homologues à celles étudiées (EBM de l'IN de MLV et domaine ET de Brd4, Crowe *et al*., 2016).

2. *Pontage covalent au BS^3 et au GTA*

L'analyse par pontage covalent a servi à caractériser les états d'oligomérisation du complexe IN_{CTD}-$Brd2_{CTD}$. Ce pontage, en présence des agents de pontage chimique BS^3 et GTA, permet de figer de façon covalente et irréversible les oligomères en solution. Les produits de la réaction sont ensuite analysés par SDS-PAGE (Figure 63).

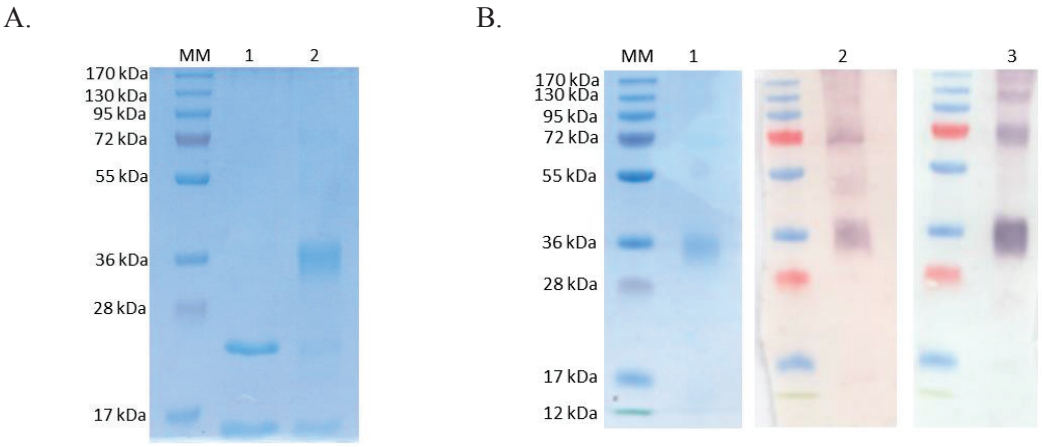


Figure 63. Analyses de l'état oligomérique du complexe IN_{CTD}-$Brd2_{CTD}$ après pontage covalent au BS^3 et au GTA. A. Analyse SDS-PAGE du pontage covalent au BS^3. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : complexe IN_{CTD}-$Brd2_{CTD}$ sans agent de pontage. Piste 2 : complexe IN_{CTD}-$Brd2_{CTD}$ incubé avec le BS^3. B. Analyse du complexe IN_{CTD}-$Brd2_{CTD}$ après pontage covalent au GTA. Piste 1 : coloration au bleu de Coomassie, piste 2 : détection de l'étiquette Strep-tag®II, piste 3 : détection de l'étiquette poly-histidine.

En présence de BS^3, l'observation d'une bande majoritaire (masse moléculaire proche de 36 kDa, Figure 63A, piste 2) suggère que le complexe IN_{CTD}-$Brd2_{CTD}$ dimérique est bien dans une stœchiométrie 1:1 en solution. Une analyse par spectrométrie de masse a confirmé la présence des deux CTD dans ce complexe purifié. Ce résultat est en accord avec celui obtenu par SEC, à savoir un hétérodimère majoritairement présent en solution. Toutefois, deux bandes minoritaires visibles sur cette même piste traduisent la présence des formes monomériques libres des partenaires (identification par analyse par spectrométrie de masse), non détectées par SEC. Elles pourraient être la conséquence d'une réaction incomplète du BS^3 due à la présence de Tris dans la solution tampon. En effet, les molécules de Tris peuvent entrer en compétition avec les fonctions amines primaires des protéines pour la réaction de pontage.

L'analyse par SDS-PAGE du pontage covalent au GTA montre la présence d'une bande majoritaire, correspondant à une masse moléculaire d'environ 36 kDa (Figure 63B, piste 1). Contrairement au pontage en présence de BS^3, la réaction avec le GTA semble totale (absence de bandes correspondant aux partenaires seuls). Des bandes de faible intensité et de plus hauts poids moléculaires sont également visibles sur ce gel d'électrophorèse. Pour vérifier la présence du complexe IN_{CTD}-$Brd2_{CTD}$ après pontage au GTA, une identification par *western-blot* a été réalisée (Figure 63B, pistes 2 et 3). Un anticorps primaire anti-histidine et de la Strep-Tactin® couplée directement à la phosphatase alcaline, reconnaissant les étiquettes poly-histidine et Strep-tag®II respectives des CTD de Brd2 et de l'IN de PERV-A/C, ont été utilisés. Sur les deux membranes d'immuno-empreinte, les deux CTD sont détectés. Ils participent à des formes oligomériques de masses moléculaires proches de 36 kDa et 72 kDa. Ces données sont en accord avec la forme dimérique du complexe (environ 30 kDa), composée d'un monomère de chaque CTD, mise en évidence par SEC et par pontage au BS^3. Toutefois, elles décrivent aussi une forme tétramérique (72 kDa) du complexe IN_{CTD}-$Brd2_{CTD}$.

3. SEC-MALS

Afin de confirmer ces observations, des expériences de SEC-MALS ont été réalisées avec une solution de complexe IN_{CTD}-$Brd2_{CTD}$ à 6 mg/ml (Figure 64).

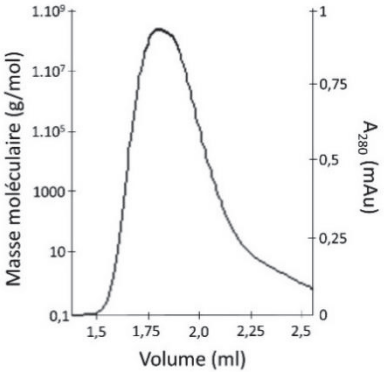


Figure 64. Chromatogramme de SEC-MALS d'une solution de complexe IN_{CTD}-$Brd2_{CTD}$ à 6 mg/ml. Axe des abscisses : volume d'élution (en ml), axes des ordonnées : (gauche) masse moléculaire (en g/mol) et (droite) absorbance à 280 nm (en mAu).

Le chromatogramme obtenu décrit un pic unique correspondant à une macromolécule de 29,6 kDa. Le complexe IN_{CTD}-$Brd2_{CTD}$, à 6 mg/ml, est donc dimérique en solution (un monomère de chaque CTD).

L'analyse de l'état d'oligomérisation du complexe IN_{CTD}-$Brd2_{CTD}$ a été réalisée avec différentes techniques, dans les mêmes conditions expérimentales (protéine pure, solution tampon, concentration protéique). Les expériences de SEC et de SEC-MALS décrivent un complexe IN_{CTD}-$Brd2_{CTD}$ dimérique, tandis que les tests de pontage covalent montrent deux formes oligomériques, à savoir un hétérodimère et un tétramère.

Lors des analyses par SEC et SEC-MALS, l'absence d'oligomère de masse moléculaire proche de 70 kDa (pontage au GTA : 72 kDa) peut s'expliquer par un manque de résolution de la colonne utilisée. Si la réaction de pontage covalent au BS^3 est incomplète, cet oligomère de haut poids moléculaire n'a sans doute pas pu être ponté dans des quantités suffisantes pour être détecté. À l'inverse, une quantité excessive de GTA par rapport à la proportion d'oligomères à ponter, a pu provoquer des pontages entre les complexes CTD IN PERV-CTD Brd2 dimériques. Ce phénomène peut engendrer la formation de l'oligomère observé (72 kDa).

Au final, les résultats obtenus ne nous ont pas permis de définir clairement l'état oligomérique du complexe IN_{CTD}-$Brd2_{CTD}$ en solution. En revanche, hormis les résultats de pontage au GTA, l'ensemble des données met en avant la monodispersité de la solution protéique d'intérêt, requise pour l'étude structurale du complexe.

2. Etude structurale du complexe IN_{CTD}-$Brd2_{CTD}$

1. Dichroïsme circulaire

Nous avons utilisé le dichroïsme circulaire pour mesurer le pourcentage de chaque élément de structure secondaire. Le spectre présente un maximum à une longueur d'onde proche de 190 nm et de deux minimas à 205 nm et 220 nm (Figure 65). Ce profil est caractéristique des protéines constituées majoritairement d'hélices α.

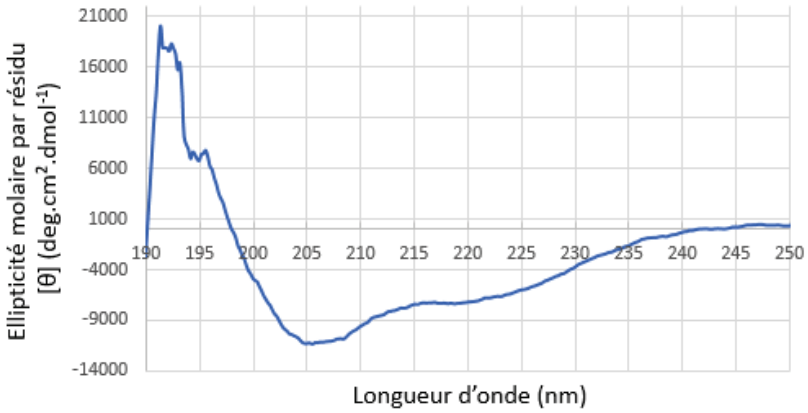


Figure 65. Spectre de dichroïsme circulaire du complexe IN_{CTD}-$Brd2_{CTD}$, représentant l'ellipticité molaire par résidu [θ] en fonction d'une gamme de longueurs d'onde dans l'UV lointain.

Les pourcentages d'hélices α, de brins β et de structures apériodiques, calculés avec l'algorithme mathématique SELCON3, sont respectivement de 29%, 23% et 48%. Ces résultats semblent cohérents avec les données de la littérature. En effet, les domaines CTD de l'IN de MLV et ET de Brd4, homologues au CTD de l'IN de PERV-A/C (312-408) et au domaine ET (676-738) contenu dans le CTD de Brd2 étudié (666-811), ont respectivement un repliement tout-β (Kvaratskhelia *et al.*, 2014) et un repliement tout-α (Crowe *et al.*, 2016) dans leurs structures RMN.

La région 739-811 de structure inconnue est prédite comme ne comportant que des hélices α d'après le serveur Psi-Pred (McGuffin *et al.*, 2000). Enfin, une structuration en deux brins β du motif EBM du CTD de l'IN de PERV-A/C est attendue lors de la formation du complexe (Crowe *et al.*, 2016). L'ensemble de ces données structurales expérimentales et prédites conduirait à des pourcentages respectifs d'hélices, de brins et de structures apériodiques pour le complexe de 27%, 15% et 58%, proches de ceux calculés par SELCON3.

2. Essais de cristallisation

Les essais ont été réalisés par la méthode de diffusion de vapeur en gouttes assise et suspendue. Près de 1 500 conditions de cristallisation ont été testées avec les kits commerciaux

PEG/Ion, PEG/Ion2, CrystalScreen (*Hampton Research*), $AmSO_4$Suite, MbClass, MPD Suite, JCSG+, PEGs II Suite (*Qiagen*), ProPlex et Morpheus (*Molecular Dimensions*). Quatre conditions ont conduit à la formation de cristaux (Tableau 7).

Kit commercial	Condition de cristallisation testée	Concentration IN_{CTD}-$Brd2_{CTD}$	Solution tampon	Rapport volumique agent précipitant/ protéine
PEGs II Suite (*Qiagen*)	0,2 M $CaCl_2$, 0,1 M Tris pH 8,5, 25% PEG 4 000 (w/v)	7 mg/ml	20 mM Tris HCl pH 8, 500 mM NaCl	1/1
	0,2 M $CaCl_2$, 0,1 M Tris pH 8,5, 20% PEG 4 000 (w/v)			
	0,1 M tri-sodium citrate, 20% PEG 4 000 (w/v), 5% isopropanol (w/v)	13 mg/ml		
$AmSO_4$Suite (*Qiagen*)	3 M $AmSO_4$, 10% glycérol (w/v)	7 mg/ml		

Tableau 7. Récapitulatif des conditions de cristallisation testées à 19°C en présence du complexe IN_{CTD}-$Brd2_{CTD}$, pour lesquelles des cristaux sont apparus.

Le pouvoir de diffraction de ces cristaux n'a pas pu être testé car ils ont fondu lorsque nous les avons pêchés ou déposés dans la solution de cryo-protection (10% éthylène glycol (v/v)) (Figure 66).

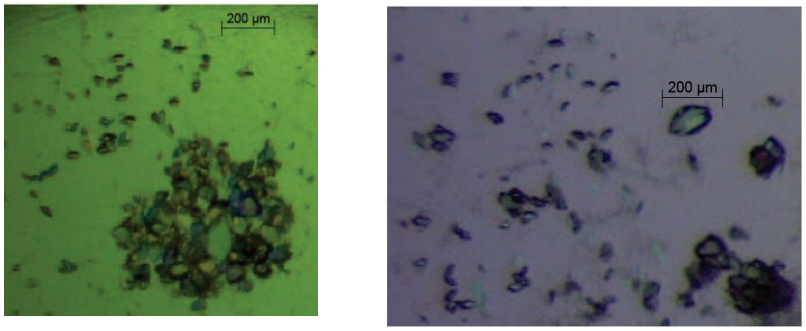


Figure 66. Cristaux obtenus avec une solution IN_{CTD}-$Brd2_{CTD}$ à 7 mg/ml et les conditions PEGII Suite (gauche) et $AmSO_4$ Suite (droite) décrites dans le tableau 7.

3. Diffusion des rayons X aux petits angles (SAXS)

Des mesures de SAXS ont été réalisées sur la ligne de lumière BM29 de l'ESRF, avec une solution de complexe IN_{CTD}-$Brd2_{CTD}$ concentrée à 6 mg/ml, en utilisant un passeur automatique d'échantillons.

La courbe de diffusion obtenue a été normalisée, après soustraction de la courbe de diffusion du tampon protéique seul (Figure 67A). La valeur de R_g calculée est de 3,1 nm. La masse moléculaire estimée est de 32 kDa, valeur proche de la masse moléculaire théorique d'un hétérodimère de complexe IN_{CTD}-$Brd2_{CTD}$ de 30 kDa. Ces résultats sont en accord avec les données précédentes obtenues par SEC, SEC-MALS et pontage covalent, pour lesquelles le complexe est dimérique en solution.

La fonction de distribution des distances P(r), calculée à partir du logiciel GNOM de la suite ATSAS, Petoukhov *et al.*, 2012), donne une valeur de D_{max} de 11 nm et confirme la valeur de R_g de 3,1 nm (Figure 67B). Cinquante modèles indépendants ont été calculés *ab initio* par le logiciel DAMMIN et superposés. L'écart-type normalisé NSD (*Normalized Spatial Discrepancy*) de 0,7 traduit l'homogénéité des enveloppes générées (Kozin *et al.*, 2001). Après moyennation, l'enveloppe SAXS résultante apparaît étendue et comprend deux parties reliées entre elles, pouvant correspondre aux CTD de Brd2 et de l'IN de PERV-A/C (Figure 67C).

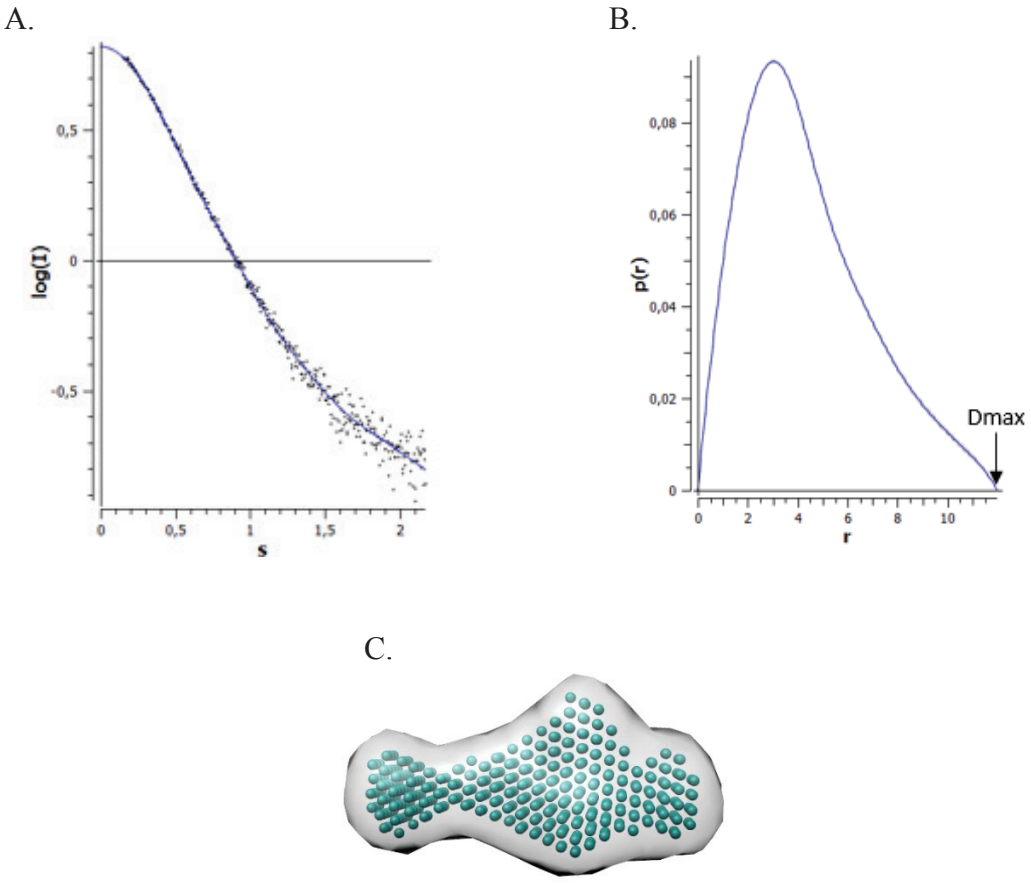


Figure 67. A. Courbe de diffusion des rayons X aux petits angles du complexe IN_{CTD}-$Brd2_{CTD}$ à 6 mg/ml : logarithme de l'intensité des rayons X diffusés I en fonction du vecteur de diffusion S (en $Å^{-1}$). B. Fonction de distribution des distances P(r) permettant de déterminer les valeurs de rayon de giration R_g et de distance interatomique maximale D_{max}. C. Modèle *ab initio* du complexe IN_{CTD}-$Brd2_{CTD}$ (6 mg/ml).

Un modèle du CTD de Brd2 réduit au domaine ET (homologue structural connu contrairement à la région 739-811 en aval) a été construit, mais la résolution de l'enveloppe SAXS n'est pas suffisante pour positionner à l'intérieur ce modèle et notre modèle du CTD de l'IN de PERV A/C.

Des enregistrements complémentaires ont été effectués à 3 et 6 mg/ml en système FPLC-SAXS (couplage entre SEC sur système FPLC et SAXS), à 3 mg/ml avec un passeur d'échantillons automatique et des résultats similaires ont été obtenus (même enveloppe SAXS avec Rg = 3,1±1 nm et D_{max} = 11±0,5 nm).

Conclusion et perspectives

La première partie de mon travail de thèse avait pour objectif de stabiliser la « nouvelle » interface dimérique du CCD de l'IN avec une nouvelle série de molécules synthétisées par l'équipe du Dr Maurin (ENS Lyon). Ces molécules ont également pour but de bloquer l'activité catalytique de l'intégrase, en agissant comme modulateurs d'oligomérisation.

Les premiers tests d'interaction protéine-ligand devaient être réalisés sur les CCD sauvage et H103C de l'IN de RAV-1. J'ai donc tenté de surproduire et purifier ces deux domaines, mais les protocoles fonctionnels auparavant mis au point n'ont pas permis d'obtenir de protéines solubles. Des problèmes de pureté d'eau et de contamination bactérienne ont été observés au même moment au sein de l'IBCP. Ceux-ci ont nécessité plusieurs mois pour être résolus. Par la suite, un phénomène d'agrégation subsistait, j'ai donc réalisé une nouvelle phase de mise au point des deux CCD, sans succès. En parallèle, nous avons essayé d'optimiser les protocoles de surproduction et de purification des CCD F185K et Y99C/F185K, pour tenter d'observer le « nouvel » assemblage dans l'IN de VIH-1 et tester les molécules candidates, mais un phénomène d'agrégation persistait.

Pour tenter d'obtenir des CCD solubles, il serait judicieux de tester d'autres tampons de purification que le Tris et/ou l'HEPES, comme par exemple des tampons phosphate ou MES à différents pH. Dans le même but, une autre approche consisterait à réaliser un criblage de tampons qui composent les conditions des kits de cristallisation actuellement disponibles (Collins *et al.*, 2005, Jancarik *et al.*, 2004). Ainsi, les gouttes seraient réalisées comme décrit dans ce manuscrit (*cf.* § IV.1, Matériels et méthodes) mais avec une solution protéique précipitée. Si des gouttes claires sont observées pour un tampon donné (disparition du précipité sous l'effet du tampon), celui-ci servira de tampon de stockage lors de la purification et de nouveaux essais de cristallisation seront ensuite envisagés. Enfin, si l'ensemble de ces tests n'aboutit pas à l'obtention de CCD solubles, nous pourrions tenter d'optimiser les constructions plasmidiques, en modifiant par exemple la position de la séquence codant l'étiquette de purification (en aval du gène codant la protéine d'intérêt). Les CCD solubles seront utilisés pour les tests d'interaction avec les molécules candidates. Celles qui se fixeront serviront aux essais de co-cristallisation et de trempage. La résolution des structures cristallines en complexe avec ces molécules pourrait permettre de valider nos hypothèses à l'échelle atomique, à savoir l'existence du « nouvel » assemblage dimérique sur le CCD de l'IN et son verrouillage par les molécules préalablement sélectionnées.

À défaut d'obtenir des CCD solubles, nous avons focalisé nos efforts sur les étapes de surproduction, de purification et de clivage de l'étiquette de purification de l'IN entière de RAV-1. Ainsi, une protéine soluble, pure avec un bon rendement de purification a pu être obtenue. En 2013, la cristallisation de l'IN entière de RSV, qui partage 95% d'identité de séquence avec l'IN de RAV-1, avait permis de résoudre une structure à deux domaines CCD+CTD (code PDB 4FW2, Shi *et al.*, 2013). Le domaine NTD n'avait pas pu être observé dans les cartes de densité électronique. Nous avons donc initié des essais de cristallisation pour tenter de résoudre la structure de l'IN entière de RAV-1, en espérant trouver de nouvelles conditions d'assemblage cristallin stabilisant ses trois domaines. Seuls des cristaux de sel ont été obtenus à ce jour.

Par ailleurs, la structure cristalline de l'IN de VIH-1 à deux domaines CCD+CTD (NTD non observé dans les cartes de densité électronique) a récemment été résolue en complexe avec une molécule inhibitrice de type ALLINI (*ALLosteric Integrase Inhibitors*, code PDB 5HOT, Gupta *et al*., 2016). Cette famille de molécules avait initialement été conçue pour bloquer l'étape d'intégration, en empêchant l'interaction entre l'IN de VIH-1 et son cofacteur cellulaire LEDGF/p75. Il est apparu, depuis, que ces molécules bloquaient également le cycle réplicatif rétroviral durant les étapes post-intégratives d'assemblage et de maturation (Le Rouzic *et al*., 2013 ; Balakrishnan *et al*., 2013). La structure cristalline du complexe révèle l'induction par les ALLINI d'une multimérisation de l'IN sous une forme inactive pour l'intégration, *via* des contacts entre les CCD et CTD de dimères adjacents. Des mutants de résistance aux ALLINI sont compatibles avec l'interface de multimérisation observée (Gupta *et al*., 2016). Ce résultat montre qu'une approche rationnelle de recherche d'inhibiteurs modulant l'oligomérisation de l'IN est viable, mais peut conduire à des effets inattendus. Il démontre également que la détermination de la structure cristalline d'un complexe IN entière-modulateur d'oligomérisation est envisageable.

Les molécules candidates ont d'abord été testées par des expériences d'intégration concertée *in vitro*, sans marquage radioactif. Celles pour lesquelles une diminution de la quantité de produits d'intégration a été observée, seront testées selon la même méthode, avec un ADN substrat radioactivement marqué et une concentration en composés plus importante. Ainsi, nous pourrons identifier des molécules qui inhibent la réaction d'intégration *in vitro*. Pour vérifier et quantifier leur pouvoir inhibiteur *in cellulo* (tests de cytotoxicité et IC_{50}), nous réaliserons des analyses sur la réplication du virus RAV-1.

Parallèlement à ces travaux, nous souhaitons avoir une meilleure compréhension du mécanisme d'intégration pour pouvoir développer de nouvelles stratégies antirétrovirales. Pour ce faire, j'ai participé en fin de thèse aux premières étapes de mise au point d'un protocole expérimental permettant, pour la première fois, de suivre en temps réel la réaction d'intégration concertée à l'échelle de la molécule unique. Cette expérience innovante, développée dans l'équipe par le Dr Fiorini, implique l'utilisation d'un microscope optique couplé à des pinces magnétiques servant à la micromanipulation de l'ADN (Figure 68).

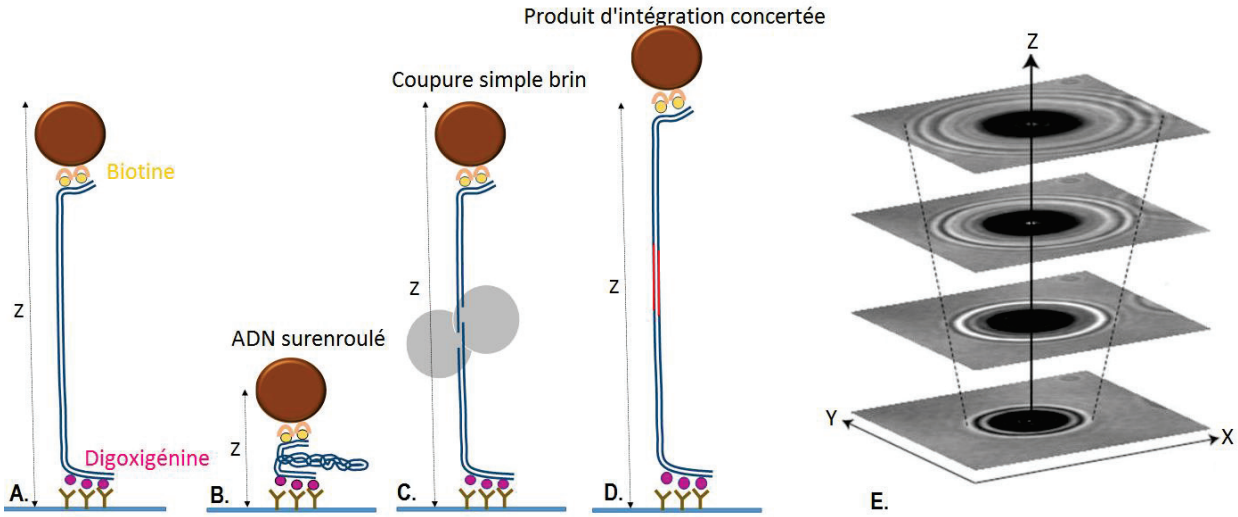


Figure 68. Étude de la réaction d'intégration concertée en molécule unique par pinces magnétiques. A. L'ADN accepteur (en bleu) est attaché au support (interaction anticorps-digoxigénine en rose) et à une bille magnétique (interaction biotine en jaune-streptavidine). B. Les mouvements de rotation et de translation des aimants mettent sous tension cet ADN qui devient surenroulé. C. L'IN (en gris) catalyse la coupure simple brin lors du transfert de brins, ce qui relâche le surenroulement de l'ADN accepteur. D. Le produit d'intégration final, contenant l'ADN viral (en rouge), a une longueur Z plus grande que celle de l'ADN initial. E. L'extension de la molécule d'ADN est suivie en temps réel, à travers la position de la bille magnétique mesurée par le microscope optique.

Cette nouvelle approche permettra de mettre en évidence des intermédiaires réactionnels difficiles à caractériser par les techniques préexistantes. Elle servira également à évaluer le pouvoir potentiellement inhibiteur des composés synthétisés sur l'activité d'intégration de l'IN de RAV-1 mais aussi de VIH-1.

La seconde partie de mes travaux de recherche a porté sur les études biochimique, biophysique et structurale de l'IN de PERV-A/C et de son cofacteur cellulaire humain Brd2.

La modélisation 3D de l'intasome de ce *Gammaretrovirus* en complexe avec le raltégravir a permis d'identifier les résidus impliqués dans cette interaction. La sensibilité de l'IN de PERV-A/C au raltégravir peut s'expliquer à travers la conservation de ces résidus dans les séquences des IN de PFV et de VIH-1 (Demange *et al*., 2015).

Afin d'étudier l'interaction entre les CTD de l'IN de PERV-A/C et de Brd2, j'ai mis au point leurs protocoles de surproduction et de purification. Les partenaires isolés purs ont ensuite été mis en contact pour former le complexe protéique d'intérêt mais, malgré les tentatives d'optimisation, l'apparition d'agrégats n'a pas permis de caractériser le complexe en solution.

Face au phénomène d'agrégation observé, une alternative a été trouvée à travers la stratégie « multiTEV » : à partir d'un plasmide unique, une polyprotéine contenant la protéase du TEV et des sites TEV a été surproduite. Ainsi la polyprotéine peut être « auto-clivée » et libérer les différentes sous-unités pour permettre l'assemblage du complexe IN_{CTD}-$Brd2_{CTD}$ directement dans les bactéries d'expression (Levy *et al.*, 2016). La présence d'étiquettes de purification distinctes sur chacun des CTD a permis de purifier le complexe d'intérêt. Les techniques biochimiques et biophysiques employées ont montré que l'échantillon était monodisperse et hétérodimérique en solution à 6 mg/ml avec une stœchiométrie 1:1. Les mesures par dichroïsme circulaire montrent que le complexe comprend 29% d'hélices et 23% de brins, ce qui est cohérent avec les résultats de modélisation. Des essais de cristallisation ont été réalisés, des cristaux ont été obtenus mais ils ont fondu lorsque nous les avons manipulés. Toutefois, des données structurales à basse résolution ont été obtenues par SAXS. Elles ont permis de calculer une enveloppe du complexe IN_{CTD}-$Brd2_{CTD}$. Cette enveloppe, de forme allongée, n'est pas assez résolutive pour placer sans ambiguïté notre modèle du CTD de l'IN de PERV-A/C et, *a fortiori*, celui du CTD de Brd2 qui contient une région sans homologue structural connu. La validation de nos résultats biochimiques et biophysiques par des expériences *in vitro* et *in cellulo* est en cours dans les équipes des Drs Ronfort (INRA Lyon) et Blanchard (ANSES Ploufagran). L'ensemble de ces expériences seront prochainement soumis à publication.

Pour tenter d'obtenir des résultats structuraux sur la formation du complexe, il pourrait être intéressant d'utiliser notre stratégie « multiTEV » en se limitant à certaines régions, à savoir le CTD de l'IN de PERV-A/C et le domaine ET de Brd2 qui forment un complexe stable (données non publiées). Ceci permettrait de positionner précisément le CTD de l'IN par rapport au domaine ET, ce qui n'est pas possible en utilisant comme guide la structure RMN du complexe minimal EBM IN MLV-ET Brd4 (Crowe *et al*., 2016). En effet, cette structure RMN comprend une extrémité carboxy-terminale déstructurée au niveau du CTD de l'IN de MLV et ne permet pas d'orienter avec précision un partenaire par rapport à l'autre.

À plus long terme, la formation du complexe pourra être étendue aux protéines entières et permettra de mieux guider la conception d'inhibiteurs d'interaction protéine-protéine par des

approches *in silico* (*docking*), *in vitro* (tests d'intégration concertée), *in cristallo* (co-cristallisation et trempage) et *in cellulo* (tests de cytotoxicité et IC_{50}).

De plus, l'étude par *histone array*, réalisée avec le CTD de l'IN de PERV-A/C, a révélé un profil de spécificité intéressant pour des queues d'histones H2B et H3 portant des modifications post-traductionnelles. En effet, les interactions détectées impliquent des queues d'histones H2B modifiées (H2BK5ac et H2BK12ac) spécifiques de sites d'initiation de la transcription, connus pour être ciblés par les *Gammaretrovirus* pour une intégration efficace. Ces informations ont également permis d'émettre une hypothèse selon laquelle les queues d'histones modifiées H2B activeraient la transcription du génome proviral. Celle-ci serait ralentie par l'interaction IN-queues d'histones H3 modifiées (H3K27me2, -me2s et -me3) au cours de l'étape d'intégration. Il apparaît essentiel de poursuivre cette étude avec le complexe IN_{CTD}-$Brd2_{CTD}$ que nous avons à disposition, pour connaître l'influence du cofacteur cellulaire humain Brd2 sur l'interaction de l'IN avec les histones. À plus long terme, les mêmes travaux pourront être menés en présence du complexe formé par les protéines entières. Ceci permettra d'identifier les mécanismes et les inhibiteurs permettant de mieux contrôler les risques de xénozoonose lors d'une xénotransplantation.

Publication

Journal of General Virology (2015), **96**, 3124–3130 DOI 10.1099/jgv.0.000236

Short Communication

Porcine endogenous retrovirus-A/C: biochemical properties of its integrase and susceptibility to raltegravir

Antonin Demange,[1]† Halima Yajjou-Hamalian,[2,3,4,5]† Kathy Gallay,[2,3,4]† Catherine Luengo,[2,3,4]† Véronique Beven,[1] Aurélie Leroux,[1] Marie-Pierre Confort,[2,3,4] Elsy Al Andary,[1,2,3,4] Patrice Gouet,[3,4,5] Karen Moreau,[2,3,4]* Corinne Ronfort[2,3,4]†† and Yannick Blanchard[1]††

[1]ANSES, Ploufragan/Plouzané Laboratory, Viral Genetics and Bio-Security Unit, Université Européenne de Bretagne, Ploufragan, France

[2]INRA, Université Lyon 1, UMR754, Rétrovirus et Pathologie Comparée, 69007 Lyon, France

[3]Université de Lyon, 69000 Lyon, France

[4]UMS3444 BioSciences Gerland Lyon Sud, 69007 Lyon, France

[5]Institut de Biologie et Chimie des Protéines, BMSSI-IBCP, UMR 5086 CNRS Université Lyon 1, 7, passage du Vercors, 69367 Lyon Cedex 07, France

Correspondence
Corinne Ronfort
ronfort@univ-lyon1.fr
Yannick Blanchard
yannick.blanchard@anses.fr

Received 16 April 2015
Accepted 3 June 2015

Porcine endogenous retroviruses (PERVs) are present in the genomes of pig cells. The PERV-A/C recombinant virus can infect human cells and is a major risk of zoonotic disease in the case of xenotransplantation of pig organs to humans. Raltegravir (RAL) is a viral integrase (IN) inhibitor used in highly active antiretroviral treatment. In the present study, we explored the potential use of RAL against PERV-A/C. We report (i) a three-dimensional model of the PERV-A/C intasome complexed with RAL, (ii) the sensitivity of PERV-A/C IN to RAL *in vitro* and (iii) the sensitivity of a PERV-A/C-IRES-GFP recombinant virus to RAL *in cellulo*. We demonstrated that RAL is a potent inhibitor against PERV-A/C IN and PERV-A/C replication with IC_{50}s in the nanomolar range. To date, the use of retroviral inhibitors remains the only way to control the risk of zoonotic PERV infection during pig-to-human xenotransplantation.

Porcine Endogenous retrovirus (PERV) is a member of the genus *Gammaretrovirus* and comprises three subtypes: PERV-A, -B and -C. The PERV-A/C recombinant virus has been shown to infect human cells *in vitro* (Patience *et al.*, 1997). Thus, the risk of zoonotic transmission has raised serious biosafety concerns about the use of pig organs in human xenotransplantation programmes (Kimsa *et al.*, 2014).

In the latest release of the pig genome (*Sus scrofa* version 10.2; GCA_000003025.4), a thorough full analysis of the PERV integration loci revealed the absence of intact full-length sequences and a considerable polymorphism in PERV loci between different pigs (Groenen *et al.*, 2012), making the genetic selection of PERV-free pigs unlikely. Furthermore, it was observed that through complementation and genetic recombination, two defective PERV genomes could give rise to an infectious virus such as the PERV-A/C recombinant virus (Bartosch *et al.*, 2004; Martin *et al.*, 2006).

Finding drugs that could be used against PERV replication is therefore highly desirable. Several drugs have been developed against human immunodeficiency virus type 1 (HIV-1) and susceptibility of PERV was demonstrated *in vitro* to the triphosphorylated nucleoside analogue of zidovudine (AZT). However, no inhibition of PERV was observed with several other drugs directed against HIV-1 (Qari *et al.*, 2001; Stephan *et al.*, 2001; Wilhelm *et al.*, 2002).

A critical step for retroviral replication is the stable integration of the provirus into the host genome. The viral enzyme integrase (IN) has a key role in this event (Engelman, 2010). Previously, we characterized the integration profile of PERV-A/C in human cells and

*Present adress: CIRI, International Center for Infectiology Research, INSERM U1111, Université Lyon 1, Ecole Normale Supérieure de Lyon, CNRS UMR5308, Lyon, France.

†These authors contributed equally to this work.

††These authors contributed equally to this work.

One supplementary table and two supplementary figures are available with the online Supplementary Material.

Downloaded from www.microbiologyresearch.org by
IP: 193.54.110.32
On: Mon, 11 Jul 2016 09:47:46
000236 © 2015 The Authors *Printed in Great Britain*

demonstrated that the integration profile of PERV-A/C showed an affinity for transcription start sites and CpG islands, and displayed common patterns of target site preference with both murine leukemia virus (MLV), another gammaretrovirus (Moalic *et al.*, 2006, 2009), and the prototype foamy virus (PFV), a spumavirus (Mitchell *et al.*, 2004; Nowrouzi *et al.*, 2006; Trobridge *et al.*, 2006). In human gene therapy protocols, the use of MLV-based vectors was associated with a high mutagenic insertion risk (Hacein-Bey-Abina *et al.*, 2003). PERV, which is otherwise insensitive to restriction by TRIM5α molecules (Wood *et al.*, 2009), is probably associated with the same risk.

Raltegravir (RAL) was the first IN inhibitor to be approved in 2007 for the treatment of HIV-1 infections. Elvitegravir and dolutegravir were approved recently. These compounds act as IN strand transfer (ST) inhibitors (INSTIs) (Serrao *et al.*, 2009) and display varying efficiency depending on the retrovirus (Koh *et al.*, 2011).

Here, we built a three-dimensional model of the PERV-A/C intasome (IN with viral DNA) in complex with RAL. We explored the potential use of RAL against PERV-A/C IN *in vitro* and against a PERV-A/C replication-competent virus *in cellulo*.

INs are modular proteins composed of three functional domains: the N-terminal domain (NTD), the catalytic core domain (CCD) and the C-terminal domain (CTD) (Fig. 1a). The individual structures of these domains have been characterized for many retroviruses (Jaskolski *et al.*, 2009), but not for PERV. Within the gammaretroviruses, PERV-A/C IN is closely related to MLV IN, sharing 50 % sequence identity in total, with 22 % for the NTD, 72 % for the CCD and 50 % for the CTD. PERV-A/C IN also shares a significant 26 % sequence identity with PFV IN, whose full-length structure is known (Hare *et al.*, 2010a; Maertens *et al.*, 2010). Sequence identity is lower with the HIV-1 and alpharetrovirus Rous-associated virus type 1 (RAV-1) INs at 17 and 18 % identity, respectively (Fig. 1b).

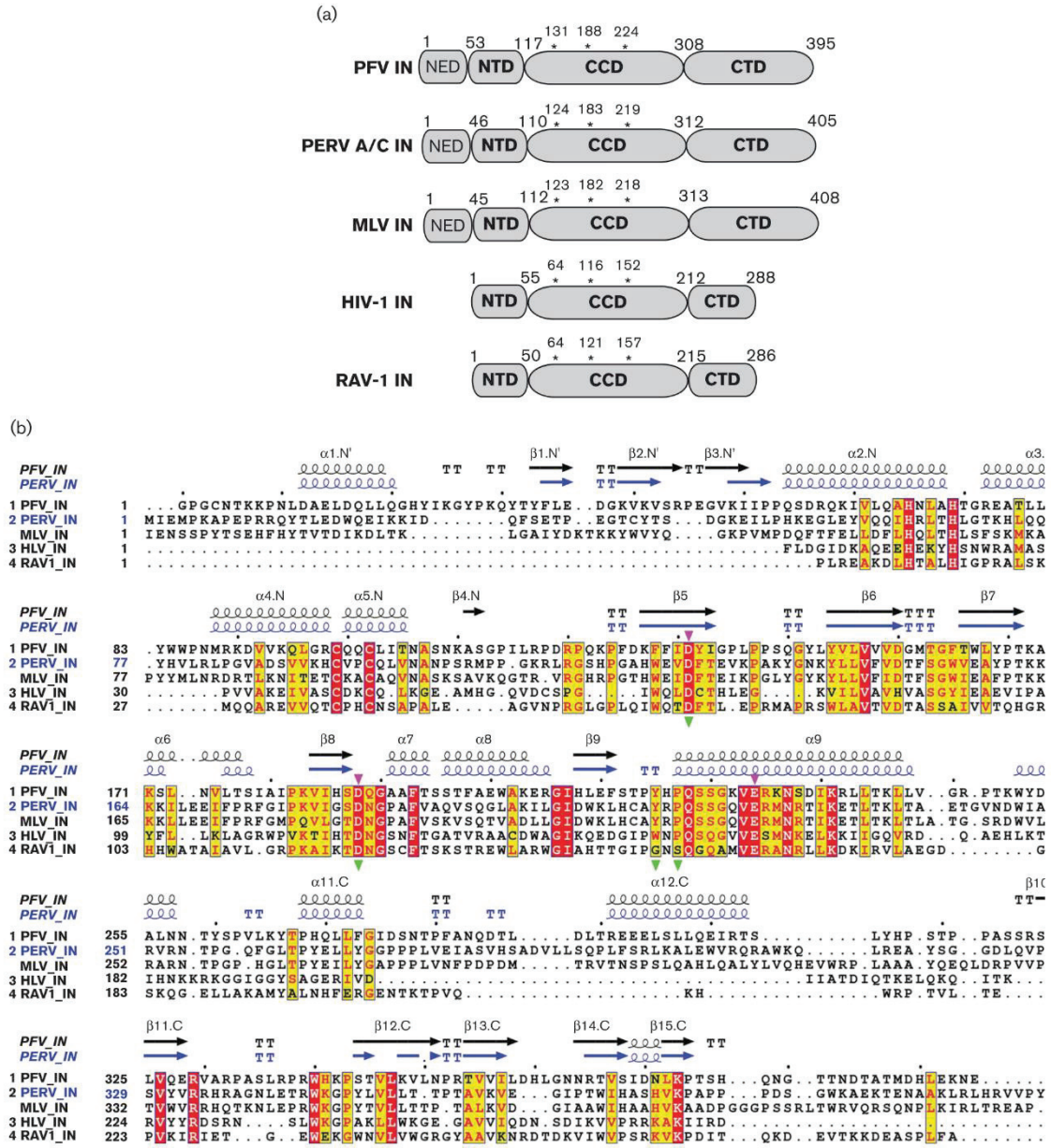
An initial model of PERV-A/C IN was obtained with the homology modelling server SWISS-MODEL (Arnold *et al.*, 2006). The structures of the NTD and CTD of MLV IN are known [Protein Data Bank (PDB) IDs: 4NZG and 2M9U (Kvaratskhelia *et al.*, 2014)] and were used to model those of PERV-A/C IN. The predicted structure of PERV-A/C IN comprises a singular 46 residue N-terminal extension domain (NED) with three β-strands, adjacent to the three α-helices bundle domain observed in all INs. This so-called NED + NTD structure is also observed in PFV IN. The CTD adopts an SH3 fold and terminates with a flexible region. The structure of the CCD of PFV IN with its highly conserved RNase H-like fold (PDB ID: 2X6N) (Réty *et al.*, 2010) was used in turn to model that of PERV-A/C. The tertiary structure of the full-length PFV IN (PDB ID: 3L2R) (Hare *et al.*, 2010a) was used to position the three domains of PERV-A/C IN with respect to each other. As a consequence, the CTD of PERV-A/C IN is positioned between the NED + NTD and the CCD in the initial model, which was further energy minimized with the program REFMAC5 (Vagin *et al.*, 2004) (for structure quality assessment, see Table S1, available in the online Supplementary Material). The resulting PERV-A/C IN model can be superimposed with the crystal structure of PFV IN (Fig. 1c) with a root-mean-square deviation of 2.1 Å on 324 Cα pairs. A model of the full-length HIV-1 IN was previously proposed by using a similar method (Krishnan *et al.*, 2010).

The crystal structure of the PFV intasome complexed with RAL (PDB ID: 3OYA) (Hare *et al.*, 2010b) was used to position bound RAL and viral DNA in our PERV-A/C IN model, after optimal superposition with the graphics program Coot (Emsley *et al.*, 2010). The resulting model of the PERV-A/C IN : RAL intasome was energy minimized with REFMAC5 (Table S1, Fig. 1c). The RAL-binding site of PFV is composed of residues Asp131, Asp188, Tyr215 and Pro217. These four residues are strictly conserved in PERV-A/C IN (corresponding to Asp124, Asp183, Tyr210 and Pro212) and well conserved in HIV-1 IN (Fig. 1b). Thus, PERV-A/C IN is expected to bind RAL, and a similar sensitivity to INSTIs is expected in spumaviruses, gammaretroviruses and lentiviruses. However, PFV Tyr215 and Pro217 are substituted by Gly and Ser in alpharetroviruses (Fig. 1a), giving rise to a significant resistance of Rous sarcoma virus IN against elvitegravir (Koh *et al.*, 2011).

We tested the sensitivity of PERV-A/C IN to RAL *in vitro*. We cloned and purified a recombinant PERV-A/C IN as we previously did for RAV-1 (Moreau *et al.*, 2002, 2003, 2004, 2009; Charmetant *et al.*, 2011) and HIV-1 INs (Cellier *et al.*, 2013; Trabaud *et al.*, 2015). The PERV IN coding sequences were subcloned into the pET14b plasmid and expressed as an N-terminal His_6-tagged protein in *Escherichia coli*. Purification was performed according to protocols previously developed for MLV (Jonsson *et al.*, 1993; Yang *et al.*, 1999). IN proteins expressed in bacteria were routinely purified to >90 % purity (Fig. S1). PERV-A/C IN activities (3′ processing and ST) were tested *in vitro* (Fig. S2) and IN was further assayed for its sensitivity to the ST inhibitor RAL (supplied by the National Institutes of Health AIDS Research and Reference Reagent Program). The IC_{50} was determined in a dose–response assay. We tested the drug from 1 nM to 10 μM and reported the ST activity as a percentage of the activity obtained in the absence of the drug (between 1 and 400 nM in Fig. 2). ST activity was almost completely inhibited at concentrations >400 nM and RAL inhibited the ST activity of the purified PERV-A/C IN *in vitro* with an IC_{50} of 40 nM. In parallel, RAL inhibited the ST activity of a purified HIV-1 IN (Cellier *et al.*, 2013) with an IC_{50} of 9 nM (data not shown), in accordance with other data (Bar-Magen *et al.*, 2009). These data showed that the RAL inhibitor can effectively inhibit PERV-A/C IN at concentrations that were similar to those needed for the inhibition of HIV-1 IN activity.

Finally, we constructed a replication-competent PERV-A/C virus encoding the GFP marker and used it to determine the susceptibility of PERV-A/C to RAL. A molecular

Downloaded from www.microbiologyresearch.org by
IP: 193.54.110.32
On: Mon, 11 Jul 2016 09:47:46



Downloaded from www.microbiologyresearch.org by
IP: 193.54.110.32
On: Mon, 11 Jul 2016 09:47:46

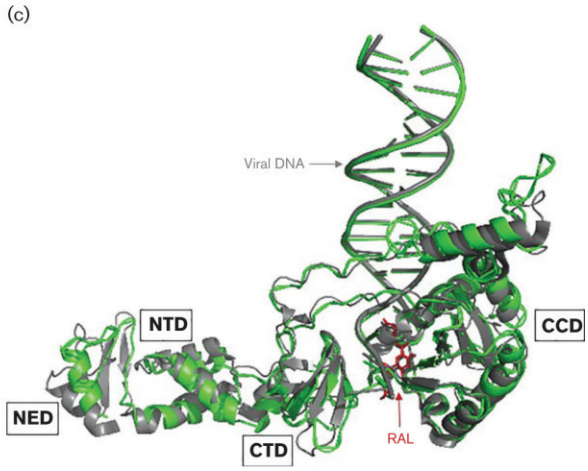


Fig. 1. Properties of the PERV-A/C IN protein. (a) Domain organization of IN proteins. The IN proteins are made of three domains: the NTD contains a zinc-binding motif and is involved in binding of viral DNA, the CCD contains the invariant catalytic triad D,D(35)E (indicated with asterisks), and the CTD is involved in non-specific DNA binding and IN oligomerization. Presented here are the spumavirus PFV IN, the gammaretroviruses PERV-A/C and MLV INs, the lentivirus HIV-1 IN, and the alpharetrovirus RAV-1 IN. PFV IN (as well as PERV-A/C and MLV INs) also contains the N-terminal extension domain (NED) which interacts with viral DNA in the PFV intasome structures. (b) Multiple sequence alignment of representative INs from (1) PFV, (2) PERV-A/C and MLV, (3) HIV-1, and (4) RAV-1. Strictly conserved residues have a red background; well-conserved residues are in red on a yellow background. Secondary structure elements are shown above sequence blocks for the crystal structure of the PFV intasome with RAL (PDB ID: 3OYA) and the SWISS-MODEL of PERV-A/C IN; the domains NED (residues 1–53), NTD (54–117) and CTD (309–395) of PFV IN are identified with N′, N and.C labels. Red triangles show the catalytic residues of the central CCD (residues 118–307); green triangles show residues in contact (distances <3.7 Å) with RAL in PFV. This alignment was calculated with CLUSTAL W and manually optimized after structural alignments with DALILITE (Holm *et al.*, 2006) between PFV IN and fragments of MLV, HIV-1 and RAV-1 INs available in the PDB (IDs: 4NZG, 2M9U, 1EX4, 1C0M and 3O4N). The figure was rendered with ESPript (Gouet *et al.*, 2003). (c) Structural superposition between the monomers of PERV-A/C IN and PFV IN with bound RAL and viral DNA. The polypeptide chains of the PERV model and the PFV crystal structure (PDB ID: 3OYA) are presented with green and grey ribbons, respectively. Viral DNA is shown as tubes (green, PERV; grey, PFV) and RAL as sticks (red, PERV; grey, PFV). The figure was created with PyMOL (DeLano, 2008).

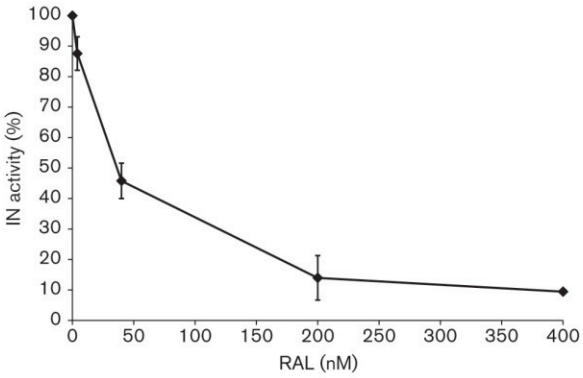


Fig. 2. Susceptibility of PERV-A/C IN to the RAL inhibitor *in vitro*. Dose–response inhibition of PERV-A/C IN ST activity by RAL (from 0 to 400 nM RAL). IN activity was determined as described in Fig. S2(b) and activities were quantified with ImageJ software (http://imagej.nih.gov/ij/). The curve represents the residual IN ST activity expressed as a percentage of the untreated assay (set as 100 %). Results are the mean ± SD of three experiments.

clone of PERV-A/C (AY570980) expressing the GFP marker (PERV-A/C-IRES-GFP) was constructed to monitor the infection process. The GFP sequence was inserted downstream of an internal ribosome entry site (IRES) sequence (Fig. 3a). Transfections, virus preparation and cell infections were performed as described previously (Moalic *et al.*, 2009).

The PERV-A/C-IRES-GFP plasmid [or an integration-defective mutant carrying a mutation in the catalytic site by substitution of D,D(35)E with D,A(35)E] was transfected into 293T cells. The GFP^+ cells were recorded by flow cytometry analysis 48 h after transfection. The percentage of GFP^+ cells was close to 20 % for both constructs. This percentage decreased to a fluorescence background level for the integration-defective PERV-A/C, whereas it remained stable for the next 2 weeks with the GFP integration- and replication-competent construction. Then, the percentage of GFP^+ cells increased progressively to reach 100 % GFP^+ cells after 1 month (Fig. 3b).

An infection experiment of HEK293 cells with the PERV-A/C-IRES-GFP was carried out with a range of RAL

Downloaded from www.microbiologyresearch.org by
IP: 193.54.110.32
On: Mon, 11 Jul 2016 09:47:46

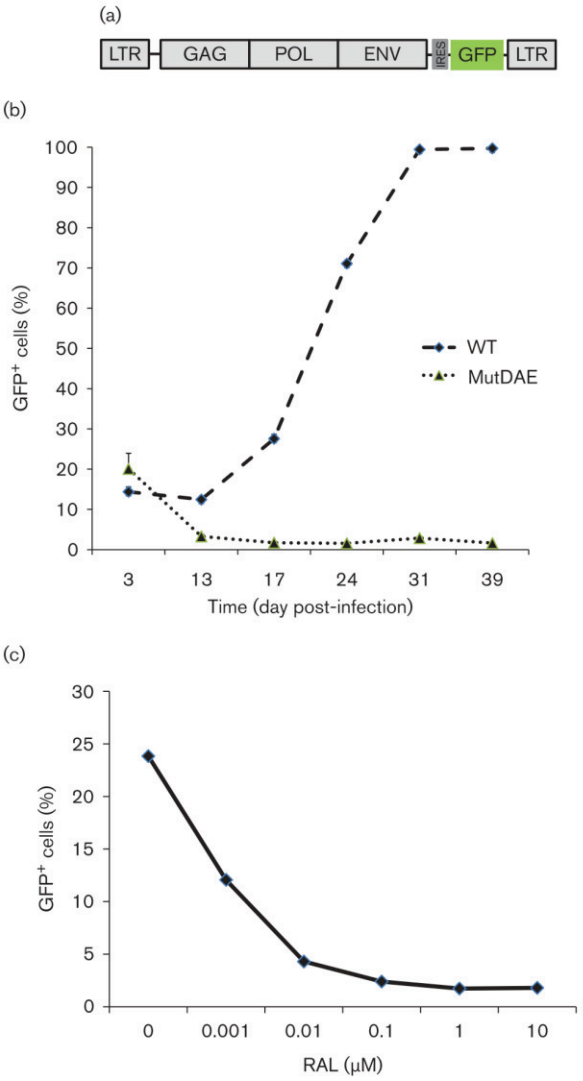


Fig. 3. Susceptibility of the PERV-A/C-IRES-GFP recombinant virus to the RAL IN inhibitor. (a) Map of the PERV-A/C-IRES-GFP recombinant virus. The IRES-EGFP cassette obtained from the pIRES2-EGFP vector (Clontech) was inserted in the PERV-A/C (GenBank accession number AY570980) backbone into the *Mlu*I and *Sal*I restriction sites downstream of the *env* gene. (b) Kinetics of PERV-A/C-IRES-GFP replication in 293T cells. 293T cells (plated at 2.5×10^5 cells per well) were transfected with 10 μg PERV-A/C-IRES-GFP plasmid [or an integration defective mutant (MutDAE)]. The GFP^+ cells were recorded by flow cytometry analysis at different times post-transfection. Results are the mean of three transfection experiments. (c) Susceptibility of PERV-A/C-IRES-GFP to the RAL IN inhibitor. HEK293 cells were plated and infected (m.o.i. 0.3) 24 h later with the PERV-A/C-IRES-GFP supernatant. At 24 h after infection, RAL (0.001–10 μM) was added. After 24 h, the cells were harvested and assessed by flow cytometry (FC500MPL; Beckman). The experiment was carried out in triplicate. Results are the mean values.

concentrations varying from 0.001 to 10 μM per well. At 48 h after infection, the percentage of GFP^+ cells displayed a twofold decrease with the first RAL concentration (0.001 μM) compared with the infected but RAL-untreated cells (Fig. 3c). This decrease of infected cells intensified with the range of drug and reached the concentration required to inhibit viral replication by 95 % at 0.1 μM. The EC_{50} could be estimated as 1 nM RAL, which is similar to the EC_{50} estimated for MLV (4.2 ± 0.1 nM) and HIV-1 (8.4 ± 2.8 nM) (Koh *et al.*, 2011). We also showed that the level of 2LTRc, which represents abortive integration events, was 7.7 times higher in cells treated with 1 μM RAL than in untreated but infected cells (data not shown). Thus, the activity profile of RAL on the replication of PERV-A/C-IRES-GFP appeared similar to that of MLV and HIV-1.

In conclusion, PERV-A/C IN is fully sensitive to the RAL inhibitor as shown by *in vitro* and *in cellulo* experiments. These data demonstrated that RAL is a potent inhibitor of PERV-A/C replication and could be useful for the treatment of PERV-A/C infections as component of a combination therapy with other drugs such as AZT. The new dolutegravir inhibitor that has recently shown more potent properties than RAL against HIV-1 (DeAnda *et al.*, 2013; Raffi *et al.*, 2013) would likely be an effective inhibitor against PERV infection.

Acknowledgements

We thank Y. Takeuchi (UCL, London, UK) for providing us with the molecular clone of PERV-A/C (AY570980). We are grateful to the 'Production and Analysis of Proteins (PAP)' platform of UMS3444 BioSciences Gerland Lyon Sud. A. D. was supported by a fellowship from Brittany and Conseil General des Côtes d'Armor; H. A. was supported by a doctorate fellowship from the Agence Nationale de Recherches sur le SIDA (ANRS). E. A. E. was supported by a doctorate fellowship from INRA and ANSES. The research leading to these results has received funding from the European Community Seventh Framework Programme (2007–2013), Research Infrastructures action, under grant agreement FP7-228394 (NADIR).

References

Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201.

Bar-Magen, T., Sloan, R. D., Faltenbacher, V. H., Donahue, D. A., Kuhl, B. D., Oliveira, M., Xu, H. & Wainberg, M. A. (2009). Comparative biochemical analysis of HIV-1 subtype B and C integrase enzymes. *Retrovirology* **6**, 103.

Bartosch, B., Stefanidis, D., Myers, R., Weiss, R., Patience, C. & Takeuchi, Y. (2004). Evidence and consequence of porcine endogenous retrovirus recombination. *J Virol* **78**, 13880–13890.

Cellier, C., Moreau, K., Gallay, K., Ballandras, A., Gouet, P. & Ronfort, C. (2013). *In vitro* functional analyses of the human immunodeficiency virus type 1 (HIV-1) integrase mutants give new insights into the intasome assembly. *Virology* **439**, 97–104.

Downloaded from www.microbiologyresearch.org by
IP: 193.54.110.32
On: Mon, 11 Jul 2016 09:47:46

Charmetant, J., Moreau, K., Gallay, K., Ballandras, A., Gouet, P. & Ronfort, C. (2011). Functional analyses of mutants of the central core domain of an Avian Sarcoma/Leukemia Virus integrase. *Virology* **421**, 42–50.

DeAnda, F., Hightower, K. E., Nolte, R. T., Hattori, K., Yoshinaga, T., Kawasuji, T. & Underwood, M. R. (2013). Dolutegravir interactions with HIV-1 integrase-DNA: structural rationale for drug resistance and dissociation kinetics. *PLoS One* **8**, e77448.

DeLano, W. L. (2008). *The PyMol Molecular Graphics Systems.* Palo Alto, CA: DeLano Scientific.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501.

Engelman, A. (2010). Reverse transcription and integration. In *Retroviruses: Molecular Biology, Genomics and Pathogenesis*, pp. 129–159. Edited by R. Kurth & N. Bannert. Norfolk: Caister Academic Press.

Gouet, P., Robert, X. & Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* **31**, 3320–3323.

Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., Rogel-Gaillard, C., Park, C., Milan, D. & other authors (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393–398.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M. P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C. S., Pawliuk, R. & other authors (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**, 415–419.

Hare, S., Gupta, S. S., Valkov, E., Engelman, A. & Cherepanov, P. (2010a). Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**, 232–236.

Hare, S., Vos, A. M., Clayton, R. F., Thuring, J. W., Cummings, M. D. & Cherepanov, P. (2010b). Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. *Proc Natl Acad Sci U S A* **107**, 20057–20062.

Holm, L., Kaariainen, S., Wilton, C. & Plewczynski, D. (2006). Using Dali for structural comparison of proteins. *Curr Protoc Bioinformatics* **14**, 5.5.1–5.5.24.

Jaskolski, M., Alexandratos, J. N., Bujacz, G. & Wlodawer, A. (2009). Piecing together the structure of retroviral integrase, an important target in AIDS therapy. *FEBS J* **276**, 2926–2946.

Jonsson, C. B., Donzella, G. A. & Roth, M. J. (1993). Characterization of the forward and reverse integration reactions of the Moloney murine leukemia virus integrase protein purified from *Escherichia coli*. *J Biol Chem* **268**, 1462–1469.

Kimsa, M. C., Strzalka-Mrozik, B., Kimsa, M. W., Gola, J., Nicholson, P., Lopata, K. & Mazurek, U. (2014). Porcine endogenous retroviruses in xenotransplantation – molecular aspects. *Viruses* **6**, 2062–2083.

Koh, Y., Matreyek, K. A. & Engelman, A. (2011). Differential sensitivities of retroviruses to integrase strand transfer inhibitors. *J Virol* **85**, 3677–3682.

Krishnan, L., Li, X., Naraharisetty, H. L., Hare, S., Cherepanov, P. & Engelman, A. (2010). Structure-based modeling of the functional HIV-1 intasome and its inhibition. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15910–15915.

Kvaratskhelia, M., Sharma, A., Larue, R. C., Serrao, E. & Engelman, A. (2014). Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Res* **42**, 10209–10225.

Maertens, G. N., Hare, S. & Cherepanov, P. (2010). The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326–329.

Martin, S. I., Wilkinson, R. & Fishman, J. A. (2006). Genomic presence of recombinant porcine endogenous retrovirus in transmitting miniature swine. *Virol J* **3**, 91.

Mitchell, R. S., Beitzel, B. F., Schroder, A. R., Shinn, P., Chen, H., Berry, C. C., Ecker, J. R. & Bushman, F. D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* **2**, E234.

Moalic, Y., Blanchard, Y., Félix, H. & Jestin, A. (2006). Porcine endogenous retrovirus integration sites in the human genome: features in common with those of murine leukemia virus. *J Virol* **80**, 10980–10988.

Moalic, Y., Félix, H., Takeuchi, Y., Jestin, A. & Blanchard, Y. (2009). Genome areas with high gene density and CpG island neighborhood strongly attract porcine endogenous retrovirus for integration and favor the formation of hot spots. *J Virol* **83**, 1920–1929.

Moreau, K., Faure, C., Verdier, G. & Ronfort, C. (2002). Analysis of conserved and non-conserved amino acids critical for ALSV (Avian leukemia and sarcoma viruses) integrase functions *in vitro*. *Arch Virol* **147**, 1761–1778.

Moreau, K., Faure, C., Violot, S., Verdier, G. & Ronfort, C. (2003). Mutations in the C-terminal domain of ALSV (Avian Leukemia and Sarcoma Viruses) integrase alter the concerted DNA integration process *in vitro*. *Eur J Biochem* **270**, 4426–4438.

Moreau, K., Faure, C., Violot, S., Gouet, P., Verdier, G. & Ronfort, C. (2004). Mutational analyses of the core domain of Avian Leukemia and Sarcoma Viruses integrase: critical residues for concerted integration and multimerization. *Virology* **318**, 566–581.

Moreau, K., Charmetant, J., Gallay, K., Faure, C., Verdier, G. & Ronfort, C. (2009). Avian sarcoma and leukemia virus (ASLV) integration *in vitro*: mutation or deletion of integrase (IN) recognition sequences does not prevent but only reduces the efficiency and accuracy of DNA integration. *Virology* **392**, 94–102.

Nowrouzi, A., Dittrich, M., Klanke, C., Heinkelein, M., Rammling, M., Dandekar, T., von Kalle, C. & Rethwilm, A. (2006). Genome-wide mapping of foamy virus vector integrations into a human cell line. *J Gen Virol* **87**, 1339–1347.

Patience, C., Takeuchi, Y. & Weiss, R. A. (1997). Infection of human cells by an endogenous retrovirus of pigs. *Nat Med* **3**, 282–286.

Qari, S. H., Magre, S., García-Lerma, J. G., Hussain, A. I., Takeuchi, Y., Patience, C., Weiss, R. A. & Heneine, W. (2001). Susceptibility of the porcine endogenous retrovirus to reverse transcriptase and protease inhibitors. *J Virol* **75**, 1048–1053.

Raffi, F., Jaeger, H., Quiros-Roldan, E., Albrecht, H., Belonosova, E., Gatell, J. M., Baril, J. G., Domingo, P., Brennan, C. & other authors (2013). Once-daily dolutegravir versus twice-daily raltegravir in antiretroviral-naive adults with HIV-1 infection (SPRING-2 study): 96 week results from a randomised, double-blind, non-inferiority trial. *Lancet Infect Dis* **13**, 927–935.

Réty, S., Reaeábková, L., Dubanchet, B., Silhán, J., Legrand, P. & Lewit-Bentley, A. (2010). Structural studies of the catalytic core of the primate foamy virus (PFV-1) integrase. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**, 881–886.

Serrao, E., Odde, S., Ramkumar, K. & Neamati, N. (2009). Raltegravir, elvitegravir, and metoogravir: the birth of "me-too" HIV-1 integrase inhibitors. *Retrovirology* **6**, 25.

Stephan, O., Schwendemann, J., Specke, V., Tacke, S. J., Boller, K. & Denner, J. (2001). Porcine endogenous retroviruses (PERVs): generation of specific antibodies, development of an immunoperoxidase assay (IPA) and inhibition by AZT. *Xenotransplantation* **8**, 310–316.

Trabaud, M. A., Cotte, L., Saison, J., Ramière, C., Ronfort, C., Venet, F., Tardy, J. C., Monneret, G. & André, P. (2015). Persistent production of an integrase-deleted HIV-1 variant with no resistance mutation

Downloaded from www.microbiologyresearch.org by
IP: 193.54.110.32
On: Mon, 11 Jul 2016 09:47:46

and wild-type proviral DNA in a treated patient. *AIDS Res Hum Retroviruses* **31**, 142–149.

Trobridge, G. D., Miller, D. G., Jacobs, M. A., Allen, J. M., Kiem, H. P., Kaul, R. & Russell, D. W. (2006). Foamy virus vector integration sites in normal human cells. *Proc Natl Acad Sci U S A* **103**, 1498–1503.

Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184–2195.

Wilhelm, M., Fishman, J. A., Pontikis, R., Aubertin, A. M. & Wilhelm, F. X. (2002). Susceptibility of recombinant porcine endogenous retrovirus reverse transcriptase to nucleoside and non-nucleoside inhibitors. *Cell Mol Life Sci* **59**, 2184–2190.

Wood, A., Webb, B. L., Bartosch, B., Schaller, T., Takeuchi, Y. & Towers, G. J. (2009). Porcine endogenous retroviruses PERV A and A/C recombinant are insensitive to a range of divergent mammalian TRIM5alpha proteins including human TRIM5alpha. *J Gen Virol* **90**, 702–709.

Yang, F., Leon, O., Greenfield, N. J. & Roth, M. J. (1999). Functional interactions of the HHCC domain of moloney murine leukemia virus integrase revealed by nonoverlapping complementation and zinc-dependent dimerization. *J Virol* **73**, 1809–1817.

Downloaded from www.microbiologyresearch.org by
IP: 193.54.110.32
On: Mon, 11 Jul 2016 09:47:46

Références bibliographiques

A

Agapkina J, Smolov M, Barbe S, Zubin E, Zatsepin T, Deprez E, Le Bret M, Mouscadet JF, Gottikh M. Probing of HIV-1 integrase/DNA interactions using novel analogs of viral DNA. J Biol Chem (2006) 281:11530-11540

Aiyar A, Hindmarsh P, Skalka AM, Leis J. Concerted integration of linear retroviral DNA by the avian sarcoma virus integrase in vitro: dependence on both long terminal repeat termini. J Virol (1996) 70:3571-3580

Aiyer S, Swapna GV, Malani N, Aramini JM, Schneider WM, Plumb MR, Ghanem M, Larue RC, Sharma A, Studamire B, Kvaratskhelia M, Bushman FD, Montelione GT, Roth MJ. Altering murine leukemia virus integration through disruption of the integrase and BET protein family interaction. Nucleic Acids Res (2014) 42:5917-5928

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol (1990) 215:403-410

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res (1997) 25:3389-3402

Andrake MD, Ramcharan J, Merkel G, Zhao XZ, Burke TR, Jr., Skalka AM. Comparison of metal-dependent catalysis by HIV-1 and ASV integrase proteins using a new and rapid, moderate throughput assay for joining activity in solution. AIDS Res Ther (2009) 6:14

Arakawa T, Tsumoto K. The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation. Biochem Biophys Res Commun (2003) 304:148-152

Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. Cold Spring Harb Perspect Med (2012) 2:a007161

Assil S, Bolze PA, Boukhali M, Cariou C, Chauveau L, Chuvin N, Dhondt K, Ducuing A, Dupont JB, Grandin C, Jarre G, Le Douce J, Lebrun D, Lechenet FX, Luther N, Milivojevic M, Peres E, Plantamura E, Sanlaville A, Schwob A, Seggio M, Serre JE, Thiebaut PA, Tirmarche S, Tshilenge KT, Vandamme C, Velrlhac R, Vinera J, Mahieux R, Journo C. Retrovirus humain XMRV : la fin d'une histoire seduisante ? Virologie (2011) 15:222-234

Azran I, Schavinsky-Khrapunsky Y, Priel E, Huleihel M, Aboud M. Implications of the evolution pattern of human T-cell leukemia retroviruses on their pathogenic virulence (Review). Int J Mol Med (2004) 14:909-915

B

Balada E, Vilardell-Tarres M, Ordi-Ros J. Implication of human endogenous retroviruses in the development of autoimmune diseases. Int Rev Immunol (2010) 29:351-370

Balakrishnan M, Yant SR, Tsai L, O'Sullivan C, Bam RA, Tsai A, Niedziela-Majka A, Stray KM, Sakowicz R, Cihlar T. Non-catalytic site HIV-1 integrase inhibitors disrupt core maturation and induce a reverse transcription block in target cells. PLoS One (2013) 8:e74163

Balasubramaniam M, Freed EO. New insights into HIV assembly and trafficking. Physiology (Bethesda) (2011) 26:236-251

Ballandras-Colas A, Brown M, Cook NJ, Dewdney TG, Demeler B, Cherepanov P, Lyumkis D, Engelman AN. Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. Nature (2016) 530:358-361

Ballandras A, Moreau K, Robert X, Confort MP, Merceron R, Haser R, Ronfort C, Gouet P. A crystal structure of the catalytic core domain of an avian sarcoma and leukemia virus integrase suggests an alternate dimeric assembly. PLoS One (2011) 6:e23032

Baltimore D. Expression of animal virus genomes. Bacteriol Rev (1971) 35:235-241

Baranova S, Tuzikov FV, Zakharova OD, Tuzikova NA, Calmels C, Litvak S, Tarrago-Litvak L, Parissi V, Nevinsky GA. Small-angle X-ray characterization of the nucleoprotein complexes resulting from DNA-induced oligomerization of HIV-1 integrase. Nucleic Acids Res (2007) 35:975-987

Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science (1983) 220:868-871

Benleulmi MS, Matysiak J, Henriquez DR, Vaillant C, Lesbats P, Calmels C, Naughtin M, Leon O, Skalka AM, Ruff M, Lavigne M, Andreola ML, Parissi V. Intasome architecture and chromatin density modulate retroviral integration into nucleosome. Retrovirology (2015) 12:13

Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. Cell (2007) 128:669-681

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res (2014) 42:W252-258

Birnboim HC, Doly J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res (1979) 7:1513-1523

Blond JL, Lavillette D, Cheynet V, Bouton O, Oriol G, Chapel-Fernandes S, Mandrand B, Mallet F, Cosset FL. An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. J Virol (2000) 74:3321-3329

Bojja RS, Andrake MD, Weigand S, Merkel G, Yarychkivska O, Henderson A, Kummerling M, Skalka AM. Architecture of a full-length retroviral integrase monomer and dimer, revealed by small angle X-ray scattering and chemical cross-linking. J Biol Chem (2011) 286:17047-17059

Brenner B, Wainberg MA. We need to use the best antiretroviral drugs worldwide to prevent HIV drug resistance. Aids (2016) 30:2725-2727

Briones MS, Dobard CW, Chow SA. Role of human immunodeficiency virus type 1 integrase in uncoating of the viral core. J Virol (2010) 84:5181-5190

Brown PO, Bowerman B, Varmus HE, Bishop JM. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. Proc Natl Acad Sci U S A (1989) 86:2525-2529

Brown R, Bennett JJ. 1866. The miscellaneous botanical works of Robert Brown. London: Published for the Ray society by R. Hardwicke

Bujacz G, Alexandratos J, Qing ZL, Clement-Mella C, Wlodawer A. The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant. FEBS Lett (1996) 398:175-178

Bukrinsky M. A hard way to the nucleus. Mol Med (2004) 10:1-5

Bushman FD, Wang B. Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding. J Virol (1994) 68:2215-2223

Busschots K, Vercammen J, Emiliani S, Benarous R, Engelborghs Y, Christ F, Debyser Z. The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. J Biol Chem (2005) 280:17841-17847

Bustin M, Reeves R. High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. Prog Nucleic Acid Res Mol Biol (1996) 54:35-100

C

Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. Nat Struct Biol (1997) 4:567-577

Calmels C, de Soultrait VR, Caumont A, Desjobert C, Faure A, Fournier M, Tarrago-Litvak L, Parissi V. Biochemical and random mutagenesis analysis of the region carrying the catalytic E152 amino acid of HIV-1 integrase. Nucleic Acids Res (2004) 32:1527-1538

Carrington JC, Dougherty WG. A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. Proc Natl Acad Sci U S A (1988) 85:3391-3395

Carteau S, Gorelick RJ, Bushman FD. Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. J Virol (1999) 73:6670-6679

Carter JB, Saunders VA. Virology : principles and applications. (2007) ISBN 978-1-118-62976-5

Cellier C, Moreau K, Gallay K, Ballandras A, Gouet P, Ronfort C. In vitro functional analyses of the human immunodeficiency virus type 1 (HIV-1) integrase mutants give new insights into the intasome assembly. Virology (2013) 439:97-104

Chen JC, Krucinski J, Miercke LJ, Finer-Moore JS, Tang AH, Leavitt AD, Stroud RM. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. Proc Natl Acad Sci U S A (2000) 97:8233-8238

Cherepanov P. LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. Nucleic Acids Res (2007) 35:113-124

Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc Natl Acad Sci U S A (2005), 102:17308-17313

Cherepanov P, Devroe E, Silver PA, Engelman A. Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. J Biol Chem (2004) 279:48883-48892

Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y, De Clercq E, Debyser Z. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J Biol Chem (2003) 278:372-381

Chow SA, Vincent KA, Ellison V, Brown PO. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. Science (1992) 255:723-726

Clavel F, Guetard D, Brun-Vezinet F, Chamaret S, Rey MA, Santos-Ferreira MO, Laurent AG, Dauguet C, Katlama C, Rouzioux C, et al. Isolation of a new human retrovirus from West African patients with AIDS. Science (1986) 233:343-346

Coffin JM, Hughes SH, Varmus HE. Retroviruses. (1997) ISBN-10: 0-87969-571-4

Coleman J, Eaton S, Merkel G, Skalka AM, Laue T. Characterization of the self association of Avian sarcoma virus integrase by analytical ultracentrifugation. J Biol Chem (1999) 274:32842-32846

Collins B, Stevens RC, Page R. Crystallization Optimum Solubility Screening: using crystallization results to identify the optimal buffer for protein crystal formation. Acta Crystallogr Sect F Struct Biol Cryst Commun (2005) 61:1035-1038

Cooper DK, Gollackner B, Sachs DH. Will the pig solve the transplantation backlog? Annu Rev Med (2002) 53:133-147

Cox MM, Doudna J, O'Donnell M. Molecular Biology: Principles and Practice. (2011) W. H. Freeman

Crowe BL, Larue RC, Yuan C, Hess S, Kvaratskhelia M, Foster MP. Structure of the Brd4 ET domain bound to a C-terminal motif from gamma-retroviral integrases reveals a conserved mechanism of interaction. Proc Natl Acad Sci U S A (2016) 113:2086-2091

D

Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J Mol Biol (2002) 319:1097-1113

De Clercq E, Li G. Approved Antiviral Drugs over the Past 50 Years. Clin Microbiol Rev (2016) 29:695-747

De Rijck J, de Kogel C, Demeulemeester J, Vets S, El Ashkar S, Malani N, Bushman FD, Landuyt B, Husson SJ, Busschots K, Gijsbers R, Debyser Z. The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. Cell Rep (2013) 5:886-894

Debyser Z, Christ F, De Rijck J, Gijsbers R. Host factors for retroviral integration site selection. Trends Biochem Sci (2015) 40:108-116

Delano WL. The PyMOL Molecular Graphics System. Palo Alto, CA, USA., DeLano Scientific LLC. (2008)

Delelis O, Carayon K, Saib A, Deprez E, Mouscadet JF. Integrase and integration: biochemical activities of HIV-1 integrase. Retrovirology (2008) 5:114

Demange A, Yajjou-Hamalian H, Gallay K, Luengo C, Beven V, Leroux A, Confort MP, Al Andary E, Gouet P, Moreau K, Ronfort C, Blanchard Y. Porcine endogenous retrovirus-A/C: biochemical properties of its integrase and susceptibility to raltegravir. J Gen Virol (2015) 96:3124-3130

Demeulemeester J, Chaltin P, Marchand A, De Maeyer M, Debyser Z, Christ F. LEDGINs, non-catalytic site inhibitors of HIV-1 integrase: a patent review (2006 - 2014). Expert Opin Ther Pat (2014) 24:609-632

Denis GV, McComb ME, Faller DV, Sinha A, Romesser PB, Costello CE. Identification of transcription complexes that contain the double bromodomain protein Brd2 and chromatin remodeling machines. J Proteome Res (2006) 5:502-511

Deprez E, Tauc P, Leh H, Mouscadet JF, Auclair C, Brochon JC. Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy. Biochemistry (2000) 39:9275-9284

Desfarges S, San Filippo J, Fournier M, Calmels C, Caumont-Sarcos A, Litvak S, Sung P, Parissi V. Chromosomal integration of LTR-flanked DNA in yeast expressing HIV-1 integrase: down regulation by RAD51. Nucleic Acids Res (2006) 34:6215-6224

Dewannieux M, Ribet D, Heidmann T. Risks linked to endogenous retroviruses for vaccine production: a general overview. Biologicals (2010) 38:366-370

Dey A, Chitsaz F, Abbasi A, Misteli T, Ozato K. The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. Proc Natl Acad Sci U S A (2003) 100:8758-8763

Dobard CW, Briones MS, Chow SA. Molecular mechanisms by which human immunodeficiency virus type 1 integrase stimulates the early steps of reverse transcription. J Virol (2007) 81:10037-10046

Drapkin R, Merino A, Reinberg D. Regulation of RNA polymerase II transcription. Curr Opin Cell Biol (1993) 5:469-476

Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science (1994) 266:1981-1986

E

Eickbush TH, Malik HS. Origins and evolution of retrotransposons. Mobile DNA II (2002) pp.1111-1144. ASM Press

Eidahl JO, Crowe BL, North JA, McKee CJ, Shkriabai N, Feng L, Plumb M, Graham RL, Gorelick RJ, Hess S, Poirier MG, Foster MP, Kvaratskhelia M. Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. Nucleic Acids Res (2013) 41:3924-3936

Eijkelenboom AP, van den Ent FM, Vos A, Doreleijers JF, Hard K, Tullius TD, Plasterk RH, Kaptein R, Boelens R. The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. Curr Biol (1997) 7:739-746

Emiliani S, Mousnier A, Busschots K, Maroun M, Van Maele B, Tempe D, Vandekerckhove L, Moisant F, Ben-Slama L, Witvrouw M, Christ F, Rain JC, Dargemont C, Debyser Z, Benarous R. Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. J Biol Chem (2005) 280:25517-25523

Engelman A. In vivo analysis of retroviral integrase structure and function. Adv Virus Res (1999) 52:411-426

Engelman A, Cherepanov P. The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. PLoS Pathog (2008) 4:e1000046

Engelman A, Cherepanov P. The structural biology of HIV-1: mechanistic and therapeutic insights. Nat Rev Microbiol (2012) 10:279-290

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

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

F

Farnet CM, Haseltine WA. Integration of human immunodeficiency virus type 1 DNA in vitro. Proc Natl Acad Sci U S A (1990) 87:4164-4168

Faure A, Calmels C, Desjobert C, Castroviejo M, Caumont-Sarcos A, Tarrago-Litvak L, Litvak S, Parissi V. HIV-1 integrase crosslinked oligomers are active in vitro. Nucleic Acids Res (2005) 33:977-986

Feng Y, Baig TT, Love RP, Chelico L. Suppression of APOBEC3-mediated restriction of HIV-1 by Vif. Front Microbiol (2014) 5:450

Filippakopoulos P, Qi J, Picaud S, Shen Y, Smith WB, Fedorov O, Morse EM, Keates T, Hickman TT, Felletar I, Philpott M, Munro S, McKeown MR, Wang Y, Christie AL, West N, Cameron MJ, Schwartz B, Heightman TD, La Thangue N, French CA, Wiest O, Kung AL, Knapp S, Bradner JE. Selective inhibition of BET bromodomains. Nature (2010) 468:1067-1073

Fischle W, Wang Y, Allis CD. Histone and chromatin cross-talk. Curr Opin Cell Biol (2003) 15:172-183

Florence B, Faller DV. You bet-cha: a novel family of transcriptional regulators. Front Biosci (2001) 6:D1008-1018

Freed EO. HIV-1 assembly, release and maturation. Nat Rev Microbiol (2015) 13:484-496

Freed EO. HIV-1 replication. Somat Cell Mol Genet (2001) 26:13-33

G

Gao K, Gorelick RJ, Johnson DG, Bushman F. Cofactors for human immunodeficiency virus type 1 cDNA integration in vitro. J Virol (2003) 77:1598-1603

Ge H, Si Y, Roeder RG. Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. EMBO J (1998) 17:6723-6729

Gerton JL, Brown PO. The core domain of HIV-1 integrase recognizes key features of its DNA substrates. J Biol Chem (1997) 272:25809-25815

Goldgur Y, Craigie R, Cohen GH, Fujiwara T, Yoshinaga T, Fujishita T, Sugimoto H, Endo T, Murai H, Davies DR. 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 (1999) 96:13040-13043

Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, Davies DR. Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. Proc Natl Acad Sci U S A (1998) 95:9150-9154

Gouet P, Robert X, Courcelle E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res (2003) 31:3320-3323

Grandgenett DP, Vora AC, Schiff RD. A 32,000-dalton nucleic acid-binding protein from avian retravirus cores possesses DNA endonuclease activity. Virology (1978) 89:119-132

Graslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park HW, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim SH, Rao Z, Shi Y, Terwilliger TC, Kim CY, Hung LW, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyi T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham K, Inouye M, Anderson S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke S, Schutz A, Heinemann U, Yokoyama S, Bussow K, Gunsalus KC. Protein production and purification. Nat Methods (2008) 5:135-146

Greene WC, Debyser Z, Ikeda Y, Freed EO, Stephens E, Yonemoto W, Buckheit RW, Este JA, Cihlar T. Novel targets for HIV therapy. Antiviral Res (2008) 80:251-265

Greenwald J, Le V, Butler SL, Bushman FD, Choe S. The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. Biochemistry (1999) 38:8892-8898

Grobler JA, Stillmock K, Hu B, Witmer M, Felock P, Espeseth AS, Wolfe A, Egbertson M, Bourgeois M, Melamed J, Wai JS, Young S, Vacca J, Hazuda DJ. Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. Proc Natl Acad Sci U S A (2002) 99:6661-6666

Gupta K, Curtis JE, Krueger S, Hwang Y, Cherepanov P, Bushman FD, Van Duyne GD. Solution conformations of prototype foamy virus integrase and its stable synaptic complex with U5 viral DNA. Structure (2012) 20:1918-1928

Gupta K, Turkki V, Sherrill-Mix S, Hwang Y, Eilers G, Taylor L, McDanal C, Wang P, Temelkoff D, Nolte RT, Velthuis E, Jeffrey J, Van Duyne GD, Bushman FD. Structural Basis for Inhibitor-Induced Aggregation of HIV Integrase. PLoS Biol (2016) 14:e1002584

Gupta SS, Maetzig T, Maertens GN, Sharif A, Rothe M, Weidner-Glunde M, Galla M, Schambach A, Cherepanov P, Schulz TF. Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. J Virol (2013) 87:12721-12736

H

Haffar OK, Popov S, Dubrovsky L, Agostini I, Tang H, Pushkarsky T, Nadler SG, Bukrinsky M. Two nuclear localization signals in the HIV-1 matrix protein regulate nuclear import of the HIV-1 pre-integration complex. J Mol Biol (2000) 299:359-368

Hare S, Di Nunzio F, Labeja A, Wang J, Engelman A, Cherepanov P. Structural basis for functional tetramerization of lentiviral integrase. PLoS Pathog (2009) 5:e1000515

Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P. Retroviral intasome assembly and inhibition of DNA strand transfer. Nature (2010) 464:232-236

Hare S, Maertens GN, Cherepanov P. 3'-processing and strand transfer catalysed by retroviral integration in crystallo. EMBO J (2012) 3020-3028

Hawkins JD. Gene Structure and Expression. (1996) Cambridge University Press

Hayward WS, Neel BG, Astrin SM. Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis. Nature (1981) 290:475-480

Hindmarsh P, Ridky T, Reeves R, Andrake M, Skalka AM, Leis J. HMG protein family members stimulate human immunodeficiency virus type 1 and avian sarcoma virus concerted DNA integration in vitro. J Virol (1999) 73:2994-3003

Hnilicova J, Hozeifi S, Stejskalova E, Duskova E, Poser I, Humpolickova J, Hof M, Stanek D. The C-terminal domain of Brd2 is important for chromatin interaction and regulation of transcription and alternative splicing. Mol Biol Cell (2013) 24:3557-3568

Huang B, Yang XD, Zhou MM, Ozato K, Chen LF. Brd4 coactivates transcriptional activation of NF-kappaB via specific binding to acetylated RelA. Mol Cell Biol (2009) 29:1375-1387

I

Ikeda T, Nishitsuji H, Zhou X, Nara N, Ohashi T, Kannagi M, Masuda T. Evaluation of the functional involvement of human immunodeficiency virus type 1 integrase in nuclear import of viral cDNA during acute infection. J Virol (2004) 78:11563-11573

J

Jacks T, Townsley K, Varmus HE, Majors J. Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus gag-related polyproteins. Proc Natl Acad Sci U S A (1987) 84:4298-4302

Jacque JM, Stevenson M. The inner-nuclear-envelope protein emerin regulates HIV-1 infectivity. Nature (2006) 441:641-645

Jancarik J, Pufan R, Hong C, Kim SH, Kim R. Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins. Acta Crystallogr D Biol Crystallogr (2004) 60:1670-1673

Janin J, Delepierre M. Biologie structurale : Principes et méthodes biophysiques. (1997) Hermann

Jaskolski M, Alexandratos JN, Bujacz G, Wlodawer A. Piecing together the structure of retroviral integrase, an important target in AIDS therapy. FEBS J (2009) 276:2926-2946

Jenkins TM, Engelman A, Ghirlando R, Craigie R. A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. J Biol Chem (1996) 271:7712-7718

Jones KA, Kadonaga JT, Rosenfeld PJ, Kelly TJ, Tjian R. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell (1987) 48:79-89

Jones KS, Coleman J, Merkel GW, Laue TM, Skalka AM. Retroviral integrase functions as a multimer and can turn over catalytically. J Biol Chem (1992) 267:16037-16040

K

Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science (1994) 266:2002-2006

Kalpana GV, Reicin A, Cheng GS, Sorin M, Paik S, Goff SP. Isolation and characterization of an oligomerization-negative mutant of HIV-1 integrase. Virology (1999) 259:274-285

Kalyanaraman VS, Sarngadharan MG, Robert-Guroff M, Miyoshi I, Golde D, Gallo RC. A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. Science (1982) 218:571-573

Katz RA, Skalka AM. A C-terminal domain in the avian sarcoma-leukosis virus pol gene product is not essential for viral replication. J Virol (1988) 62:528-533

Katzman M, Sudol M. Nonspecific alcoholysis, a novel endonuclease activity of human immunodeficiency virus type 1 and other retroviral integrases. J Virol (1996) 70:2598-2604

King AMQ, Lefkowitz E, Adams MJ, Carstens EB. (2011) Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses: Elsevier

Koh Y, Matreyek KA, Engelman A. Differential sensitivities of retroviruses to integrase strand transfer inhibitors. J Virol (2011) 85:3677-3682

Kornberg RD, Thomas JO. Chromatin structure; oligomers of the histones. Science (1974) 184:865-868

Kouzarides T. Chromatin modifications and their function. Cell (2007) 128:693-705

Kozin M, Svergun D. Automated matching of high- and low-resolution structural models. J Appl Cryst (2001) 34:33-41

Krishnan L, Engelman A. Retroviral integrase proteins and HIV-1 DNA integration. J Biol Chem (2012) 287:40858-40866

Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol Cell Biol (1992) 12:2331-2338

Kulkosky J, Katz RA, Merkel G, Skalka AM. Activities and substrate specificity of the evolutionarily conserved central domain of retroviral integrase. Virology (1995) 206:448-456

Kurth R, Bannert N. (2010) Retroviruses: Molecular Biology, Genomics and Pathogenesis: Caister Academic Press

Kvaratskhelia M, Sharma A, Larue RC, Serrao E, Engelman A. Molecular mechanisms of retroviral integration site selection. Nucleic Acids Res (2014) 42:10209-10225

L

Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (1970) 227:680-685

Le Grice SF. Human immunodeficiency virus reverse transcriptase: 25 years of research, drug discovery, and promise. J Biol Chem (2012) 287:40850-40857

Le Rouzic E, Bonnard D, Chasset S, Bruneau JM, Chevreuil F, Le Strat F, Nguyen J, Beauvoir R, Amadori C, Brias J, Vomscheid S, Eiler S, Levy N, Delelis O, Deprez E, Saïb A, Zamborlini A, Emiliani S, Ruff M, Ledoussal B, Moreau F, Benarous R. Dual inhibition of HIV-1 replication by integrase-LEDGF allosteric inhibitors is predominant at the post-integration stage. Retrovirology (2013) 10:1186-1742

Lee MS, Craigie R. A previously unidentified host protein protects retroviral DNA from autointegration. Proc Natl Acad Sci U S A (1998) 95:1528-1533

Lee SK, Potempa M, Swanstrom R. The choreography of HIV-1 proteolytic processing and virion assembly. J Biol Chem (2012) 287:40867-40874

Lee SP, Xiao J, Knutson JR, Lewis MS, Han MK. Zn2+ promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. Biochemistry (1997) 36:173-180

LeRoy G, Rickards B, Flint SJ. The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. Mol Cell (2008) 30:51-60

Lesbats P, Botbol Y, Chevereau G, Vaillant C, Calmels C, Arneodo A, Andreola ML, Lavigne M, Parissi V. Functional coupling between HIV-1 integrase and the SWI/SNF chromatin remodeling complex for efficient in vitro integration into stable nucleosomes. PLoS Pathog (2011) 7:e1001280

Lesbats P, Metifiot M, Calmels C, Baranova S, Nevinsky G, Andreola ML, Parissi V. In vitro initial attachment of HIV-1 integrase to viral ends: control of the DNA specific interaction by the oligomerization state. Nucleic Acids Res (2008) 36:7043-7058

Levin A, Hayouka Z, Friedler A, Loyter A. Transportin 3 and importin alpha are required for effective nuclear import of HIV-1 integrase in virus-infected cells. Nucleus (2010) 1:422-431

Levy N, Eiler S, Pradeau-Aubreton K, Maillot B, Stricher F, Ruff M. Production of unstable proteins through the formation of stable core complexes. Nat Commun (2016) 7:10932

Lieberman-Blum SS, Fung HB, Bandres JC. Maraviroc: a CCR5-receptor antagonist for the treatment of HIV-1 infection. Clin Ther (2008) 30:1228-1250

Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res (2003) 31:3701-3708

Llano M, Delgado S, Vanegas M, Poeschla EM. Lens epithelium-derived growth factor/p75 prevents proteasomal degradation of HIV-1 integrase. J Biol Chem (2004) 279:55570-55577

Llano M, Saenz DT, Meehan A, Wongthida P, Peretz M, Walker WH, Teo W, Poeschla EM. An essential role for LEDGF/p75 in HIV integration. Science (2006) 314:461-464. United States

Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, Clore GM, Gronenborn AM. Solution structure of the DNA binding domain of HIV-1 integrase. Biochemistry (1995) 34:9826-9833

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature (1997) 389:251-260

Lutzke RA, Plasterk RH. 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 (1998) 72:4841-4848

Lv Z, Chu Y, Wang Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. HIV AIDS (2015) 7:95-104

M

Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, Debyser Z, Engelborghs Y. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. J Biol Chem (2003) 278:33528-33539

Maertens GN, Hare S, Cherepanov P. The mechanism of retroviral integration from X-ray structures of its key intermediates. Nature (2010) 468:326-329

Mahmood T, Yang PC. Western blot: technique, theory, and trouble shooting. N Am J Med Sci (2012) 4:429-434

Maignan S, Guilloteau JP, Zhou-Liu Q, Clement-Mella C, Mikol V. Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. J Mol Biol (1998) 282:359-368

Maillot B, Levy N, Eiler S, Crucifix C, Granger F, Richert L, Didier P, Godet J, Pradeau-Aubreton K, Emiliani S, Nazabal A, Lesbats P, Parissi V, Mely Y, Moras D, Schultz P, Ruff M. Structural and functional role of INI1 and LEDGF in the HIV-1 preintegration complex. PLoS One (2013) 8:e60734

Mandal D, Prasad VR. Analysis of 2-LTR circle junctions of viral DNA in infected cells. Methods Mol Biol (2009) 485:73-85

Mantovani F, Drost J, Voorhoeve PM, Del Sal G, Agami R. Gene regulation and tumor suppression by the bromodomain-containing protein BRD7. Cell Cycle (2010) 9:2777-2781

Marchand C, Johnson AA, Karki RG, Pais GC, Zhang X, Cowansage K, Patel TA, Nicklaus MC, Burke TR, Jr., Pommier Y. Metal-dependent inhibition of HIV-1 integrase by beta-diketo acids and resistance of the soluble double-mutant (F185K/C280S). Mol Pharmacol (2003) 64:600-609

Martin GS. The road to Src. Oncogene (2004) 23:7910-7917

Maskell DP, Renault L, Serrao E, Lesbats P, Matadeen R, Hare S, Lindemann D, Engelman AN, Costa A, Cherepanov P. Structural basis for retroviral integration into nucleosomes. Nature (2015) 523:366-369

Matuo Y, Matsui S, Nishi N, Wada F, Sandberg AA. Quantitative solubilization of nonhistone chromosomal proteins without denaturation using zwitterionic detergents. Anal Biochem (1985) 150:337-344

McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics (2000) 16:404-405

Metifiot M, Marchand C, Pommier Y. HIV integrase inhibitors: 20-year landmark and challenges. Adv Pharmacol (2013) 67:75-105

Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, Keith JC, Jr., McCoy JM. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature (2000) 403:785-789

Michel F, Crucifix C, Granger F, Eiler S, Mouscadet JF, Korolev S, Agapkina J, Ziganshin R, Gottikh M, Nazabal A, Emiliani S, Benarous R, Moras D, Schultz P, Ruff M. Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor. EMBO J (2009) 28:980-991

Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, Ecker JR, Bushman FD. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol (2004) 2:E234

Moalic Y, Blanchard Y, Felix H, Jestin A. Porcine endogenous retrovirus integration sites in the human genome: features in common with those of murine leukemia virus. J Virol (2006) 80:10980-10988

Molteni V, Greenwald J, Rhodes D, Hwang Y, Kwiatkowski W, Bushman FD, Siegel JS, Choe S. Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain. Acta Crystallogr D Biol Crystallogr (2001) 57:536-544

Moreau K, Faure C, Violot S, Gouet P, Verdier G, Ronfort C. Mutational analyses of the core domain of Avian Leukemia and Sarcoma Viruses integrase: critical residues for concerted integration and multimerization. Virology (2004) 318:566-581

Moreau K, Faure C, Violot S, Verdier G, Ronfort C. Mutations in the C-terminal domain of ALSV (Avian Leukemia and Sarcoma Viruses) integrase alter the concerted DNA integration process in vitro. Eur J Biochem (2003) 270:4426-4438

Moriniere J, Rousseaux S, Steuerwald U, Soler-Lopez M, Curtet S, Vitte AL, Govin J, Gaucher J, Sadoul K, Hart DJ, Krijgsveld J, Khochbin S, Muller CW, Petosa C. Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. Nature (2009) 461:664-668

Mulder LC, Chakrabarti LA, Muesing MA. The SET complex acts as a barrier to autointegration of HIV-1. J Biol Chem (2002) 5:27489-27493

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol (1986) 51 Pt 1:263-273

N

Nakamura Y, Umehara T, Nakano K, Jang MK, Shirouzu M, Morita S, Uda-Tochio H, Hamana H, Terada T, Adachi N, Matsumoto T, Tanaka A, Horikoshi M, Ozato K, Padmanabhan B, Yokoyama S. Crystal structure of the human BRD2 bromodomain: insights into dimerization and recognition of acetylated histone H4. J Biol Chem (2007) 282:4193-4201

Navia MA, Fitzgerald PM, McKeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature (1989) 337:615-620

Neamati N, Wang B. (2011) HIV-1 Integrase: Mechanism and Inhibitor Design: Wiley

Nisole S, Stoye JP, Saib A. TRIM family proteins: retroviral restriction and antiviral defence. Nat Rev Microbiol (2005) 3:799-808

O

Ornstein L. Disc electrophoresis. I. Background and therory. Ann N Y Acad Sci (1964) 121:321-349

Oz I, Avidan O, Hizi A. Inhibition of the integrases of human immunodeficiency viruses type 1 and type 2 by reverse transcriptases. Biochem J (2002) 361:557-566

P

Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. Protein Sci (1995) 4:2411-2423

Patience C, Takeuchi Y, Weiss RA. Infection of human cells by an endogenous retrovirus of pigs. Nat Med (1997) 3:282-286

Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HD, Konarev PV, Svergun DI. New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Crystallogr (2012) 45:342-350

Poeschla EM. Integrase, LEDGF/p75 and HIV replication. Cell Mol Life Sci (2008) 65:1403-1424

Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc Natl Acad Sci U S A (1980) 77:7415-7419

Porath J, Carlsson J, Olsson I, Belfrage G. Metal chelate affinity chromatography, a new approach to protein fractionation. Nature (1975) 258:598-599

Primrose SB, Twyman R. (2006) Principles of Gene Manipulation and Genomics: Wiley-Blackwell

Pruss D, Reeves R, Bushman FD, Wolffe AP. The influence of DNA and nucleosome structure on integration events directed by HIV integrase. J Biol Chem (1994) 269:25031-25041

Pryciak PM, Varmus HE. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell (1992) 69:769-780

Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys (2007) 40:191-285

R

Reddy KK, Singh P, Singh SK. Blocking the interaction between HIV-1 integrase and human LEDGF/p75: mutational studies, virtual screening and molecular dynamics simulations. Mol Biosyst (2014) 10:526-536

Rety S, Reaeabkova L, Dubanchet B, Silhan J, Legrand P, Lewit-Bentley A. Structural studies of the catalytic core of the primate foamy virus (PFV-1) integrase. Acta Crystallogr Sect F Struct Biol Cryst Commun (2010) 66:881-886

Rice PA, Baker TA. Comparative architecture of transposase and integrase complexes. Nat Struct Biol (2001) 8:302-307

Rous P. A sarcoma of the fowl transmissible by an agent separable from the tumor cells. J Exp Med (1911) 13:397-411

Ruprecht K, Mayer J, Sauter M, Roemer K, Mueller-Lantzsch N. Endogenous retroviruses and cancer. Cell Mol Life Sci (2008) 65:3366-3382

S

Sachs DH. The pig as a potential xenograft donor. Vet Immunol Immunopathol (1994) 43:185-191

Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. Evaluation of comparative protein modeling by MODELLER. Proteins (1995) 23:318-326

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A (1977) 74:5463-5467

Scherdin U, Rhodes K, Breindl M. Transcriptionally active genome regions are preferred targets for retrovirus integration. J Virol (1990) 64:907-912

Schmidt TG, Skerra A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. Nat Protoc (2007) 2:1528-1535

Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. Cell (2002) 110:521-529

Semenova EA, Marchand C, Pommier Y. HIV-1 integrase inhibitors: update and perspectives. Adv Pharmacol (2008) 56:199-228

Serrao E, Engelman AN. Sites of retroviral DNA integration: From basic research to clinical applications. Crit Rev Biochem Mol Biol (2016) 51:26-42

Sherman PA, Fyfe JA. Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. Proc Natl Acad Sci U S A (1990) 87:5119-5123

Shi K, Pandey KK, Bera S, Vora AC, Grandgenett DP, Aihara H. A possible role for rhe asymmetric C-terminal domain dimer of Rous sarcoma virus integrase in viral DNA binding. PLoS One (2013) 8:e56892

Singh DP, Kimura A, Chylack LT, Jr., Shinohara T. Lens epithelium-derived growth factor (LEDGF/p75) and p52 are derived from a single gene by alternative splicing. Gene (2000) 242:265-273

Singh PK, Plumb MR, Ferris AL, Iben JR, Wu X, Fadel HJ, Luke BT, Esnault C, Poeschla EM, Hughes SH, Kvaratskhelia M, Levin HL. LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. Genes Dev (2015) 29:2287-2297

Sinha A, Faller DV, Denis GV. Bromodomain analysis of Brd2-dependent transcriptional activation of cyclin A. Biochem J (2005) 387:257-269

Smith DB, Johnson KS. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene (1988) 67:31-40

Studier FW. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif (2005) 41:207-234

Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol (1986) 189:113-130

Sundquist WI, Krausslich HG. HIV-1 assembly, budding, and maturation. Cold Spring Harb Perspect Med (2012) 2:a006924

Suzuki Y, Craigie R. The road to chromatin – nuclear entry of retroviruses. Nat Rev Microbiol (2007) 5:187-196. England

Svergun D, Koch MHJ. Small angle scattering studies of biological macromolecules in solution. Reports on Progression in Physics. (2003) 66:1735-1782

T

Tarlinton RE, Meers J, Young PR. Retroviral invasion of the koala genome. Nature (2006) 442:79-81

Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol (2007) 14:1025-1040

Taylor RG, Walker DC, McInnes RR. *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. Nucleic Acids Res (1993) 21:1677-1678

Tegel H, Ottosson J, Hober S. Enhancing the protein production levels in *Escherichia coli* with a strong promoter. FEBS J (2011) 278:729-739

Temin HM, Mizutani S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature (1970) 226:1211-1213

Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol (2003) 60:523-533

Trobridge GD, Miller DG, Jacobs MA, Allen JM, Kiem HP, Kaul R, Russell DW. Foamy virus vector integration sites in normal human cells. Proc Natl Acad Sci U S A (2006) 103:1498-1503

Turlure F, Devroe E, Silver PA, Engelman A. Human cell proteins and human immunodeficiency virus DNA integration. Front Biosci (2004) 9:3187-3208

Turlure F, Maertens G, Rahman S, Cherepanov P, Engelman A. A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs, mediates the association of LEDGF/p75 with chromatin in vivo. Nucleic Acids Res (2006) 34:1653-1665

U

Umehara T, Nakamura Y, Jang MK, Nakano K, Tanaka A, Ozato K, Padmanabhan B, Yokoyama S. Structural basis for acetylated histone H4 recognition by the human BRD2 bromodomain. J Biol Chem (2010) 285:7610-7618

V

Van Maele B, Busschots K, Vandekerckhove L, Christ F, Debyser Z. Cellular co-factors of HIV-1 integration. Trends Biochem Sci (2006) 31:98-105

Vandegraaff N, Devroe E, Turlure F, Silver PA, Engelman A. Biochemical and genetic analyses of integrase-interacting proteins lens epithelium-derived growth factor (LEDGF)/p75 and hepatoma-derived growth factor related protein 2 (HRP2) in preintegration complex function and HIV-1 replication. Virology (2006) 346:415-426

Vink C, Lutzke RA, Plasterk RH. Formation of a stable complex between the human immunodeficiency virus integrase protein and viral DNA. Nucleic Acids Res (1994) 22:4103-4110

Violot S, Hong SS, Rakotobe D, Petit C, Gay B, Moreau K, Billaud G, Priet S, Sire J, Schwartz O, Mouscadet JF, Boulanger P. The human polycomb group EED protein interacts with the integrase of human immunodeficiency virus type 1. J Virol (2003) 77:12507-12522

Vranckx LS, Demeulemeester J, Saleh S, Boll A, Vansant G, Schrijvers R, Weydert C, Battivelli E, Verdin E, Cereseto A, Christ F, Gijsbers R, Debyser Z. LEDGIN-mediated Inhibition of Integrase-LEDGF/p75 Interaction Reduces Reactivation of Residual Latent HIV. EBioMedicine (2016) 8:248-264

W

Wang F, Liu H, Blanton WP, Belkina A, Lebrasseur NK, Denis GV. Brd2 disruption in mice causes severe obesity without Type 2 diabetes. Biochem J (2009) 425:71-83

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

Weidner-Glunde M, Ottinger M, Schulz TF. WHAT do viruses BET on? Front Biosci (Landmark Ed) (2010) 15:537-549

Weiss RA. The discovery of endogenous retroviruses. Retrovirology (2006) 3:67

Weiss RA, Vogt PK. 100 years of Rous sarcoma virus. J Exp Med (2011) 208:2351-2355

Wensing AM, Calvez V, Gunthard HF, Johnson VA, Paredes R, Pillay D, Shafer RW, Richman DD. Update of the drug resistance mutations in HIV-1. Top Antivir Med (2014) 22:642-650

Whitmore L, Wallace BA. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res (2004) 32:W668-673

Wielens J, Crosby IT, Chalmers DK. A three-dimensional model of the human immunodeficiency virus type 1 integration complex. J Comput Aided Mol Des (2005) 19:301-317

Wilen CB, Tilton JC, Doms RW. HIV: cell binding and entry. Cold Spring Harb Perspect Med (2012)

Wilkinson TA, Chow SA, Neamati N. Functional interaction between human immunodeficiency virus type 1 reverse transcriptase and integrase, in HIV-1 Integrase, Mechanism and Inhibitor Design. (2011):95-103 Wiley

Wlodawer A, Erickson JW. Structure-based inhibitors of HIV-1 protease. Annu Rev Biochem (1993) 62:543-585

Wright K. AIDS therapy. First tentative signs of therapeutic promise. Nature (1986) 323:283

Wu SY, Chiang CM. The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. J Biol Chem (2007) 282:13141-13145

Y

Yamashita M, Emerman M. Retroviral infection of non-dividing cells: old and new perspectives. Virology (2006) 344:88-93

Yang F, Roth MJ. Assembly and catalysis of concerted two-end integration events by Moloney murine leukemia virus integrase. J Virol (2001) 75:9561-9570

Yang ZN, Mueser TC, Bushman FD, Hyde CC. Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. J Mol Biol (2000) 296:535-548

Yin Z, Shi K, Banerjee S, Pandey KK, Bera S, Grandgenett DP, Aihara H. Crystal structure of the Rous sarcoma virus intasome. Nature (2016) 530:362-366

Z

Zennou V, Petit C, Guetard D, Nerhbass U, Montagnier L, Charneau P. HIV-1 genome nuclear import is mediated by a central DNA flap. Cell (2000) 101:173-185

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

Zimm BH. The dependence of the scattering of light on angle and concentration in linear polymer solutions. J Phys Colloid Chem (1948) 52:260-267