



HAL
open science

Étude moléculaire et structurale de l'intégrase du virus porcin PERV et de son partenaire cellulaire humain Brd2, dans le cadre de la prévention de xénozoonoses rétrovirales

Margaux Chahpazoff

► **To cite this version:**

Margaux Chahpazoff. Étude moléculaire et structurale de l'intégrase du virus porcin PERV et de son partenaire cellulaire humain Brd2, dans le cadre de la prévention de xénozoonoses rétrovirales. Biochimie, Biologie Moléculaire. Université de Lyon, 2020. Français. NNT : 2020LYSE1003 . tel-02613815

HAL Id: tel-02613815

<https://theses.hal.science/tel-02613815>

Submitted on 20 May 2020

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

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



N°d'ordre NNT : 2020LYSE1003

THESE de DOCTORAT DE L'UNIVERSITE DE LYON

Opérée au sein de

L'Université Claude Bernard Lyon 1

École Doctorale N° 205

EDISS – École Doctorale Interdisciplinaire Sciences-Santé

Spécialité de doctorat : Biochimie

Soutenue publiquement le 14/01/2020, par :

Margaux CHAHPAZOFF

**Étude moléculaire et structurale de l'intégrase du virus
porcin PERV et de son partenaire cellulaire humain Brd2,
dans le cadre de la prévention de zoonoses rétrovirales**

Devant le jury composé de :

MIELE Adriana	PR	Université Claude Bernard Lyon 1	Examinatrice
ANDREOLA Marie-Line	DR	CNRS, Université de Bordeaux	Rapporteuse
CHERRIER Mickaël	MC	Université Grenoble Alpes	Rapporteur
REGUERA Juan	CR	INSERM, Marseille	Rapporteur
BLANCHARD Yannick	IR	ANSES, Ploufragan - Plouzané	Examineur
GOUET Patrice	PR	Université Claude Bernard Lyon 1	Directeur de thèse
FIORINI Francesca	CR	CNRS, Lyon	Invitée

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université

Président du Conseil Académique

Vice-président du Conseil d'Administration

Vice-président du Conseil des Études et de la vie Universitaire

Vice-président de la Commission Recherche

Directeur Général des Services

M. Frédéric FLEURY

M. Hamda BEN HADID

M. Didier REVEL

M. Philippe CHEVALIER

M. Damien VERHAEGHE

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard

Doyen : M. Gilles RODE

Faculté de Médecine et Maïeutique Lyon Sud – Charles Mérieux

Doyenne : Mme Carole BURILLON

UFR d'Odontologie

Doyenne : Mme Dominique SEUX

Institut des Sciences Pharmaceutiques et Biologiques

Directrice : Mme Christine

VINCIGUERRA

Institut des Sciences et Techniques de la Réadaptation

Directeur : M. Xavier PERROT

Département de Formation et Centre de Recherche en Biologie Humaine

Directrice : Mme Anne-Marie SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

UFR Biosciences

Directrice : Mme Kathrin GIESELER

Département Génie Electrique et des Procédés (GEP)

Directrice : Mme Rosaria FERRIGNO

Département Informatique

Directeur : M. Behzad SHARIAT

Département Mécanique

Directeur : M. Marc BUFFAT

UFR – Faculté des Sciences

Administrateur provisoire : M. Bruno ANDRIOLETTI

UFR (STAPS)

Directeur : M. Yannick VANPOULLE

Observatoire de Lyon

Directrice : Mme Isabelle DANIEL

École Polytechnique Universitaire Lyon 1

Directeur : M. Emmanuel PERRIN

École Supérieure de Chimie, Physique, Électronique (CPE Lyon)

Directeur : M. Gérard PIGNAULT

Institut Universitaire de Technologie de Lyon 1

Directeur : M. Christophe VITON

Institut de Science Financière et d'Assurance

Directeur : M. Nicolas LEBOISNE

ESPE

Administrateur provisoire : M. Pierre CHAREYRON

REMERCIEMENTS

Le travail réalisé au cours de ma thèse n'aurait pas été le même sans l'aide, scientifique ou non, des personnes suivantes.

Je tiens tout d'abord à remercier la Région Auvergne – Rhône-Alpes pour m'avoir attribuée une allocation de recherche.

Je souhaite remercier le Dr. Christophe Geourjon, Directeur de l'Institut de Biologie et Chimie des Protéines et le Dr. Jean-Michel Jault, Directeur de l'Unité « Microbiologie Moléculaire et Biochimie Structurale », pour leur accueil durant ces quatre années.

Je tiens à remercier le Pr. Adriana Miele, le Dr. Mickaël Cherrier, le Dr. Juan Reguera, le Dr. Marie-Line Andreola et le Dr. Yannick Blanchard qui m'ont fait l'honneur de juger mon travail de thèse. Je remercie également le Dr. Jean-Michel Jault et le Dr. Pierre Jalinot pour leur écoute et leurs conseils lors de mon suivi de thèse.

J'exprime ma profonde reconnaissance à mon directeur de thèse, le Pr. Patrice Gouet, pour toute la confiance qu'il m'a accordée dès mon stage de Master, ainsi que pour sa patience et sa présence.

J'adresse également mes remerciements à l'équipe du Dr. Yannick Blanchard et à Aurélie Leroux pour leur disponibilité et l'aide apportée au cours de ce projet.

Tout du long de ma thèse j'ai pu utiliser les équipements de la plateforme *Protein Science Facility* PSF (SFR Biosciences, UMS3444). Je tiens à remercier Virginie Gueguen-Chaignon, Éric Diesis, Roland Montserret et Isabelle Zanela-Cléon. Je remercie également Daphna Fenel (Plateforme de Microscopie Électronique, UMR5075, IBS Grenoble) et le Dr. Paolo Swuec (Plateforme de Cryo-Microscopie, Université de Milan) pour leur accueil dans leurs laboratoires respectifs.

Je souhaite remercier les membres de l'équipe « Biocristallographie et Biologie Structurale des Cibles Thérapeutiques » du Dr. Nushin Aghajari. Merci à Fred pour tous nos fous rires et moments à ré-imaginer la science ; à Seb pour tous tes bons conseils cinéma et ta bienveillance et à Lionel pour toutes tes blagues et pour m'avoir laissée t'embêter autant que tu m'as embêtée. Merci à Léa pour ta bonne humeur et de m'aider chaque jour un peu plus à rester jeune dans ma tête ! Merci également à Loïc, notre thésard en chef parti bien vite de l'autre côté de la Manche. Merci également à tous ceux avec qui j'aurais pu partager le labo plus ou moins longtemps : Hollie, Asako, Vinesh et Minh Chau.

Cette expérience n'aurait pas été la même sans la présence quotidienne de tous les membres de mon équipe. Je tiens donc à remercier Xav, pour les boosts de confiance, nos débats sur la vie et le monde ainsi que pour tout ton soutien. Francesca, merci pour toute ton aide et pour m'avoir poussée à continuer d'y croire jusqu'au bout. Merci à Christophe pour tes précieux conseils et pour le partage de ton inépuisable culture générale. Merci à Marie pour ta gentillesse. Merci à Cecilia et Mathieu pour tous les bons moments ; à vous la relève ! Enfin merci à Halima et Stelle

pour ces belles amitiés que nous avons créées. Halima, merci de m'avoir accueillie si chaleureusement, ton soutien m'a mené jusqu'ici et je ne t'en remercierai jamais assez. Stelle, merci d'avoir été continuellement présente toutes ces années, pour nos soirées sur la guinguette, nos séjours chez Mickey et toutes nos aventures. Vivement les prochaines !

Ces trois années de thèse auraient été bien tristes sans la présence à l'IBCP de Marine (on aura vécu ces montagnes russes en même temps et on s'en sortira ensemble !), Sylvain (merci de nous emmener respirer l'air frais de la nature si souvent), Stéphanie, Julien, Arthur, Maxime, Alexis, Guillaume, Célia, Thomas et Alexandre. Merci à vous tous pour les sessions Synchrotron, les soirées au Bureau, Croustibat, les pub quizz et les retraites MMSB. Merci également à ceux que j'ai eu le plaisir de croiser quotidiennement au détour d'un couloir, Marine, Frank, Anaïs, Veronica, Benjamin et Maty.

Je ne peux pas écrire de remerciements sans citer Prune, Cassandre, Aurélie, Line, Camille, Marie-Laure, Caroline et Aline. Il m'est difficile d'imaginer une époque où vous n'étiez pas dans ma vie et je suis heureuse de penser à tout ce que l'on a vécu ensemble. Merci à vous d'avoir toujours été là pour me sortir la tête du labo, que ce soit pour un resto, un week-end d'exploration ou juste un ciné.

À mes Magnifiques, merci. Pour tout. Je suis chanceuse de vous avoir dans ma vie, de vous avoir eu à mes côtés lors de ces trois années et de pouvoir compter sur vous à tout moment. Pour toute votre attention, pour vos bons conseils et pour faire de moi une personne meilleure chaque jour depuis notre rencontre, merci.

Mes derniers remerciements ne sont pas les moindres puisqu'ils vont à toute ma famille. Merci pour tout ce que vous m'avez donné et apporté. Vous avez toujours cru en moi, vous m'avez aidée et poussée à persévérer et je vous en serai pour toujours reconnaissante. Je vous aime.

Merci à tous de m'avoir aidée à en arriver jusque-là.

TABLE DES MATIÈRES

REMERCIEMENTS	5
TABLE DES MATIÈRES	9
LISTE DES ABRÉVIATIONS	13
INTRODUCTION	17
I. Les rétrovirus	19
I. 1. Généralités	19
I. 2. Classification des rétrovirus	20
I. 2. 1. Taxonomie des virus	20
I. 2. 2. Virus endogènes et exogènes	21
I. 2. 3. Complexité du génome	23
I. 3. Organisation génomique et structurale	23
I. 3. 1. La particule rétrovirale	23
I. 3. 2. Le génome rétroviral	24
I. 3. 3. Les rétrovirus complexes	26
I. 4. Le cycle réplcatif rétroviral	27
I. 4. 1. La phase pré-intégrative	27
I. 4. 2. La phase post-intégrative	29
II. L'intégrase	30
II. 1. Fonctions de l'intégrase	30
II. 1. 1. Complexe de pré-intégration	30
II. 1. 2. Intégration concertée	32
II. 1. 3. Autres fonctions de l'intégrase	34
II. 2. Structure de l'intégrase	35
II. 2. 1. Le domaine amino-terminal	36
II. 2. 2. Le domaine central catalytique	37
II. 2. 3. Le domaine carboxy-terminal	38
II. 2. 4. Structures connues	39
II. 2. 5. Intasomes et états oligomériques	40
II. 2. 6. Intégrase et nucléosome	46
II. 3. Partenaires cellulaires	48
II. 3. 1. LEDGF/p75	48
II. 3. 2. Les protéines BET	50
II. 4. Inhibiteurs de l'intégrase	53
II. 4. 1. Les INSTI – Inhibiteur du transfert de brin	53
II. 4. 2. Les ALLINI – Inhibiteurs allostériques	56
PROJET DOCTORAL	57

MATÉRIELS ET MÉTHODES	63
I. Techniques de biologie moléculaire.....	65
I. 1. Constructions plasmidiques	65
I. 1. 1. Les constructions plasmidiques MultiTEV	66
I. 1. 2. Construction plasmidique de la protéine Brd2 (667-808).....	68
I. 1. 3. Les constructions plasmidiques de l'IN de PERV-A/C	68
I. 2. Mutagenèse dirigée par PCR.....	70
I. 3. Transformation bactérienne	72
I. 4. Extraction de l'ADN plasmidique.....	73
I. 5. Dosage de l'ADN.....	73
I. 6. Séquençage de l'ADN.....	74
II. Surproduction et purification des protéines d'intérêt	74
II. 1. MultiTEV / MultiTEV Minimal	74
II. 1. 1. Surproduction en système procaryote.....	74
II. 1. 2. Lyse bactérienne	76
II. 1. 3. Purification des protéines d'intérêt.....	76
II. 1. 4. Concentration protéique.....	79
II. 2. Brd2 (667-808)	80
II. 2. 1. Surproduction en système procaryote.....	80
II. 2. 2. Lyse bactérienne	80
II. 2. 3. Purification de Brd2 (667-808).....	81
II. 2. 4. Clivage de l'étiquette d'affinité par la protéase du TEV	81
II. 3. L'intégrase entière de PERV-A/C	82
II. 3. 1. Tests de surproduction en système bactérien.....	82
II. 3. 2. Lyse bactérienne	83
II. 3. 3. Solubilisation de l'IN de PERV-A/C.....	85
II. 3. 4. Purification de l'IN de PERV-A/C	85
II. 3. 5. Préparation du substrat d'ADN de l'intasome	87
II. 3. 6. Assemblage de l'intasome de PERV-A/C avec Brd2 (667-808).....	88
III. Techniques analytiques	90
III. 1. Électrophorèse sur gel discontinu de polyacrylamide en conditions dénaturantes.....	90
III. 1. 1. Révélation au bleu de Coomassie.....	92
III. 1. 2. Révélation au nitrate d'argent	92
III. 2. Immunodétection après électrotransfert sur membrane de nitrocellulose	92
III. 3. Pontage covalent.....	94
III. 3. 1. Pontage covalent au BS ³	94
III. 3. 2. Pontage covalent au GTA.....	94
III. 4. Diffusion dynamique de la lumière (DLS).....	94
IV. Mesure d'interaction	95

IV. 1. Titration calorimétrique isotherme (ITC).....	95
IV. 2. Essais de type pull down	97
V. Étude structurale	97
V. 1. Diffusion des rayons X aux petits angles	97
V. 2. Cristallographie aux rayons X	101
V. 2. 1. Cristallogenèse	101
V. 2. 2. Criblage des conditions de cristallisation	103
V. 2. 3. Diffraction aux rayons X	104
V. 3. Criblage virtuel de chimiothèque	106
V. 4. Microscopie électronique en transmission	107
RÉSULTATS ET DISCUSSION	111
I. Étude de l'interaction entre l'IN de PERV-A/C et Brd2.....	113
I. 1. Le CTD de l'IN PERV-A/C et Brd2 (667-808) – Stratégie MultiTEV	116
I. 1. 1. Surproduction et purification du complexe MultiTEV	116
I. 1. 2. Étude in vitro de la dissociation du complexe MultiTEV	123
I. 1. 3. Étude structurale aux rayons X du complexe MultiTEV	126
I. 2. Le CTD de l'IN PERV-A/C et Brd2 (667-735) – Stratégie MultiTEV Minimal.....	129
I. 2. 1. Surproduction et purification du complexe MultiTEV Minimal.....	129
I. 2. 2. Caractérisation du complexe MultiTEV Minimal.....	131
I. 2. 3. Études structurales du complexe MultiTEV Minimal.....	134
I. 3. Publication	143
II. Étude de l'intasome de PERV-A/C	156
II. 1. Expression et purification de l'IN de PERV-A/C.....	156
II. 1. 1. Tests de surproduction en système bactérien.....	156
II. 1. 2. Solubilisation de l'IN de PERV-A/C.....	159
II. 1. 3. Tests de purification de l'IN de PERV-A/C	161
II. 2. Surproduction et purification de Brd2 (667-808)	164
II. 2. 1. Purification de Brd2 (667-808).....	164
II. 2. 2. Clivage de l'étiquette d'affinité	166
II. 3. Assemblage et purification de l'intasome de PERV-A/C.....	168
II. 3. 1. Tests d'assemblage de l'intasome de PERV-A/C.....	168
II. 3. 2. Étude structurale par microscopie électronique à transmission	178
II. 4. Criblage de molécules compétitrices de l'intasome de PERV-A/C.....	181
II. 4. 1. Identification des molécules	181
II. 4. 2. Test des molécules par <i>pull down</i>	183
II. 5. Discussion.....	185
CONCLUSION ET PERSPECTIVES	187

LISTE DES ABRÉVIATIONS

A₂₆₀	Absorbance à 260 nm
ADN	Acide DésoxyriboNucléique
ALV	<i>Avian Leukosis Virus</i>
ALLINI	<i>Allosteric Integrase Inhibitors</i> – Inhibiteurs allostériques de l'intégrase
APS	<i>Ammonium PerSulfate</i> – Persulfate d'ammonium
ASLV	<i>Avian Sarcoma Leukosis Virus</i>
ATP	Adénosine TriPhosphate
ARN	Acide RiboNucléique
ARNm	Acide RiboNucléique messenger
BCIP	5-Bromo-4-Chloro-3-Indolyl Phosphate
BD	Bromodomaines
BET	Bromodomaines et domaine Extra-Terminal
BS³	Bis(SulfoSuccinimidyl) Subérate
CA	protéine de CApside
CAPS	Acide 3-(CyclohexylAmino)-1-PropaneSulfonique
CCD	<i>Catalytic Core Domain</i> – Cœur Central Catalytique
CDC	<i>Clived Donor Complex</i>
CHAPS	3-[(3-Cholamidopropyl)diméthylammonio]-1-propanesulfonate
CIC	<i>Conserved Intasome Core</i>
CTD	<i>Carboxy-Terminal Domain</i> – Domaine carboxy-terminal
DLS	<i>Dynamic Light Scattering</i> – Diffusion dynamique de la lumière
D_{max}	Distance interatomique maximale
ddNTP	Didésoxyribonucléotide triphosphate
dNTP	Désoxyribonucléotide triphosphate
DTG	Dolutégravir
EBM	<i>ET Binding Motif</i>
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	<i>EthleneDiamineTetraAcetic Acid</i>
ESRF	<i>European Synchrotron Radiation Facility</i>
ET	domaine Extra-Terminal
EVG	Elvitégravir
FDA	<i>Food and Drug Administration</i>
FeLV	<i>Feline Leukemia Virus</i>
FPLC	<i>Fast Protein Liquid Chromatography</i> – Chromatographie liquide à basse pression
H3K36me3	Lysine 36 triméthylée de la queue d'histone 3
HAART	<i>Highly Active AntiRetroviral Therapy</i> – Thérapie antirétrovirale hautement active
HEPES	Acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
HERV	<i>Human Endogenous RetroVirus</i>

HRP	<i>Hepatoma-derived growth factor Related Protein</i>
HTLV	<i>Human T-Lymphotropic Virus</i>
Gag	<i>Group-specific antigene</i>
GST	Glutathion-S-Transférase
GTA	Glutaraldéhyde
IBD	<i>Integrase Binding Domain</i> – Domaine de liaison à l'intégrase
IN	Intégrase
INI1	<i>Integrase Interactor 1</i>
INSTI	<i>Integrase Strand Transfer Inhibitor</i> – Inhibiteur du transfert de brin
IPTG	IsoPropylThioGalactoside
ITC	<i>Isothermal Titration Calorimetry</i> – Titration calorimétrique isotherme
KSHV	<i>Kaposi's Sarcoma-associated HerpesVirus</i>
LAV	<i>Lymphadenopathy-Associated Virus</i>
LB	<i>Lysogeny Broth</i>
LEDGF	<i>Lens Epithelium-Derived Growth Factor</i>
LTR	<i>Long Terminal Repeat</i>
MA	protéine de MAtrice
MCS	<i>Multiple Cloning Site</i> – Site multiple de clonage
MLV	<i>Murine Leukemia Virus</i> – Virus de la leucémie murine
MMTV	<i>Mouse Mammary Tumor Virus</i>
MVV	<i>Maedi-Visna Virus</i>
NBT	NitroBlue Tetrazolium
NC	NucléoCapside
NED	<i>N-terminal Extension Domain</i>
Nef	<i>Negative regulatory factor</i>
NHS	N-HydroxySuccinimide
NLS	<i>Nuclear Localization Signal</i> – Signal de localisation nucléaire
NTA	<i>NitriloTriacetic Acid</i> – Acide nitrilotriacétique
NTD	<i>amiNo-Terminal Domain</i> – Domaine amino-terminal
PAGE	<i>PolyAcrylamide Gel Electrophoresis</i>
PCR	<i>Polymerase Chain Reaction</i>
PdI	<i>Polydispersity Index</i> – Indice de polydispersité
PEG	Polyéthylène glycol
PERV	<i>Porcin Endogenous RetroVirus</i> – Virus endogène porcin
PES	PolyEther Sulfone
PFV	<i>Prototype Foamy Virus</i>
PIC	<i>Pre-Integration Complex</i> – Complexe de pré-intégration
PR	PRotéase
RAL	Raltégravir
REV	<i>avian ReticuloEndotheliosis Virus</i>
Rev	<i>Regulator of expression of virion proteins</i>
R_g	Rayon de giration
R_H	Rayon hydrodynamique

RING	<i>Really Interesting - New Gene</i>
RMN	Résonance Magnétique Nucléaire
RNase H	Ribonucléase H
RSV	<i>Rous Sarcoma Virus</i>
RT	<i>Reverse Transcriptase – Transcriptase Inverse</i>
RTC	<i>Reverse Transcription Complex – Complexe de transcription inverse</i>
SAXS	<i>Small-Angle X-ray Scattering – Diffusion des rayons X aux petits angles</i>
SDS	<i>Sodium Dodecyl Sulfate</i>
SEC	<i>Size Exclusion Chromatography – Chromatographie d'exclusion de taille</i>
SIDA	Syndrome d'ImmunoDéficiency Acquis
SSC	<i>Stable Synaptic Complex</i>
STC	<i>Strand Transfer Complex</i>
SU	glycoprotéine de SURface
TAE	Tris-Acétate/EDTA
Tat	<i>Transactivator of transcription</i>
TCC	<i>Target Capture Complex</i>
TEM	<i>Transmission Electron Microscopy – Microscopie électronique en transmission</i>
TEMED	N,N,N',N'-tétraméthyl-éthylènediamine
TEV	<i>Tobacco Etch Virus</i>
TM	glycoprotéine TransMembranaire
Vif	<i>Virion infectivity factor</i>
VIH	Virus de l'Immunodéficiency Humaine
Vpr	<i>Viral protein R</i>
Vpu	<i>Viral protein U</i>

INTRODUCTION

I. Les rétrovirus

I. 1. Généralités

Les rétrovirus sont des virus globulaires, enveloppés, de la famille des *Retroviridae*. Leur génome est formé de deux copies d'acide ribonucléique (ARN) monocaténaire, à polarité positive. Au cours du cycle de réplication des rétrovirus, un double brin d'acide désoxyribonucléique (ADN) est synthétisé à partir du génome viral ARN simple brin. Cette réaction, propre à cette famille de virus et appelée « rétro-transcription », est catalysée par une enzyme, la transcriptase inverse (RT, pour *Reverse Transcriptase*), présente au sein de la particule rétrovirale. L'ADN viral peut ensuite être incorporé au génome de la cellule infectée lors de l'intégration, une réaction catalysée par l'enzyme virale intégrase (IN). Les rétrovirus se distinguent des autres virus par le fait que le progénome viral devienne part intégrale du génome de l'hôte.

Les rétrovirus furent découverts au début du XX^{ème} siècle lors de deux études sur les maladies tumorales des poulets. En effet, c'est en 1908 que l'équipe de V. Ellermann et O. Bang montra que les leucémies du poulet étaient d'origine virale et causées par un virus nommé ALV (pour *Avian Leukosis Virus*) (Ellermann & Bang, 1908). Trois ans plus tard, l'équipe de P. Rous mit en évidence une transmission des sarcomes chez les poulets, permettant ainsi la découverte du *Rous Sarcoma Virus* (RSV) (Rous, 1911). Plus tard, ces deux virus spécifiques aux espèces aviaires furent réunis sous une même appellation, *Avian Sarcoma Leukosis Viruses* ou ASLV.

Il fallut encore quelques décennies avant que les rétrovirus oncogènes ne soient décrits chez les mammifères. C'est en 1936 puis 1957 que furent découverts pour la première fois des agents viraux causant des carcinomes et leucémies chez la souris (Bittner, 1936; Gross, 1957). Dans les années qui suivirent, des agents rétroviraux similaires furent identifiés chez les primates, les espèces bovines, les félins, etc.

C'est la découverte de la RT, en 1970, qui permit de valider les hypothèses selon lesquelles une copie d'ADN était générée d'après le génome viral ARN avant d'être intégrée dans le génome de l'hôte (Baltimore, 1970; Temin & Mizutani, 1970).

Au cours des années 1970, les recherches se sont tournées vers l'identification de rétrovirus oncogènes chez l'Homme. Ce n'est qu'en 1977 que la suspicion d'un agent pathogène rétroviral ayant une implication dans les leucémies à cellules T est apparue (Uchiyama *et al.*, 1977). Cet

agent fut décrit en 1980 comme le premier rétrovirus oncogène humain et nommé HTLV-1 (pour *Human T-Lymphotropic Virus type 1*) (Poiesz *et al.*, 1980), suivi deux ans plus tard par l'identification d'HTLV-2, moins pathogène (Kalyanaraman *et al.*, 1982).

À cette même période une épidémie, associée à un syndrome d'immunodéficience acquise (SIDA) et entraînant des maladies opportunistes et des cancers, est apparue dans plusieurs pays développés. Un rétrovirus isolé en 1983 chez des patients atteints de lymphoadénopathies et d'abord appelé LAV (pour *LymphAdenopathy-associated Virus*) fut découvert par F. Barré-Sinoussi, J.-C. Chermann et L. Montagnier (Barre-Sinoussi *et al.*, 1983). Sa capacité à se répliquer dans les cellules du système immunitaire, les lymphocytes T CD4+, et à les détruire permit d'établir des hypothèses sur son rôle dans l'apparition du SIDA. L'année suivante, un rétrovirus nommé alors HTLV-3 (pour *Human T-Lymphotropic Virus type 3*) fut isolé à partir de sérum de patients atteints du SIDA (Gallo *et al.*, 1984). LAV et HTLV-3 furent caractérisés comme étant identiques et, en 1986, le terme « Virus de l'Immunodéficience Humaine de type 1 » (VIH-1) fut adopté par la communauté scientifique. Plus tard cette année-là, l'équipe de L. Montagnier isola un virus nommé VIH-2 (Clavel *et al.*, 1986). VIH-1 et VIH-2 sont, à ce jour, les seuls rétrovirus du genre *Lentivirus* découverts chez l'Homme.

À l'heure actuelle, le SIDA est l'un des principaux problèmes de santé publique dans le monde avec plus de 40 millions de morts depuis le début de l'épidémie. Fin 2018, on estime à 38 millions le nombre de personnes dans le monde vivant avec le VIH et près d'un million de personnes seraient décédées des suites du SIDA dans le courant de l'année (ONUSIDA, fiche d'information 2019, <https://www.unaids.org/fr/resources/fact-sheet>).

I. 2. Classification des rétrovirus

I. 2. 1. Taxonomie des virus

Selon le Comité International sur la Taxonomie des Virus (King *et al.*, 2012), les rétrovirus sont divisés en deux sous-familles et sept genres. Cette classification repose sur les homologies de séquence du gène viral *pol* (**Figure 1**). Les *Spumaretrovirinae* contiennent uniquement le genre *Spumavirus*, dont fait partie le *Prototype Foamy Virus* (PFV), tandis que les *Orthoretrovirinae* regroupent six genres :

- Les *Alpharétrovirus* (dont ASLV),
- Les *Bêtarétrovirus* (dont MMTV, pour *Mouse Mammary Tumor Virus*),

- Les *Gammarrétrovirus* (dont MLV, pour *Murine Leukemia Virus*),
- Les *Deltarétrovirus* (dont HTLV-1),
- Les *Epsilon-rétrovirus*,
- Les *Lentivirus* (dont VIH-1).

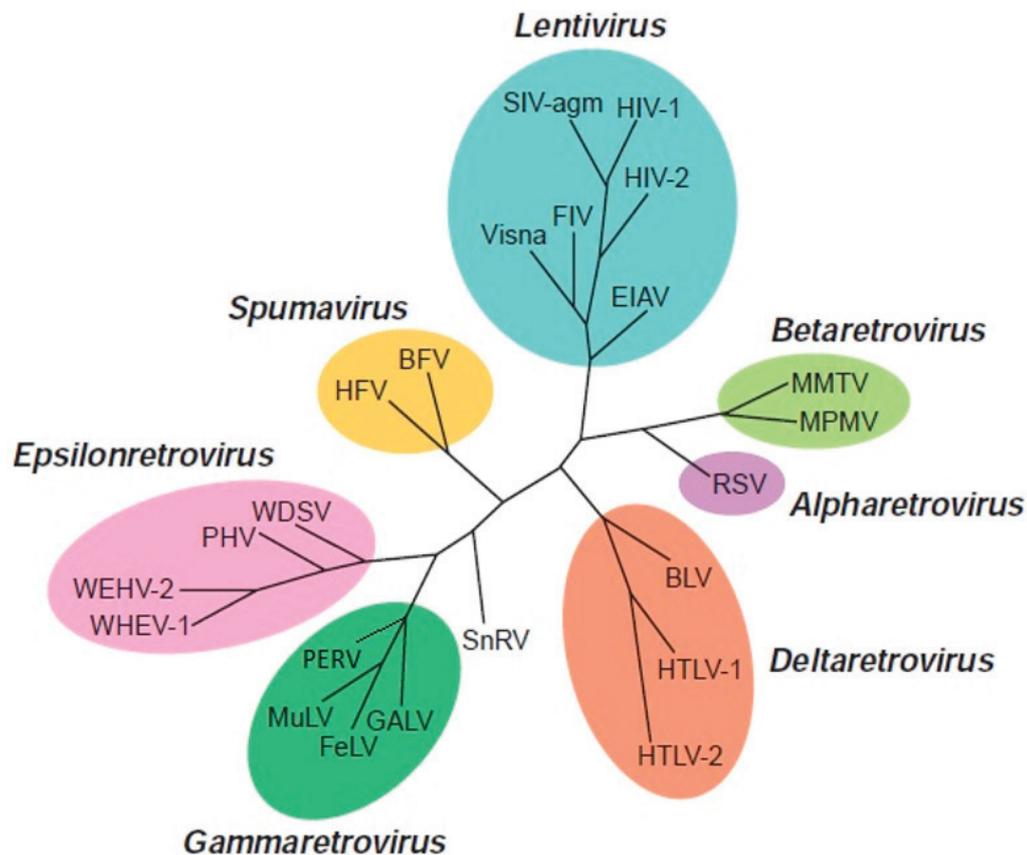


Figure 1 | Arbre phylogénétique des rétrovirus, basé sur les alignements de séquences du gène viral *pol* (King *et al.*, 2012).

I. 2. 2. Virus endogènes et exogènes

Les rétrovirus peuvent également être classifiés selon leur caractère endogène ou exogène. Les rétrovirus exogènes sont naturellement absents de l'organisme. Ils ont besoin de l'infecter pour produire de nouvelles particules infectieuses. Ils se transmettent d'un individu à l'autre de façon horizontale ou bien de façon verticale (pendant l'allaitement par exemple). Contrairement aux rétrovirus exogènes, les rétrovirus endogènes sont intégrés dans le génome des cellules germinales par un processus d'« endogénisation ». Ce sont des vestiges d'infection ancestrale et ils sont transmis à la descendance (**Figure 2**).

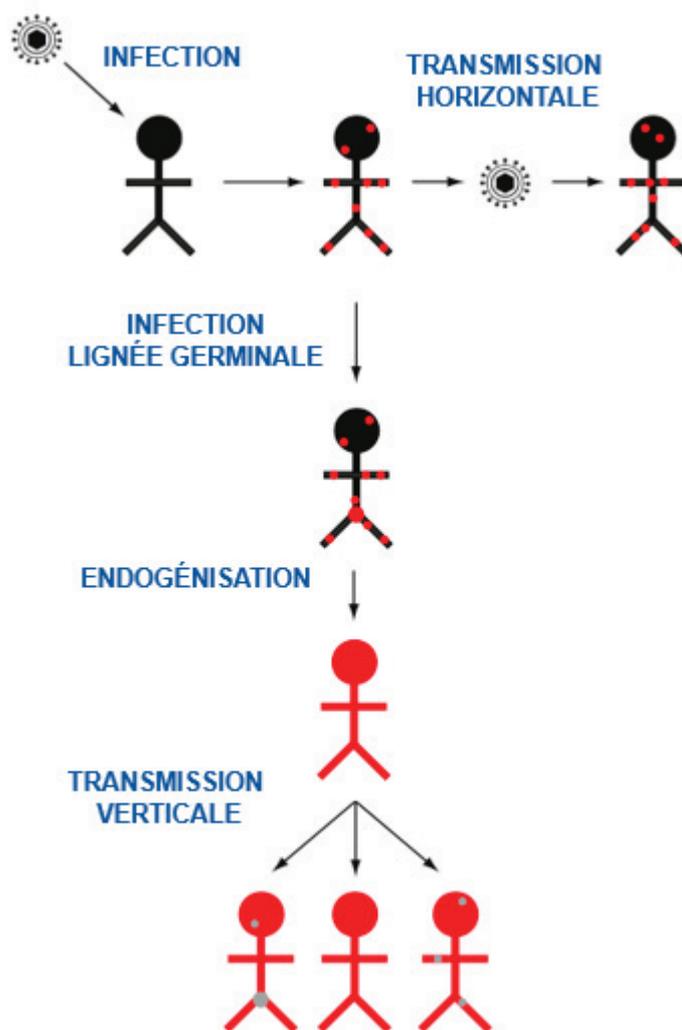


Figure 2 | Principes de transmissions et d'endogénéisation des rétrovirus (adaptée de (Dewannieux *et al.*, 2010)).

Les rétrovirus endogènes humains HERV (pour *Human endogenous retroviruses*) représentent 8% du génome humain (Nelson *et al.*, 2003). Bien que la majorité des rétrovirus endogènes soient non répliquatifs en raison d'un grand nombre de délétions et mutations, certains d'entre eux ont la capacité de former des virions infectieux.

Un autre exemple concerne le rétrovirus endogène porcin PERV (pour *Porcine Endogenous RetroVirus*) qui fait partie intégrante du génome des porcs et qui est présent en proportions différentes selon l'animal, le type de tissu ou le sous-type viral. Cette variabilité montre que PERV est un virus endogène actif pouvant se réintégrer dans le génome porcin. De la même façon que MLV, PERV appartient au genre *Gammarrétrovirus* et, bien qu'il ait conservé sa capacité de répllication, aucune pathologie ne lui a été associée. Néanmoins, il a été montré

qu'une transmission de PERV était possible de cellules porcines à cellules humaines, *in vitro* (Patience *et al.*, 1997). Cette potentialité détermine un risque d'infection pour l'Homme lors de xénotransplantation, c'est-à-dire d'une transplantation de cellules, tissus ou organes d'une espèce à une autre. En effet, la xénotransplantation est étudiée comme une alternative face au manque mondial de donneurs humains. En raison de son élevage aisé, de sa nombreuse progéniture, de la ressemblance de ses organes avec ceux de l'Homme et de sa compatibilité physiologique, le porc semble être le donneur idéal dans le cadre de xéno greffes (Cooper *et al.*, 2018). Cependant, il est important de déterminer les risques que ces transplantations peuvent poser si PERV venait à infecter l'Homme ou à se recombiner avec d'autres rétrovirus déjà présents.

I. 2. 3. Complexité du génome

Un dernier critère de classification repose sur la complexité du génome du virus. On différencie alors les rétrovirus dits « simples », dont le génome est composé uniquement de gènes codant des protéines structurales et enzymatiques, des rétrovirus dits « complexes » qui possèdent des gènes supplémentaires, codants des protéines régulatrices et accessoires.

I. 3. Organisation génomique et structurale

I. 3. 1. La particule rétrovirale

Les virions sont des particules virales matures d'environ 100 nanomètres de diamètre (**Figure 3**). L'enveloppe virale est composée de lipides et de protéines transmembranaires issus de la membrane plasmique de la cellule infectée. Intercalées dans celle-ci se trouvent également deux types de glycoprotéines virales : les glycoprotéines de surface (SU) et les glycoprotéines transmembranaires (TM). Enfin, les protéines de matrice (MA), assemblées sous forme sphérique, tapissent la face interne de l'ensemble.

L'enveloppe renferme la capsid virale, formée par la multimérisation de la protéine CA. La capsid contient différentes enzymes virales (la protéase (PR), l'IN et la RT) ainsi que le génome rétroviral, en deux copies à ARN simple brin. L'ensemble formé par la capsid, le génome et des protéines associées à ce dernier (dites nucléoprotéines) est appelé nucléocapsid (NC).

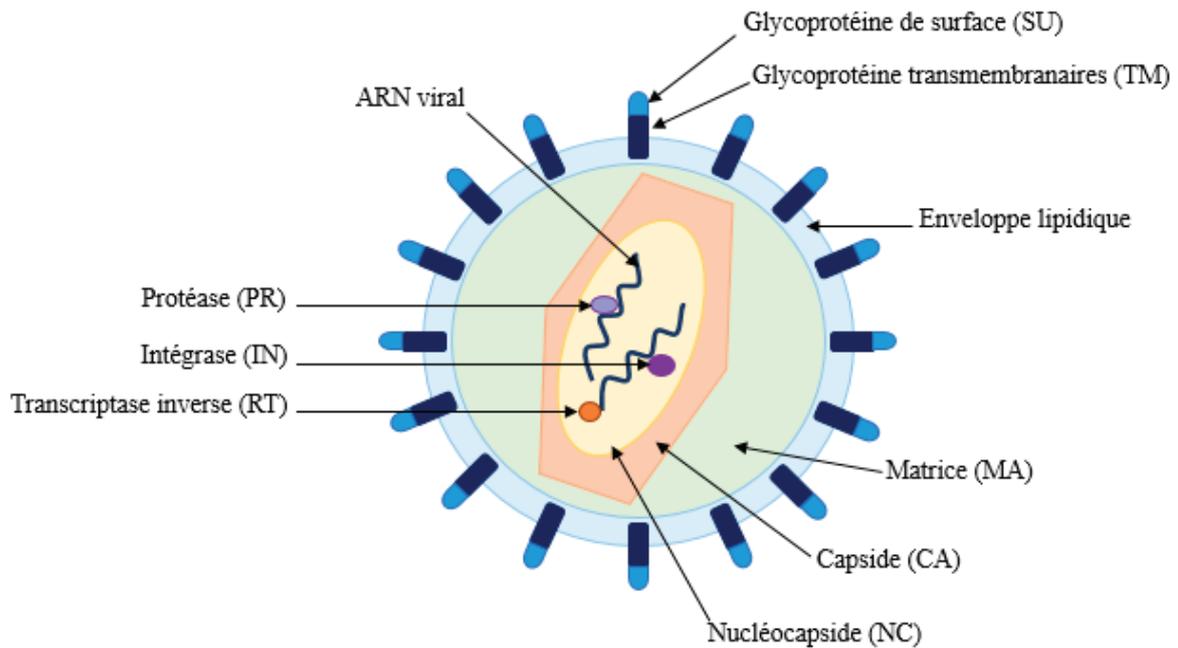


Figure 3 | Organisation structurale d'une particule rétrovirale (adaptée de (Lopata *et al.*, 2018)).

Certains rétrovirus, et notamment les *Lentivirus* et les rétrovirus dits « complexes », comportent en leur sein des protéines supplémentaires qui ne sont pas présentes chez les rétrovirus dits « simples » comme les *Gammarrétrovirus*.

I. 3. 2. Le génome rétroviral

Le génome des rétrovirus est constitué de deux brins d'ARN simple brin à polarité positive coiffés et poly-adenylés, d'une taille allant de 7 à 13 kb. Il est composé de trois gènes communs à tous, *Gag*, *Pol* et *Env*, séparés en deux phases de lecture, l'une regroupant *Gag/Pol* et l'autre *Env* (**Figure 4**). Après transcription inverse, l'ADN viral est intégré dans le génome de la cellule. On parlera alors de provirus.

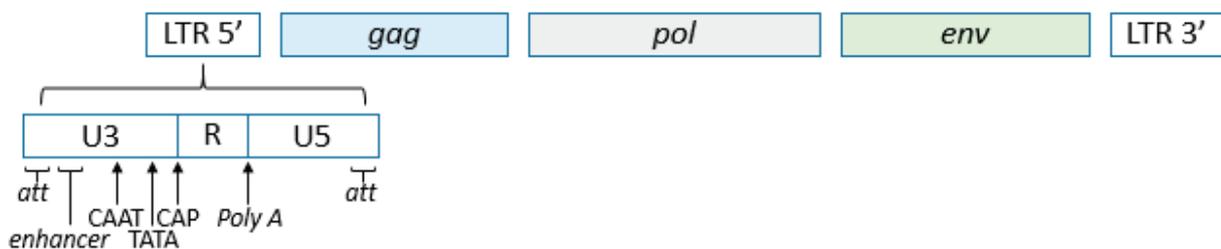


Figure 4 | Représentation de l'organisation génomique des rétrovirus (adaptée de (King *et al.*, 2012)). LTR pour *Long Terminal Repeat* ; att, le site d'attachement ; CAP, le site permettant l'ajout de la coiffe en 5' et Poly A, le signal de polyadénylation.

L'ADN proviral est encadré par deux séquences non codantes identiques, les *Long Terminal Repeat* (LTR), qui sont importantes pour l'intégration du génome dans la cellule hôte et pour sa transcription. Les deux LTR sont constitués de trois régions :

- La région U3, U pour *Unique*, a une taille variable et possède à son extrémité 5' un site d'attachement, nommé *att*, qui est reconnu spécifiquement par l'intégrase. Elle contient également des séquences promotrices, dont la région *enhancer* qui est composée de sites de fixation de facteurs de transcription permettant la régulation du niveau d'expression des gènes proviraux. On trouve également, dans cette région, les boîtes CAAT et TATA en amont du site d'initiation de la transcription. Ces deux séquences servent au recrutement de facteurs de transcription qui guideront la fixation de l'ARN polymérase de type II.

- La région R, pour *Repeat*, intervient lors de l'étape de transcription inverse. Elle contient également le signal de polyadénylation en 3' (*PolyA*) et le site permettant l'ajout de la coiffe en 5' (*CAP*) sur l'ARN messager viral.

- La région U5 possède, tout comme la région U3, la séquence *att* qui est reconnue par l'intégrase. Cette région contient également un site de fixation de l'ARN de transfert qui servira à amorcer la transcription inverse.

La région codante du génome rétroviral est composée de trois gènes :

- Le gène *Gag* (pour *group-specific antigen*) est transcrit sous la forme du précurseur polyprotéique Gag de 55 kDa qui sera secondairement clivé par la protéase virale en trois protéines matures : les protéines de capsid (CA), de nucléocapsid (NC) et de matrice (MA). Ces dernières ont un rôle dans l'organisation structurale du virus (Freed, 1998). Le gène *Gag* code le domaine p6 et deux peptides espaceur SP1 et SP2 chez VIH-1.

- Le gène *Pol* code les enzymes virales : la protéase (PR), la transcriptase inverse (RT) et l'intégrase (IN). Elles sont la résultante du clivage par la protéase virale du précurseur polyprotéique Gag-Pol de 160 kDa (Jacks *et al.*, 1987). Celui-ci est produit lors d'un décalage programmé du cadre de lecture pendant la transcription de *Gag* qui se produit dans 5% des cas et qui permet la régulation de l'expression des enzymes virales (Jacks & Varmus, 1985).

- Le gène *Env*, quant à lui, code un précurseur polyprotéique qui sera clivé par des protéases cellulaires. En résulte la production de deux types de glycoprotéines d'enveloppe : les

glycoprotéines transmembranaires (TM) et de surface (SU), qui sont impliquées dans le mécanisme d'entrée du virus dans la cellule hôte (Vogt, 1997).

I. 3. 3. Les rétrovirus complexes

Les rétrovirus complexes correspondent à tous les virus dont le génome code des protéines additionnelles, dites régulatrices et accessoires. On retrouve parmi ceux-là les *Lentivirus*, les *Deltarétrovirus* et les *Epsilon-rétrovirus*.

Ces protéines supplémentaires permettent aux rétrovirus complexes un contrôle autonome sur des fonctions pour lesquelles les rétrovirus simples dépendent de l'hôte.

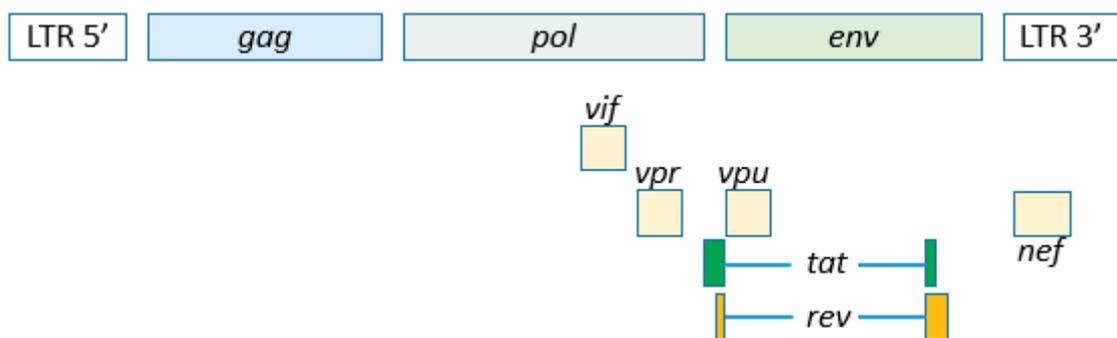


Figure 5 | Représentation de l'organisation génomique des rétrovirus complexes. Exemple du *Lentivirus* VIH-1 et mise en évidence de ses protéines régulatrices et accessoires.

Dans le génome des *Lentivirus*, on retrouve six gènes supplémentaires, codant deux protéines régulatrices Tat (*Transactivator of transcription*) et Rev (*Regulator of expression of virion proteins*), et quatre protéines accessoires : Vif (*Virion infectivity factor*), Vpr (*Viral protein R*), Vpu (*Viral protein U*) et Nef (*Negative regulatory factor*) (**Figure 5**).

Chez VIH-1, la protéine Tat est un transactivateur de la transcription et a une influence sur l'élongation en interagissant avec les extrémités 5' des ARN en cours de synthèse (Gaynor, 1995). Quant à elle, la protéine Rev dirige l'export des ARN messagers viraux non épissés du noyau vers le cytoplasme où ils seront traduits.

Les protéines accessoires de VIH-1 ont chacune une fonction qui leur est propre :

- La protéine Vif contre l'action des protéines du système immunitaire, APOBEC3G et APOBEC3F, qui sont des facteurs de restriction du cycle de réplication rétroviral. En se fixant à ces protéines, Vif provoque leur dégradation et la reprise de la réplication rétrovirale.

- La protéine Vpu est une protéine transmembranaire qui intervient dans le bourgeonnement des nouvelles particules virales.
- La protéine Vpr joue un rôle dans l'importation du complexe de pré-intégration (PIC, pour *pre-integration complex*) dans le noyau de la cellule.
- La protéine Nef amplifie l'infectiosité et la réplication virale, favorisant ainsi la persistance de l'infection (Faust *et al.*, 2017).

Un autre exemple concerne les *Deltarétrovirus* HTLV dont le génome code plusieurs protéines accessoires dont un trans-activateur, nommé Tax (*Review* (Cullen, 1992)). Tax active la transcription de HTLV-1 à travers son interaction avec des facteurs de transcription cellulaire dont l'activité est augmentée à son contact. Cette protéine est également un oncogène qui induit le processus de transformation des cellules infectées (Gessain & Mahieux, 2012).

I. 4. Le cycle répliatif rétroviral

Le cycle répliatif des rétrovirus leur permet de se répliquer afin de former de nouveaux virus infectieux qui pourront se propager et infecter d'autres cellules. Il est divisé en deux étapes communes à tous les rétrovirus : la phase pré-intégrative et la phase post-intégrative (Freed, 2001) (**Figure 6**).

I. 4. 1. La phase pré-intégrative

La première phase de l'infection correspond à l'entrée du rétrovirus dans la cellule hôte. Cette étape implique la reconnaissance de la glycoprotéine SU située à la surface de l'enveloppe virale par des récepteurs membranaires de la cellule. L'interaction spécifique entre la glycoprotéine SU et le récepteur permet de définir le caractère infectieux et le tropisme du virus. Cette interaction entraîne un changement de conformation des glycoprotéines qui vont ensuite diriger la fusion des bicouches lipidiques virale et cellulaire. Ce phénomène va permettre la libération du matériel génique viral encapsidé dans le cytoplasme de l'hôte (Wilén *et al.*, 2012).

Lors de la décapsidation dans le cytoplasme, l'ARN viral monocaténaire est converti en un ADN bicaténaire linéaire par la RT, au cours de la transcription inverse. C'est lors de cette étape que sont générés les LTR, non présents dans leur intégralité dans l'ARN viral. La catalyse de la transcription inverse est entièrement dépendante de deux activités enzymatiques propres à la RT : une activité 'polymérase ADN' pouvant aussi bien se servir d'ARN que d'ADN comme guide ainsi qu'une activité nucléasique (ribonucléase H ou RNase H), spécifique de l'ARN dans

les duplexes ARN-ADN. Cette étape se déroule au sein du complexe de transcription inverse (RTC, pour *Reverse Transcription Complex*), un complexe nucléoprotéique contenant différentes protéines contribuant au bon déroulement de cette étape.

L'ADN néo-synthétisé s'associe à l'IN au sein du PIC. C'est dans ce complexe nucléoprotéique qu'est initiée la réaction d'intégration par l'IN. Il est ensuite transporté au noyau grâce à un mécanisme dépendant du genre viral. En effet, certains rétrovirus comme les *Gammarrétrovirus* infectent seulement les cellules en divisions et le PIC ne pourra entrer dans le noyau que lors de la cassure de la membrane nucléaire (Yamashita & Emerman, 2006). Inversement, les *Lentivirus* peuvent également infecter des cellules qui ne sont pas en division et l'import nucléaire du PIC se fait par transport actif à travers les pores nucléaires.

Enfin, l'ADN viral est incorporé de façon stable au sein du génome de la cellule hôte lors de l'étape d'intégration. Le génome viral ainsi intégré est appelé provirus. L'étape d'intégration sera détaillée dans la suite de ce manuscrit.

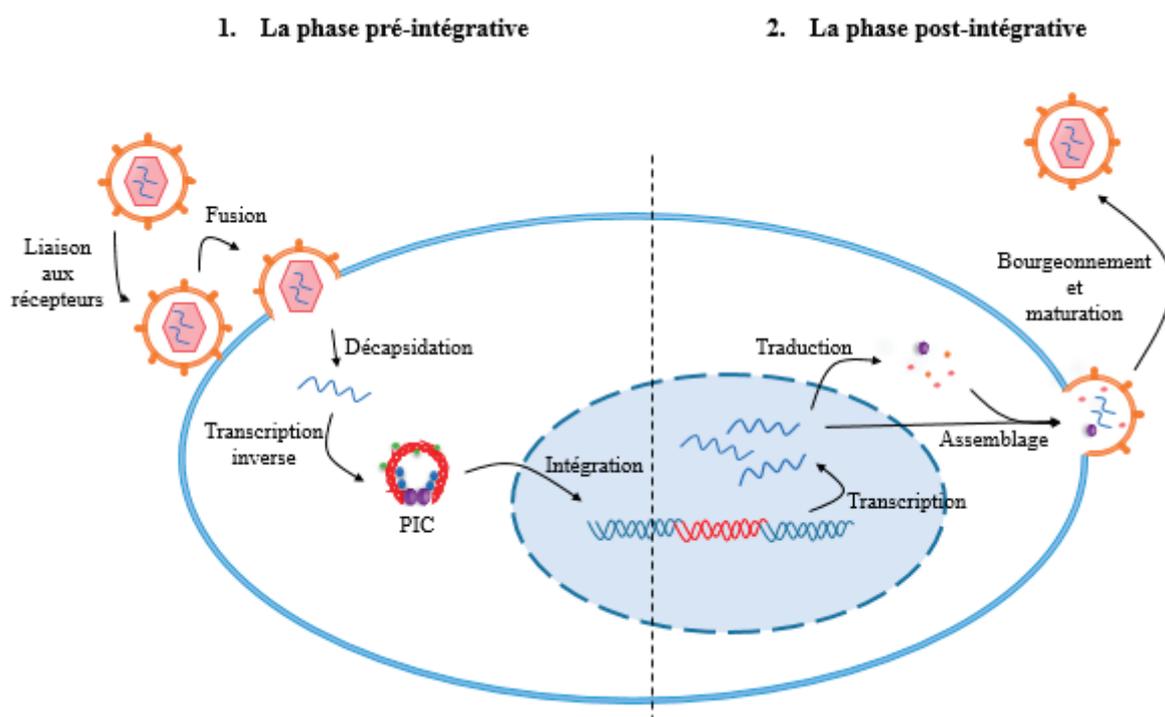


Figure 6 | Le cycle répliatif rétroviral.

I. 4. 2. La phase post-intégrative

Après son intégration dans le génome cellulaire, le provirus sert de matrice pour la synthèse des ARN messagers (ARNm) codant les protéines virales. C'est l'ARN polymérase de type II cellulaire qui exécute la transcription rétrovirale. Une fois transcrit, l'ARNm viral est sujet à l'ajout d'une coiffe en 5', d'une polyadénylation en 3' et, pour une partie des molécules, d'un épissage permettant la formation de plusieurs ARN matures.

Ces ARNm sont exportés hors du noyau et traduits par la machinerie cellulaire. Alors que le précurseur Gag n'est pas clivé, le domaine MA participe à l'adressage de la polyprotéine à la membrane plasmique de la cellule. Le domaine CA va alors entraîner la multimérisation de Gag pendant l'assemblage alors que le domaine NC recrute deux brins d'ARN viral.

Les particules virales ainsi assemblées bourgeonnent alors pour former de nouveaux virions immatures (Sundquist & Krausslich, 2012) qui deviendront infectieux lors de l'étape de maturation, après clivage des polyprotéines Gag et Gag-Pol par la protéase virale (Freed, 2015) (**Figure 7**).

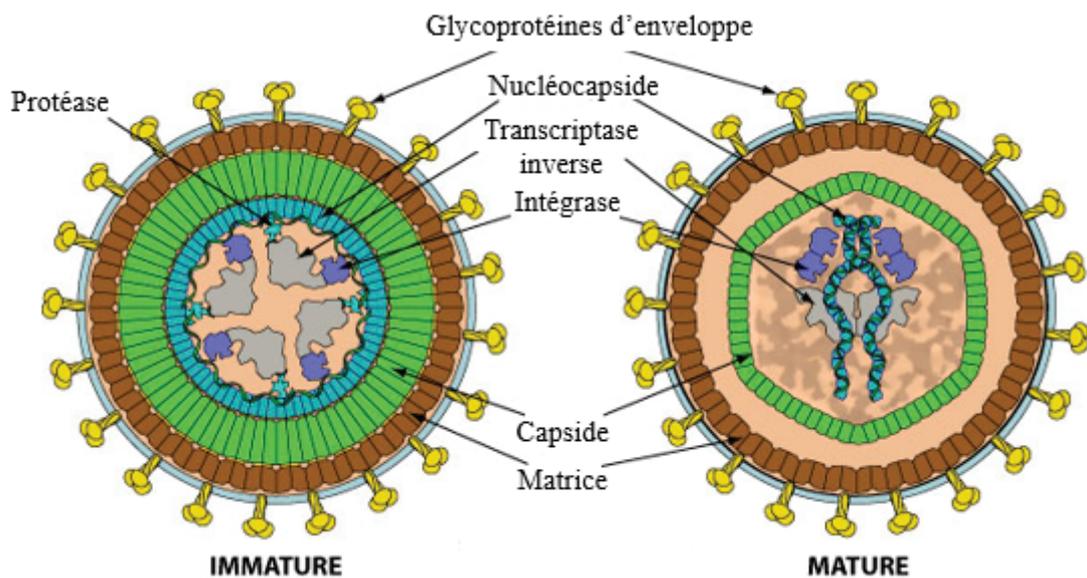


Figure 7 | Schématisation des particules virales immature et mature d'un *Gammarétrovirus* (ViralZone, Swiss Institute of bioinformatics, 2009, <https://www.unaids.org/fr/resources/factsheet>).

II. L'intégrase

II. 1. Fonctions de l'intégrase

II. 1. 1. Complexe de pré-intégration

La molécule d'ADN viral fraîchement rétro-transcrite existe sous forme d'un complexe de pré-intégration qui contient à la fois des protéines virales et des protéines cellulaires (Miller *et al.*, 1997; Raghavendra *et al.*, 2010) l'aidant à accomplir ou réguler différentes fonctions.

Outre l'IN et l'ADN viral, le PIC peut contenir d'autres protéines virales telles que les protéines RT, CA, NC et MA mais aussi des cofacteurs cellulaires dont les fonctions initiales dans la cellule sont détournées par le virus. Dans le cas du VIH-1, il a notamment été montré que son intégrase interagissait spécifiquement avec les protéines INI1 (pour *Integrase Interactor 1*) ou encore LEDGF/p75 (pour *Lens Epithelium-Derived Growth Factor*) via leurs domaines IBD (pour *Integrase Binding Domain*) respectifs (Maillot *et al.*, 2013). Ces protéines associées vont avoir un rôle de régulateurs positifs ou négatifs de l'activité de l'IN. La protéine INI1 est un activateur de transcription, composant central du complexe de remodelage de la chromatine SWI/SNF, et fut le premier partenaire cellulaire de l'IN à être caractérisé (Kalpana *et al.*, 1994). Bien qu'elle ait un rôle qui reste controversé dans le cycle de réplication du VIH-1, il semblerait qu'elle agisse à la fois lors des étapes de pré-intégration et de post-intégration. Son rôle dans le choix du site d'intégration a été proposé en raison de la capacité du complexe SWI/SNF à restaurer l'intégration *in vitro* par remodelage de la chromatine (Lesbats *et al.*, 2011). Les données structurales également obtenues permettent de penser qu'INI1 stabilise l'IN, ceci l'empêchant de catalyser des réactions d'intégration non spécifiques ou d'auto-intégration pendant le transfert du PIC jusqu'au site d'intégration de l'ADN viral (**Figure 8**) (Maillot *et al.*, 2013). LEDGF/p75 est, quant à lui, le partenaire de l'IN du VIH-1 le plus étudié et le principal cofacteur de l'intégration du VIH-1. Il a un rôle essentiel lors du choix du site d'intégration de l'ADN viral en guidant l'IN vers les régions où la chromatine est transcriptionnellement active (Serrao & Engelman, 2016) (développé dans **Introduction §II.3**).

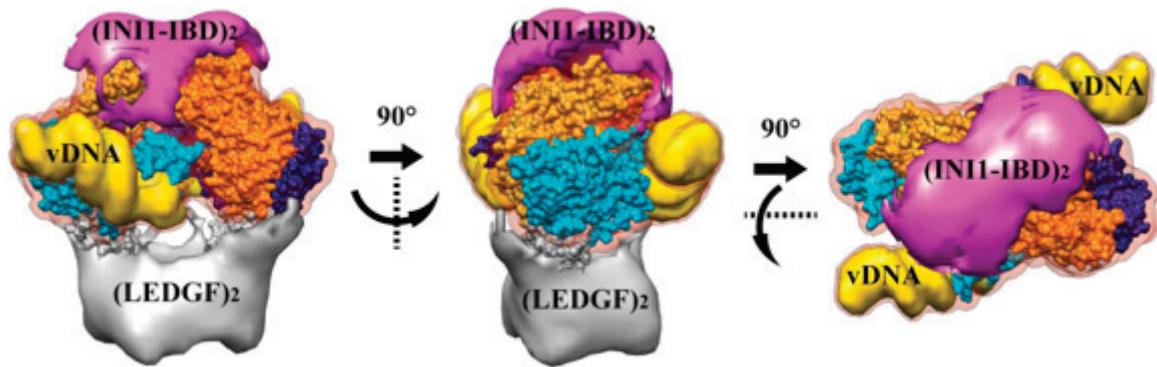


Figure 8 | Représentation 3D du complexe formé par l'IN du VIH-1, LEDGF/p75, l'IBD d'INI1 et l'ADN viral, obtenue par cryo-microscopie électronique (Maillot *et al.*, 2013).

Le transfert du PIC du cytoplasme au noyau de la cellule se déroule de manière distincte selon le genre rétroviral. Pour les *Gammarétrovirus*, l'infection se fait lors de la division cellulaire ; ainsi l'import du PIC dans le noyau a lieu lors de la rupture de la membrane nucléaire. À l'inverse, les *Lentivirus* infectent des cellules quiescentes et le PIC peut franchir la membrane nucléaire au travers des pores nucléaires. Cet import est possible grâce aux signaux de localisation nucléaire présents chez certaines des protéines composants le PIC. Il s'agit des protéines IN, MA et Vpr qui reconnaissent ainsi les importines cellulaires, protéines impliquées dans le transport actif de molécules entre le cytoplasme et le noyau (Review (Suzuki & Craigie, 2007)).

Selon la localisation de l'intégration dans le génome hôte, l'environnement chromatinien du provirus peut entraîner l'activation virale ou l'extinction transcriptionnelle. Pour cette raison, le choix du site d'intégration a un impact sur le niveau de réplication viral et peut contribuer à la mise en place de réservoirs viraux latents (Maldarelli *et al.*, 2014). L'alignement des sites d'intégration a montré un consensus de séquences palindromiques spécifique à chaque virus et se trouvant à quelques paires de bases des sites d'intégration (Serrao *et al.*, 2015). Les différents rétrovirus ont des préférences pour des caractéristiques du génome cellulaire qui leur sont propres (**Figure 9**) (Review (Kvaratskhelia *et al.*, 2014)). Ainsi, HIV-1 et les *Lentivirus* ont une préférence pour les unités de transcription et plus particulièrement les gènes fortement exprimés et les zones riches en introns (Mitchell *et al.*, 2004), alors que MLV et les *Gammarétrovirus* privilégient les régions promotrices et les îlots CpG (Moalic *et al.*, 2006). Les *Spumavirus*, comme PFV, défavorisent les gènes et les sites de transcription active (Nowrouzi *et al.*, 2006). Enfin, les *Alpharétrovirus* et *Bêtarétrovirus* ont une intégration aléatoire (Mitchell *et al.*, 2004).

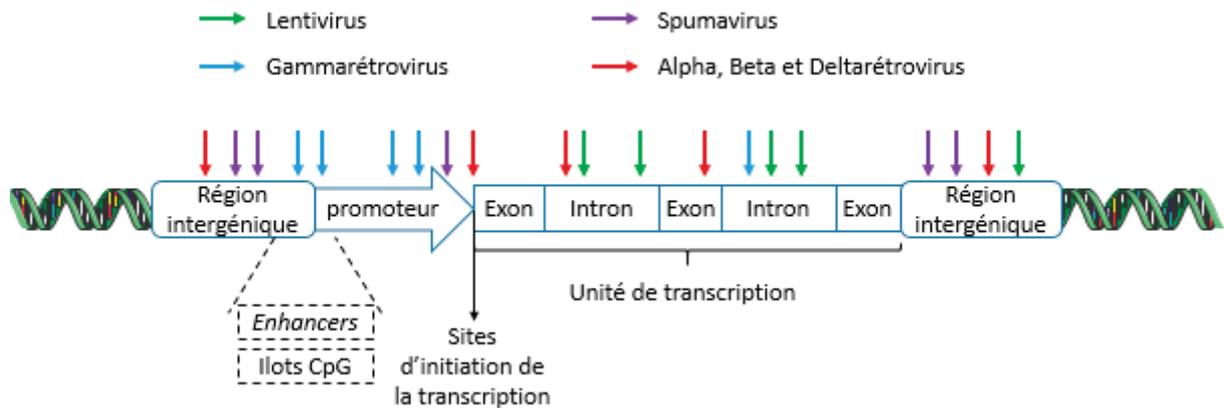


Figure 9 | Représentation des sites d'intégrations préférentiels pour les différents genres rétroviraux (adaptée de (Serrao & Engelman, 2016)).

II. 1. 2. Intégration concertée

Lors du cycle réplcatif rétroviral, l'IN joue un rôle majeur en catalysant la réaction d'intégration et en assurant l'association stable entre l'ADN viral et l'ADN hôte (Grandgenett *et al.*, 2015). Cette réaction est dite « concertée » car les deux extrémités de l'ADN viral s'insèrent en un même site de l'ADN cellulaire.

Cette réaction se fait au sein d'un complexe nucléoprotéique composé de l'IN, de l'ADN viral et, plus tardivement, de l'ADN cible. Il est communément appelé intasome. Si la réaction d'intégration en elle-même est maintenant comprise, des questions demeurent sur la participation d'éventuels partenaires cellulaires ainsi que sur l'état oligomérique de l'IN.

L'intégration concertée est divisée en trois étapes successives : le clivage des extrémités de l'ADN viral ou « 3'-processing », le transfert de brins entre ADN viral et ADN cellulaire 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 cellulaires pour la réparation de l'ADN (**Figure 10**).

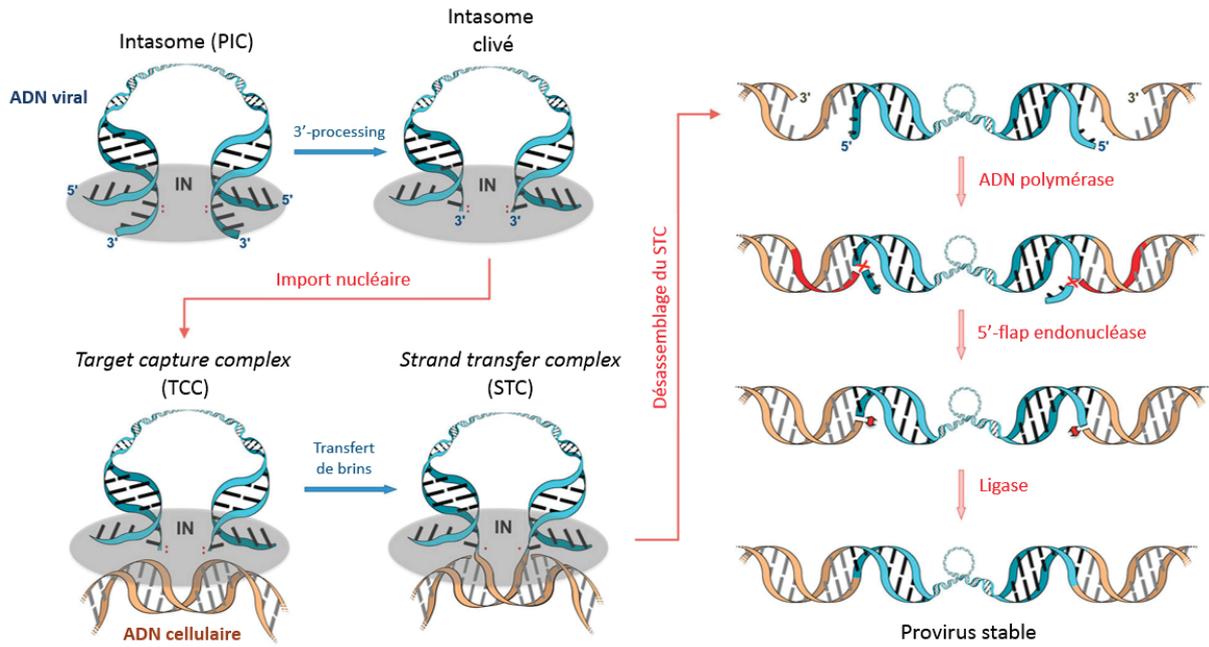


Figure 10 | Schématisation des étapes de l'intégration rétrovirale (Lesbats *et al.*, 2016).

Lors du « 3'-processing », l'IN catalyse le clivage simultané de deux nucléotides terminaux aux extrémités 3' de l'ADN double brin viral (**Figure 10**, première étape). Il faut donc que la synthèse de l'ADN viral par la RT soit achevée pour que cette étape puisse être réalisée. L'IN va alors former un complexe avec l'ADN viral, le SSC (pour *Stable Synaptic Complex*). En présence de cofacteur métallique, Mg^{2+} ou Mn^{2+} , l'IN hydrolyse une liaison phosphodiester à chaque extrémité de l'ADN viral, libérant ainsi des groupes hydroxyles en 3' attachés à un dinucléotide 5'-CA-3' invariant. Cette coupure aboutit ainsi à la création de deux extrémités CA-3'-OH conservées et le SSC devient alors le CDC (pour *Clived Donor Complex*).

Après import nucléaire, l'intasome lie l'ADN cible, formant ainsi le TCC (pour *Target Capture Complex*). Les deux extrémités CA-3'-OH générées se liguent à l'ADN cellulaire lors du transfert de brins. Pour cela, l'IN utilise les extrémités 3' de l'ADN viral pour réaliser une attaque nucléophile sur l'ADN cellulaire. On parle alors enfin de STC, pour *Strand Transfer Complex* (**Figure 10**, troisième étape). Il en résulte une brèche de quelques nucléotides de part et d'autre de l'ADN viral ainsi qu'un petit surplomb en 5'.

Des enzymes cellulaires réparent les brèches et reforment les liaisons phosphodiester pour permettre l'insertion définitive du génome viral (Skalka & Katz, 2005). Trois activités enzymatiques sont nécessaires pour la réparation de l'ADN : ADN polymérase, 5' « flap » endonucléase et ligase. Les enzymes cellulaires, ADN polymérase β , AP-endonucléase 1 et

ligase I, venant du mécanisme de réparation par excision de base de la cellule, ont ce rôle (Yoder & Bushman, 2000).

Une fois cette étape achevée, la cellule devient porteur permanent du génome viral, alors appelé provirus. Celui-ci est flanqué d'une courte duplication de base de la séquence de l'ADN cible, dont la longueur est dépendante du virus. Elle peut aller de quatre paires de bases pour MLV et PFV, à six paires de bases pour HTLV et HIV-1 (Vincent *et al.*, 1990; Kim *et al.*, 2010); *Review* : (Serrao & Engelman, 2016).

II. 1. 3. Autres fonctions de l'intégrase

Outre sa fonction principale dans l'intégration, l'IN est impliquée dans d'autres étapes du cycle réplcatif telles que la transcription inverse, l'import nucléaire de l'ADN viral, l'assemblage et la maturation des virions (Masuda, 2011). La plupart de ces fonctions secondaires ont été mises en évidence par la génération de mutants dits de « type 2 » qui, contrairement aux mutants de « type 1 », n'ont pas d'influences sur l'activité enzymatique de l'IN.

Tout d'abord, l'IN joue un rôle lors de l'étape de transcription inverse. Des études ont montré une stimulation de l'activité de la RT résultant de son interaction avec l'IN (Dobard *et al.*, 2007). En effet, six mutants uniques au niveau du domaine carboxy-terminal (CTD, pour *Carboxy-Terminal Domain*) de l'IN du VIH-1 sont responsables d'une diminution de l'activité de la RT tout en conservant celle de l'IN (Tekeste *et al.*, 2015). Cependant, le rôle précis de l'IN lors de cette étape n'est pas entièrement compris.

De même, il a été montré l'importance de l'IN dans l'import nucléaire, notamment lors du transport actif dans les cellules quiescentes. Une étude a en effet mis en évidence que des mutants réduisant la reconnaissance de l'ADN cellulaire par l'IN affectent l'import du matériel génétique dans le noyau de la cellule hôte (Ikeda *et al.*, 2004). *In vivo*, il a également été montré que l'IN du VIH-1 interagit spécifiquement avec l'importine 7 de l'hôte (Ao *et al.*, 2007), ceci soutenant son rôle dans le transport de l'ADN viral à travers la membrane nucléaire.

D'autres études menées sur des mutants de l'IN du VIH-1 ont également montré son implication dans les étapes d'assemblage et de relargage des virions. Le mécanisme par lequel l'IN influence ces étapes n'est cependant pas encore compris (Engelman, 2011).

Enfin, l'IN interagit spécifiquement avec l'ARN viral, jouant ainsi un rôle essentiel dans le processus de morphogénèse viral et notamment pour l'incorporation de l'ARN dans la capside. Des études utilisant des inhibiteurs allostériques de l'IN (ALLINI, pour *Allosteric Integrase Inhibitor*) ou des mutants de l'IN ont mis en évidence la formation de virions excentriques et non infectieux lorsque l'interaction de l'IN avec l'ARN est altérée (Fontana *et al.*, 2015; Kessl *et al.*, 2016). Un agrégat formé entre autres de l'ARN et des protéines NC est alors observé en dehors de la capside (**Figure 11**).

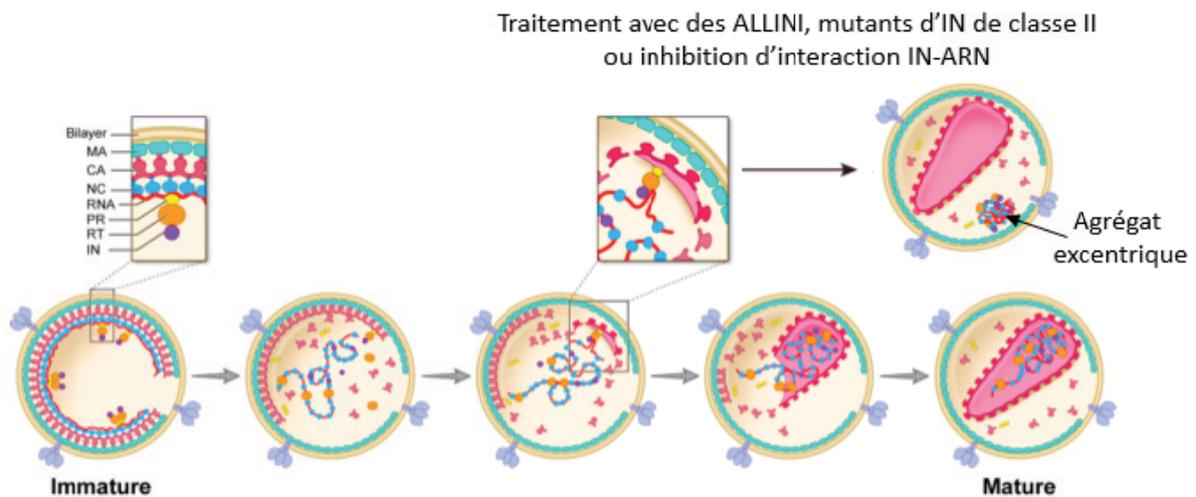


Figure 11 | Modèle de la maturation d'un virion et rôle de l'IN dans l'encapsidation de l'ARN viral (Fontana *et al.*, 2015).

II. 2. Structure de l'intégrase

L'intégrase rétrovirale est composée d'environ 300 résidus d'acides aminés. Par exemple, l'IN de VIH-1 comprend 288 résidus et celle du *Gammarétrovirus* PERV-A/C possède 405 résidus d'acides aminés. Elle comporte trois domaines fonctionnels : un domaine amino-terminal, un domaine central catalytique et un domaine carboxy-terminal, reliés par des boucles flexibles (**Figure 12**).

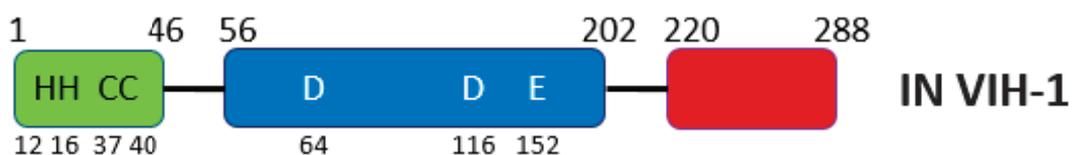


Figure 12 | Organisation schématique de l'IN du VIH-1. Le domaine amino-terminal NTD est représenté en vert, le domaine central catalytique CCD en bleu, le domaine carboxy-terminal CTD en rouge.

À l'échelle de la protéine entière, la structure primaire des IN des différents rétrovirus est relativement peu conservée. En effet, l'identité de séquence de l'IN de PERV-A/C avec les IN de VIH-1 et PFV est respectivement de 18% et 26% (Demange *et al.*, 2015). Le domaine carboxy-terminal présente la structure primaire la moins conservée chez les IN rétrovirales.

Les boucles flexibles reliant les trois domaines de l'IN rendent difficile l'étape de cristallogénèse de la protéine entière. Afin d'obtenir des informations structurales, la première stratégie adoptée fut d'étudier séparément les domaines de cette protéine modulaire.

II. 2. 1. Le domaine amino-terminal

Le domaine amino-terminal (NTD, pour *amiNo-Terminal Domain*) a vu sa structure, dimérique en solution, être tout d'abord résolue par résonance magnétique nucléaire (RMN) (**Figure 13**) (Cai *et al.*, 1997). Il est formé de quatre hélices α et son repliement est conservé parmi les IN rétrovirales. Ce domaine lie l'ADN viral par l'intermédiaire d'un motif hélice-tour-hélice, commun aux protéines liant les acides nucléiques. Il contient également une séquence conservée composée de deux résidus d'histidine et deux résidus de cystéine : le motif HHCC (Eijkelenboom *et al.*, 1997). Ce dernier, proche d'un domaine à doigt de zinc, établit des liaisons de coordination avec un ion Zn^{2+} qui stabilisent sa structure.

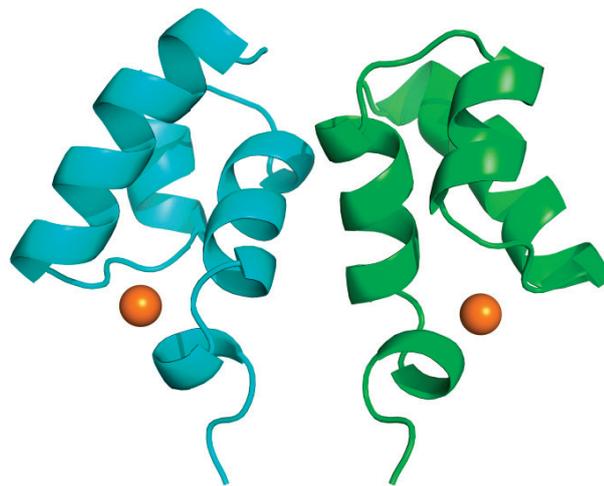


Figure 13 | Structure du NTD de l'IN du VIH-1 sous forme dimérique résolue par RMN (code PDB : 1WJC). En orange, les ions Zn^{2+} .

Le NTD n'a pas de rôle direct dans l'activité enzymatique de l'IN (Bushman & Wang, 1994). Cependant, il est impliqué dans l'oligomérisation de l'IN, élément essentiel pour la réaction d'intégration.

Chez certains rétrovirus, un sous-domaine nommé NED (pour *N-terminal Extension Domain*) peut être présent dans la structure des IN rétrovirales, en amont du NTD. Ce sous-domaine protéique additionnel d'environ quarante acides aminés est retrouvé dans l'IN des *Gammarrétrovirus* et des *Spumavirus*. Il comporte une hélice α et un feuillet de trois brins β antiparallèles (Maertens *et al.*, 2010).

II. 2. 2. Le domaine central catalytique

Le CCD (pour *Catalytic Core Domain*), cœur catalytique de l'enzyme, contient la triade catalytique. Elle est formée de deux résidus d'acide aspartique et d'un résidu d'acide glutamique -D, D(35), E- qui sont essentiels aux activités de clivage et de transfert de brins de l'IN (Kulkosky *et al.*, 1992). De par son repliement, le CCD permet le rapprochement de ces trois résidus d'acides aminés. Ce domaine catalytique présente un repliement caractéristique des membres de la superfamille des intégrases rétrovirales, de type RNase H. Celui-ci se compose d'un feuillet β à cinq brins et de six hélices α . Le CCD peut lier des cations divalents métalliques, dont Mg^{2+} , qui sont essentiels à l'activité enzymatique de l'IN.

En absence des autres domaines de l'IN, le CCD garde sa capacité à réaliser les réactions enzymatiques telles que le 3'-*processing*, mais également de reconnaître les extrémités de l'ADN viral et de catalyser la réaction de désintégration *in vitro*.

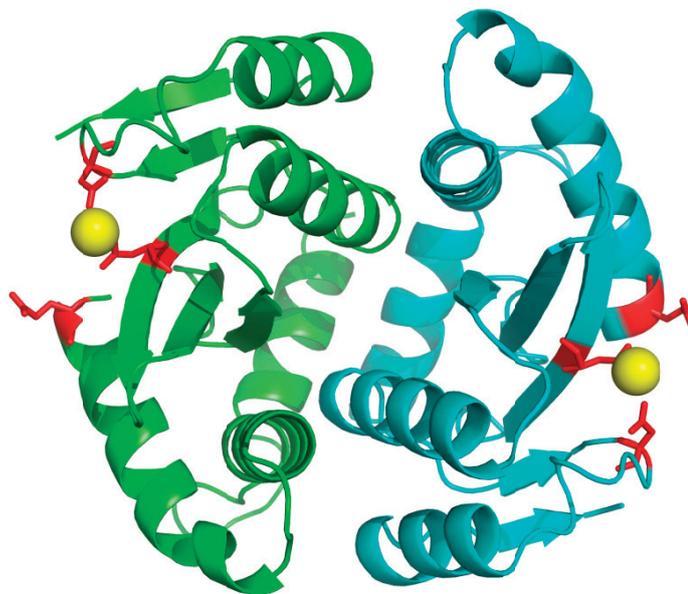


Figure 14 | Structure cristalline du CCD de l'IN du VIH-1 sous forme dimérique (code PDB : 1BL3). En rouge, la triade catalytique D,D(35),E ; en jaune, les ions Mg^{2+} .

Plus d'une trentaine d'études par cristallographie aux rayons X ont été menées sur le CCD des IN de VIH-1, ASLV et PFV. Pour le VIH-1, celles-ci ont pu être réalisées grâce à la mutation F185K sur le CCD qui augmente de façon significative la solubilité du domaine sans nuire à l'activité de la protéine (Jenkins *et al.*, 1996). Ces structures cristallines ont permis d'étudier le CCD sous sa forme *apo*, en complexe avec des cations divalents Mg^{2+} et Mn^{2+} , ou avec des molécules inhibitrices. De plus, ces structures ont mis en évidence une formation dimérique du CCD faisant intervenir des interactions entre des tandems d'hélices α de chaque monomère (**Figure 14**) (Maignan *et al.*, 1998).

II. 2. 3. Le domaine carboxy-terminal

Le CTD est le domaine le moins conservé chez les IN rétrovirales. Sa structure a été obtenue par RMN avec l'IN du VIH-1 et montre la formation d'un dimère en solution (**Figure 15**) (Lodi *et al.*, 1995). D'un point de vue structural, ce domaine est riche en brins β organisés en tonneau, semblables aux domaines SH3 retrouvés notamment dans des protéines liant des acides nucléiques (Jaskolski *et al.*, 2009). La boucle flexible reliant le CTD au CCD n'est pas conservée et a une longueur variable selon le genre rétroviral.

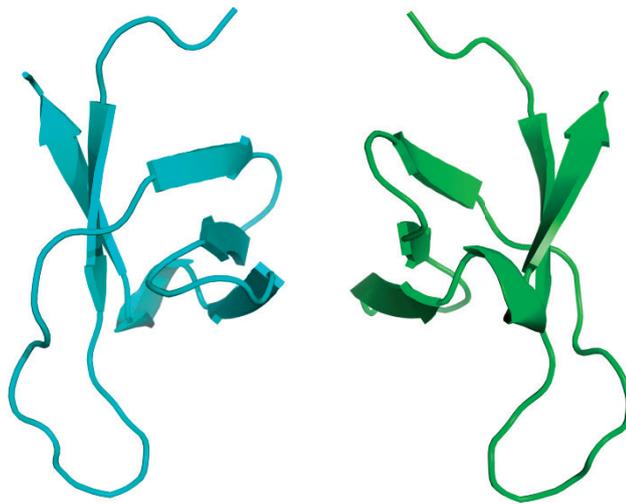


Figure 15 | Structure du CTD de l'IN du VIH-1 résolue par RMN (code PDB : 1IHV).

Le CTD est, d'une manière générale, impliqué dans la fixation de l'ADN viral et de l'ADN cellulaire, dans l'oligomérisation fonctionnelle de l'IN entière et dans l'assemblage de l'intasome (Maertens *et al.*, 2010).

II. 2. 4. Structures connues

La structure d'un fragment de deux domaines NTD-CCD de l'IN du VIH-1 a été résolue au début des années 2000 (Wang *et al.*, 2001). Cette structure cristalline montre un arrangement dimérique, avec une interface au niveau des CCD similaire à celle observée pour ce domaine isolé. La boucle reliant le NTD au CCD est flexible et n'est pas observée dans la structure, tandis que deux dimères NTD-CCD interagissent dans le réseau cristallin (**Figure 16A**).

La structure cristalline du fragment NTD-CCD de l'IN du VIH-2 a été obtenue en 2009 (**Figure 16B**) (Hare *et al.*, 2009). Celle-ci a été résolue en présence du domaine IBD de LEDGF/p75, qui interagit avec un dimère du fragment d'IN. Ici, le NTD et le CCD forment une structure plus compacte que celle observée précédemment, mettant en évidence la modularité de l'IN.

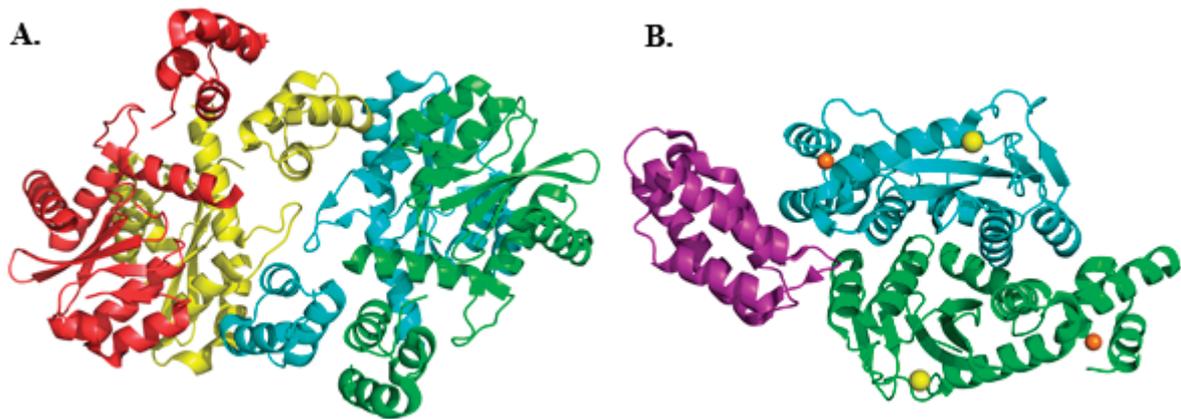


Figure 16 | **A.** Structure cristalline du fragment NTD-CCD de l'IN du VIH-1 (code PDB : 1K6Y). Les sous-unités du premier dimère sont représentées en jaune et rouge ; les sous-unités du deuxième dimère, en vert et bleu. **B.** Structure cristalline du fragment NTD-CCD de l'IN du VIH-2 sous forme dimérique (en vert et bleu) en complexe avec le domaine IBD de LEDGF/p75 (en violet) (code PDB : 3F9K). Mg²⁺ est représenté par des sphères jaunes et Zn²⁺ par des sphères orange.

Concernant la structure du fragment à deux domaines CCD-CTD, celle-ci a été résolue simultanément pour l'IN du VIH-1 (**Figure 17A**) (Chen *et al.*, 2000) et pour l'IN de RSV (**Figure 17B**) (Yang *et al.*, 2000).

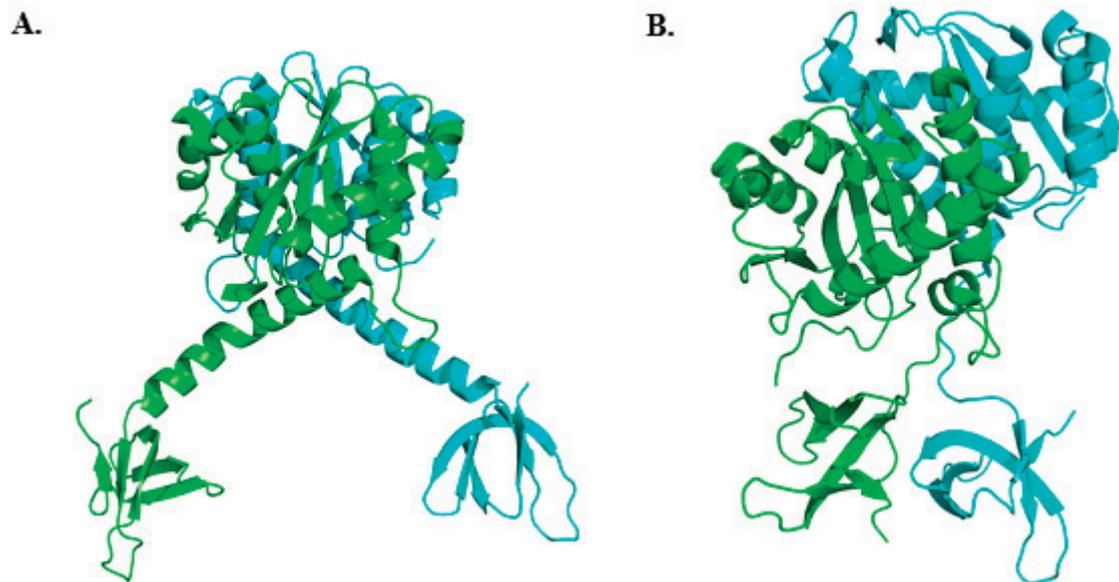


Figure 17 | **A.** Structure cristalline du fragment CCD-CTD de l'IN du VIH-1 sous forme dimérique (code PDB : 1EX4). **B.** Structure cristalline du fragment CCD-CTD de l'IN de RSV sous forme dimérique (code PDB : 1C0M).

Bien que l'organisation dimérique des CCD soit de nouveau conservée dans les deux structures cristallines, on retrouve des différences non négligeables au niveau des CTD. En effet, la boucle flexible séparant le CTD du CCD de l'IN du VIH-1 contient une dizaine d'acides aminés de plus que celle de l'IN de RSV et permet la formation d'une hélice α qui entraîne un espacement des deux CTD. Chez l'IN de RSV, la boucle flexible, plus courte, permet le rapprochement des CTD qui interagissent sous forme dimérique.

II. 2. 5. Intasomes et états oligomériques

La première structure d'intasome déterminée à l'échelle atomique a été publiée en 2010. Elle concerne l'intasome du *Spumavirus* PFV obtenue par cristallographie aux rayons X (Maertens *et al.*, 2010).

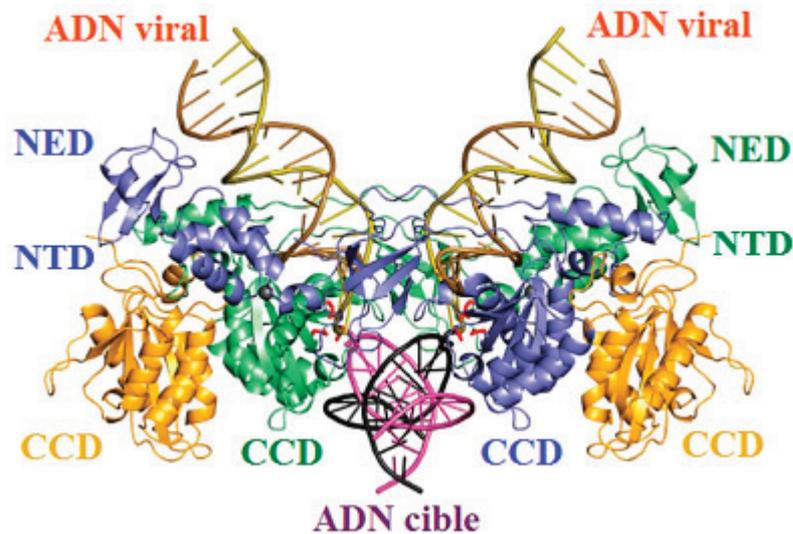
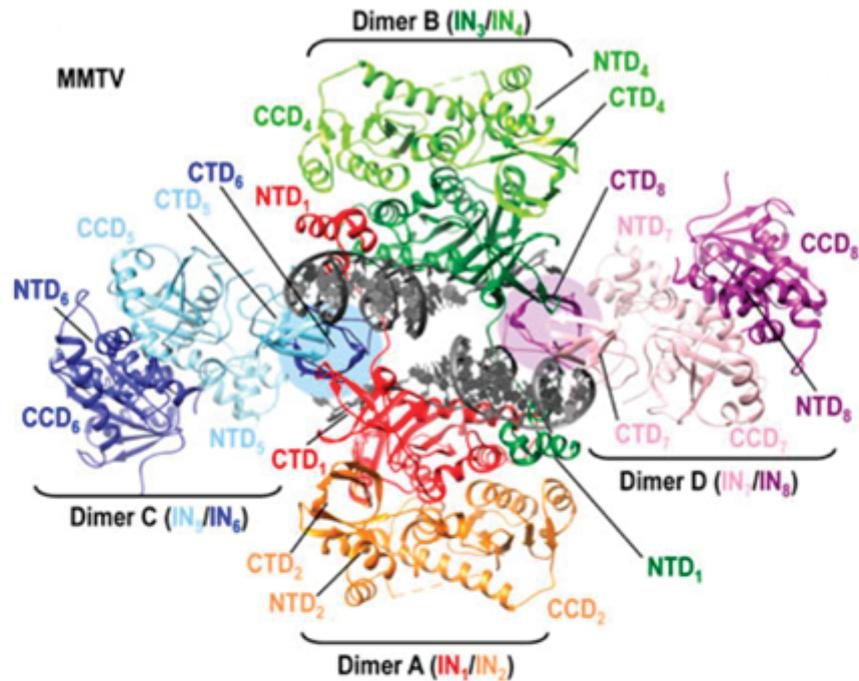


Figure 18 | Structure cristalline de l'intasome de PFV avec son ADN cible (code PDB : 3OS2). L'intégrase forme un tétramère. Un premier dimère est coloré en vert et jaune ; le deuxième, en bleu et jaune. Alors que les monomères internes (bleu et vert) sont entièrement visibles dans les cartes de densité électronique, seul le domaine CCD des monomères externes (en jaune) est observé, les domaines restants étant trop flexibles.

L'intasome de PFV est formé d'un tétramère d'intégrase avec un assemblage en dimère de dimères (**Figure 18**). Les sous-unités internes de chaque dimère sont responsables des interactions avec l'ADN viral et de la position des sites actifs pour la catalyse. Elles interagissent entre elles en formant des contacts NTD-CCD. Les deux sous-unités externes interagissent avec leurs sous-unités internes respectives grâce à la dimérisation des CCD. Cependant, les structures des NED, NTD et CTD de ces monomères n'ont pu être observées en densité électronique en raison de leur flexibilité. Des études par diffusion des rayons X aux petits angles (SAXS, pour *Small-Angle X-ray Scattering*) ont permis de proposer une position pour ces domaines de chaque côté de l'intasome (Gupta *et al.*, 2012). La résolution de la structure de l'intasome avec l'ADN cible montre que ce dernier se positionne dans la crevasse formée entre les CCD des sous-unités internes (**Figure 18**).

Par la suite, les structures de deux nouveaux intasomes ont été résolues par cryo-microscopie électronique et cristallographie aux rayons X. Il s'agit de celles du *Bêtarétrovirus* MMTV (Ballandras-Colas *et al.*, 2016) et de l'*Alpharétrovirus* RSV (Yin *et al.*, 2016), qui montrent toutes deux la présence de huit molécules d'IN.

A.



B.

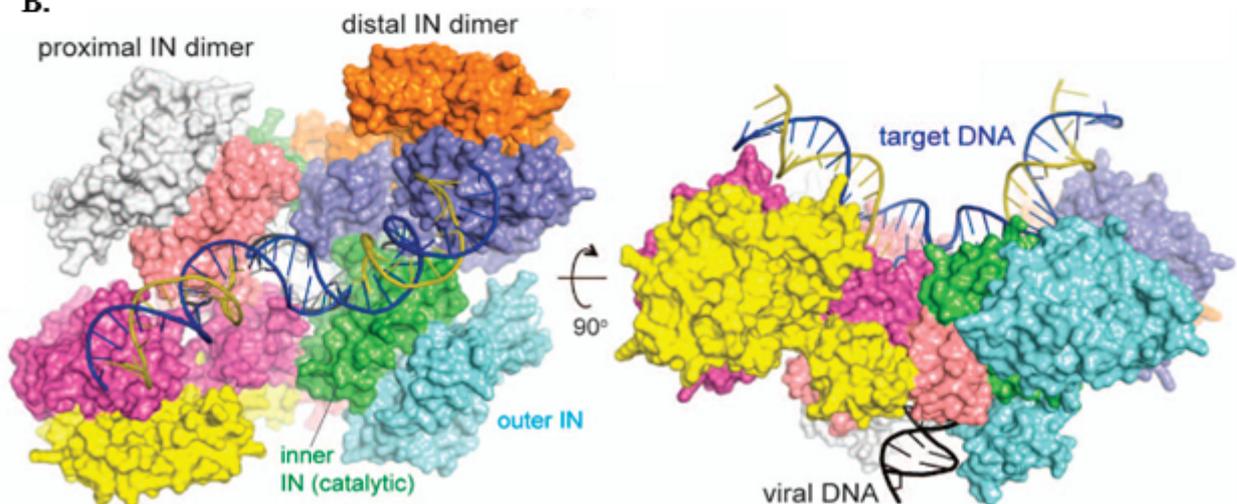


Figure 19 | **A.** Structure de l'intasome de MMTV résolue par cryo-microscopie électronique (code PDB : 3JCA). **B.** Structure cristalline de l'intasome de RSV (code PDB : 5EJK).

L'intasome de MMTV, assemblé avec un ADN dont les extrémités 3' avaient été clivées, représente le CDC (**Figure 19A**), alors que celui de RSV correspond au STC (**Figure 19B**). En effet, l'ADN utilisé pour la formation de ce dernier est composé de trois branches hybridées entre elles et formant une structure unique mimant l'intégration de l'ADN viral dans l'ADN cellulaire. L'architecture tétramérique de l'intasome de PFV a servi de base pour la description de ces intasomes octamériques.

Ces deux intasomes présentent des similarités architecturales. En effet, ils sont tous deux composés d'un « cœur » conservé, le CIC (pour *Conserved Intasome Core*), formé par deux dimères d'IN et similaire à l'intasome de PFV. Autour de cette région se trouvent deux autres dimères d'IN. Chaque dimère se forme par une interface entre les CCD des deux molécules d'IN le composant.

De façon analogue à PFV, les monomères d'IN internes dans le CIC interagissent avec l'ADN viral et sont actifs pour la catalyse de l'intégration. À l'inverse de l'intasome de PFV, le CIC est entouré par d'autres dimères d'IN. Alors que la boucle reliant le NTD au CCD est étendue pour les IN internes du CIC, ceci leur permettant un contact avec les CCD opposés, elle est contractée dans les autres monomères, montrant la nécessité d'une flexibilité entre ces deux domaines.

L'une des plus grandes différences entre l'intasome de PFV et ceux de MMTV et RSV vient de la position des CTD des IN des dimères externes. En effet, ces CTD sont ici proches du CIC, leur permettant d'interagir avec l'ADN viral. Cette différence peut s'expliquer par la variabilité de taille de la boucle reliant le CTD au CCD. Celle-ci est composée de 50 résidus d'acides aminés chez l'IN de PFV alors qu'elle n'en compte que 8 pour les IN de MMTV et RSV (Ballandras-Colas *et al.*, 2016).

De nouvelles structures d'intasomes de lentivirus ont été récemment publiées. L'équipe du Dr. Lyumkis a ainsi résolu la structure cryo-EM de l'intasome du VIH-1 (Passos *et al.*, 2017) après fusion de l'IN à une protéine liant l'ADN, Sso7d, ceci permettant d'augmenter sa solubilité *in vitro* tout en conservant son activité *in vivo* (Li *et al.*, 2014). Cet intasome est composé d'un dimère de dimères avec une architecture similaire à celle de PFV (**Figure 20A**). Les deux sous-unités internes s'enroulent autour de l'ADN viral et se positionnent sur l'ADN cible formant ainsi le STC. Les CTD des sous-unités externes interagissent partiellement avec l'ADN viral et se rapprochent des CTD des sous-unités internes, donnant ainsi des informations qui n'étaient pas visibles dans l'intasome de PFV. Les deux NTD des sous-unités externes ainsi que la protéine de fusion SSo7d ne sont pas observés dans la structure.

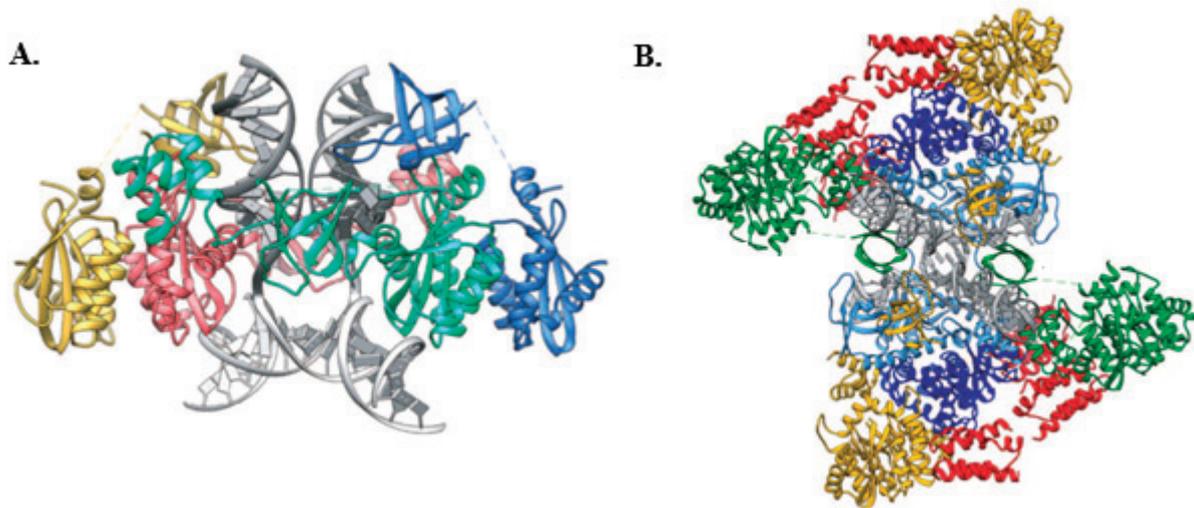


Figure 20 | **A.** Structure de l'intasome du VIH-1 résolue par cryo-microscopie électronique (code PDB : 5U1C). **B.** Structure de l'intasome du VIH-1 en complexe avec un domaine IBD de LEDGF/p75 (en rouge) résolue par cryo-microscopie électronique.

L'équipe du Dr. Lyumkis a également déterminé la structure cryo-EM de l'intasome de VIH-1 associée au domaine IBD de LEDGF/p75 (**Figure 20B**) (Passos *et al.*, 2017). Cette structure permet d'observer un dodécamère d'IN avec des zones additionnelles de densité laissant supposer une forme globale hexadécamérique. Les IBD sont observés sur les régions externes de l'intasome (**Figure 20B**). Ils interagissent avec les domaines NTD et CCD des IN concernées, formant une interface intermoléculaire similaire à celle observée dans la structure cristalline du complexe IBD et fragment NTD-CCD d'IN VIH-2 (Hare *et al.*, 2009).

Enfin, la structure de l'intasome du *Lentivirus*, MVV (pour *Maedi-Visna Virus*) a été récemment résolue par cryo-microscopie électronique (Ballandras-Colas *et al.*, 2017). Celui-ci est formé de seize unités d'IN organisées en tétramère de tétramères et est structuralement très similaire à l'intasome du VIH-1 (**Figure 21**). Cette conservation de structure, malgré la faible identité de séquences entre les IN de MVV et du VIH-1 (moins de 30%), semble confirmer la présence d'une organisation hexadécamérique pour les *Lentivirus* (Engelman & Cherepanov, 2017).

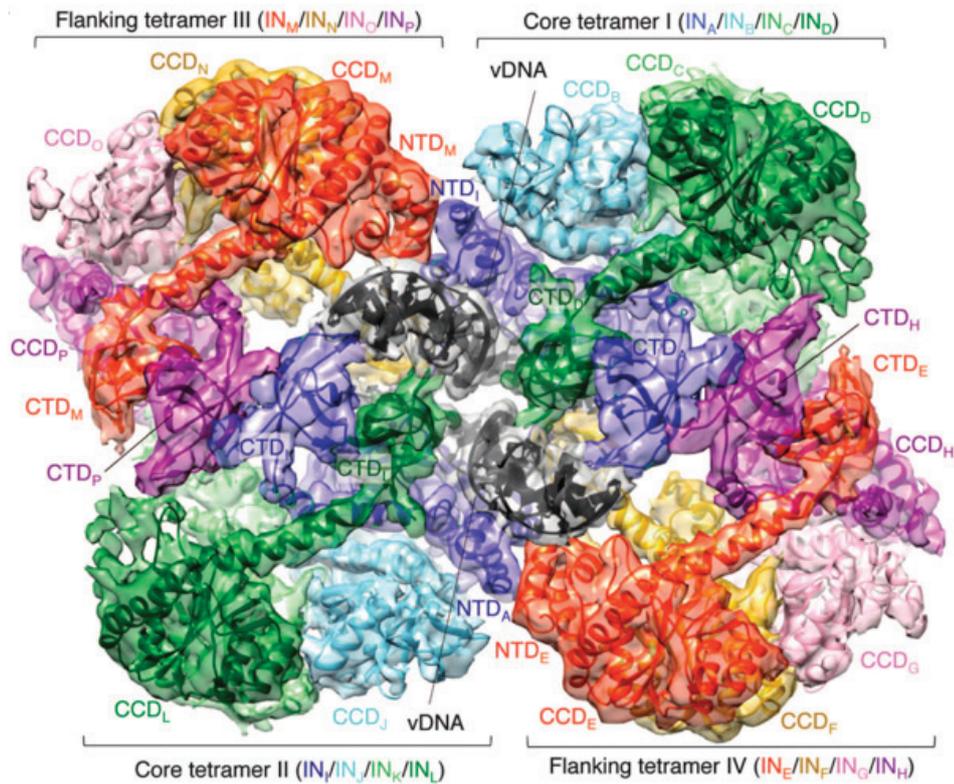


Figure 21 | Structure de l'intasome de MVV résolue par cryo-microscopie électronique (code PDB : 5M0Q).

L'ensemble de ces informations sur l'architecture des intasomes rétroviraux constitue une avancée majeure dans la compréhension du fonctionnement de l'IN et le développement d'inhibiteurs spécifiques. Il semble en effet que l'intasome soit constitué d'un CIC invariable, similaire à l'intasome de PFV et représenté en couleur sur la **Figure 22**. Cependant, il existe une variabilité dans l'oligomérisation de l'IN qui pourrait dépendre du genre rétroviral avec des formes tétramériques pour les *Spumavirus*, octamériques pour les *Alpha-* et les *Bêtarétrovirus* et hexadécamérique pour les *Lentivirus*. Cette organisation semble liée à la longueur du segment connectant les domaines CCD et CTD. Néanmoins, aucune information structurale concernant les *Gammarétrovirus*, les *Deltarétrovirus* et les *Epsilon-rétrovirus* n'a à ce jour été publiée.

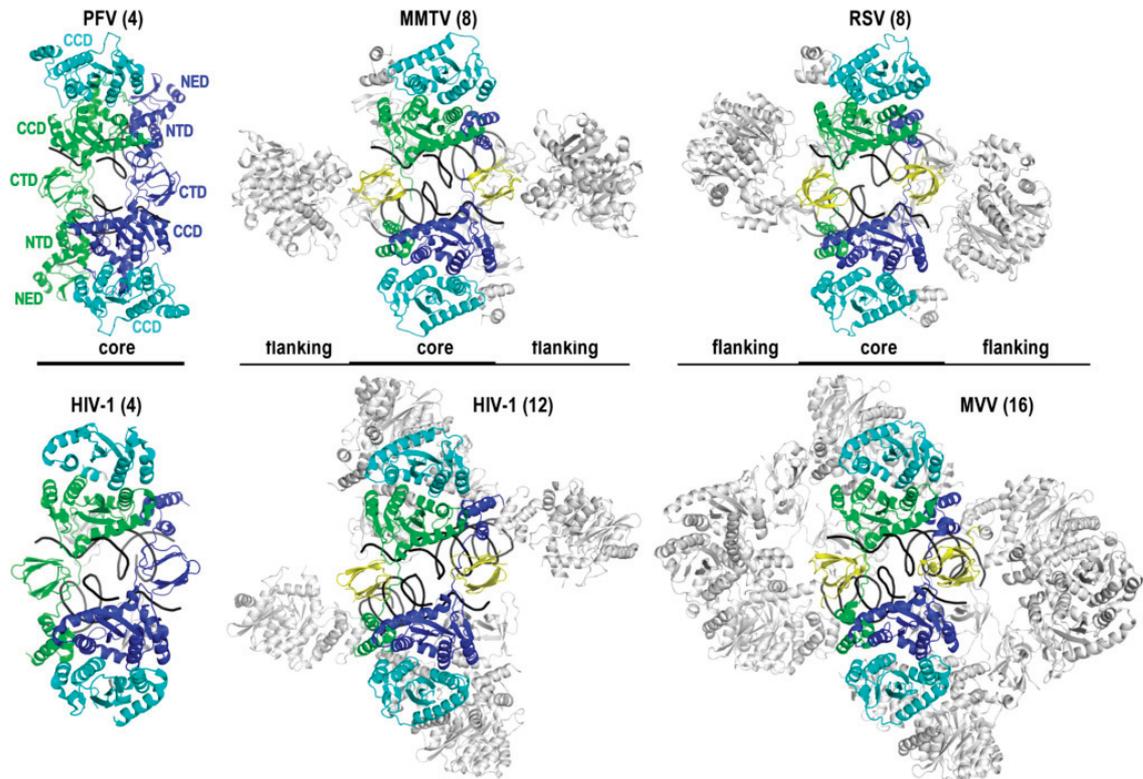


Figure 22 | Structures des intasomes rétroviraux de PFV, MMTV, RSV, VIH-1 et MVV. En couleur sont représentés les IN appartenant au CIC (Engelman & Cherepanov, 2017).

II. 2. 6. Intégrase et nucléosome

Le nucléosome est un complexe nucléoprotéique formé d'un fragment d'ADN double brin d'une longueur d'environ 146 paires de nucléotides enroulé autour d'un cœur protéique. Ce dernier est composé d'un octamère d'histones dans lequel se retrouvent deux exemplaires de chaque histone : H2A, H2B, H3 et H4. Le nucléosome correspond au premier niveau de compactage de l'ADN et est le constituant de base de la chromatine, support de l'information génétique. Les extrémités amino-terminales des histones sont désordonnées et pointent vers l'extérieur du nucléosome, leur valant le nom de « queue d'histone ». Les queues d'histones sont sujettes à des modifications post-traductionnelles qui permettent de recruter différentes protéines à la chromatine.

Parmi les facteurs ayant une influence sur la sélection du site d'intégration se trouve la chromatine. En effet, bien que la plupart des IN utilisent des facteurs cellulaires pour les guider vers des sites d'intégration privilégiés, l'interaction finale a lieu entre l'IN et le nucléosome. Cette interaction est dépendante de la compaction de la chromatine et semble varier selon le

genre rétroviral. Ainsi, l'IN du VIH-1 favorise les régions de faible densité alors que MMLV et les *Spumavirus* favorisent la chromatine compacte (Review (Engelman & Singh, 2018)).

La première structure atomique de nucléosome a été résolue par cristallographie aux rayons X en 1997 (Luger *et al.*, 1997) et la première structure d'un mononucléosome en complexe avec un intasome, celui de PFV, a été déterminée par cryo-microscopie électronique en 2015 (Maskell *et al.*, 2015). Cette structure met en évidence une interface d'interaction formée de trois sous-unités d'IN, de deux tours d'ADN nucléosomique et de l'hétérodimère d'histones H2A-H2B (Figure 23). La queue amino-terminale de l'histone H2A vient se positionner contre le CTD d'une IN interne et apparaît essentielle à la formation du complexe. En accord avec cette observation structurale, la délétion de cette queue d'histone diminue la liaison du nucléosome à l'intasome ainsi que le processus de transfert de brin dans les nucléosomes (Maskell *et al.*, 2015).

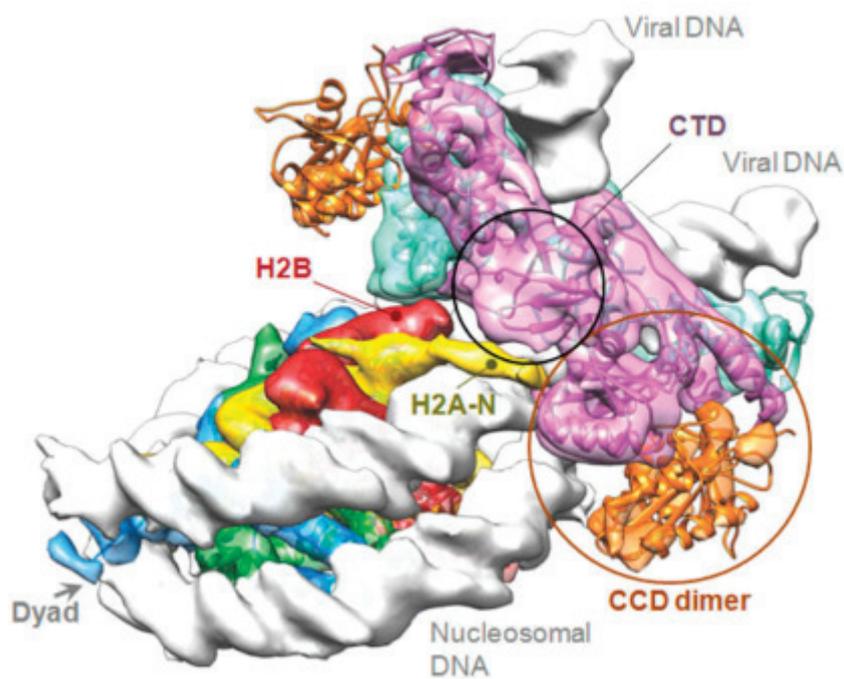


Figure 23 | Modèle de l'intasome de PFV en complexe avec un nucléosome obtenu par cryo-microscopie électronique (code EMDB : 13005). Le dimère d'IN interne est représenté en rose et bleu clair et les CCD des IN externes sont représentés en orange. Les ADN cellulaire et viral sont représentés en densité grise, et l'histone en bleu, jaune, rouge et vert (Maskell *et al.*, 2015).

II. 3. Partenaires cellulaires

Différentes études ont montré l'importance de cofacteurs cellulaires dans l'étape d'intégration *in vivo* (Review (Debyser *et al.*, 2015)). Ces cofacteurs semblent spécifiques à un genre rétroviral ; la protéine LEDGF/p75, qui a déjà été citée dans ce manuscrit, est ainsi associée aux IN lentivirales, tandis que les protéines de la famille BET (Bromodomaines et domaine Extra-Terminal) sont associées aux IN gammarétrovirales. Les sites d'intégration dépendent des spécificités d'interaction de ces cofacteurs avec les queues d'histone modifiées post-traductionnellement (**Figure 9** et **Figure 24**) comme plus amplement détaillé ci-après. À ce jour, aucun cofacteur dirigeant l'intégration des *Spumavirus* ou des *Alpharétrovirus* n'a encore été rapporté, même si le facteur de liaison FACT (*FAcilitates Chromatin Transcription*) pourrait être un bon candidat pour l'IN d'ASLV (Winans *et al.*, 2017).

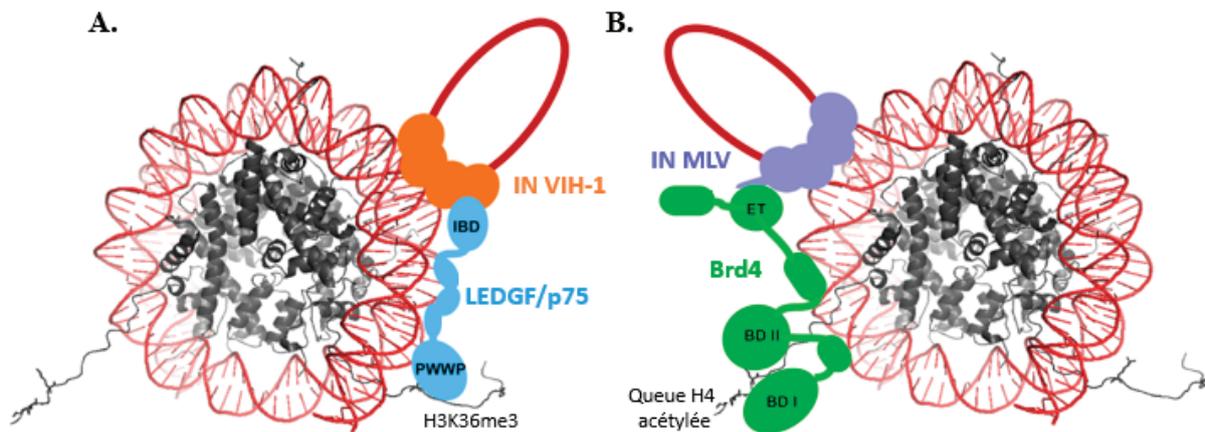


Figure 24 | Modèle de l'interaction d'un nucléosome **A.** avec l'intasome du lentivirus VIH-1 (en orange) et son partenaire cellulaire LEDGF/p75 (en bleu) et **B.** avec l'intasome du gammarétrovirus MLV (en violet) et son partenaire cellulaire Brd4 (en vert). LEDGF/p75 interagit à la fois avec la queue d'histone H3 tri-méthylée sur le nucléosome et avec l'intasome du VIH-1 grâce à son IBD. Brd4 interagit quant à elle avec la queue d'histone H4 acétylée et avec l'intasome de MLV grâce à son domaine ET. Les histones sont représentées en gris et l'ADN nucléosomique en rouge (Kvaratskhelia *et al.*, 2014).

II. 3. 1. LEDGF/p75

Chez les *Lentivirus*, et notamment VIH-1, l'IN requiert la présence de LEDGF/p75 pour une intégration efficace (Ciuffi *et al.*, 2005; Engelman & Cherepanov, 2008). LEDGF/p75 est une protéine cellulaire ubiquitaire de la famille HRP (pour *Hepatoma-derived growth factor Related Proteins*) ayant la capacité de s'associer à la chromatine (Nishizawa *et al.*, 2001). Elle est

originellement connue comme étant un coactivateur de transcription mais elle a aussi la capacité de lier l'IN du VIH-1 et joue un rôle dans sa localisation nucléaire (Maertens *et al.*, 2003; Busschots *et al.*, 2005). L'épissage alternatif de l'ARN messager codant LEDGF/p75 permet la synthèse de l'isoforme p52 (Ge *et al.*, 1998). Contrairement à LEDGF/p75, p52 ne possède pas de domaine IBD (**Figure 25**) et n'a pas la capacité d'interagir avec l'IN de VIH-1 ou d'effet sur l'intégration.

LEDGF/p75 est composé de 530 résidus d'acides aminés et contient deux domaines structurés (**Figure 25**). Le domaine PWWP se trouve à l'extrémité amino-terminale. Il comprend en son centre une signature Pro-Pro-Trp-Trp, appartient à la famille Tudor et est capable d'interagir avec la chromatine (Eidahl *et al.*, 2013). À l'extrémité carboxy-terminale se trouve un domaine IBD qui lie le CCD et le NTD de l'IN des *Lentivirus* (Cherepanov *et al.*, 2004; Hare *et al.*, 2009), le CCD étant essentiel pour la liaison et le NTD renforçant l'interaction.

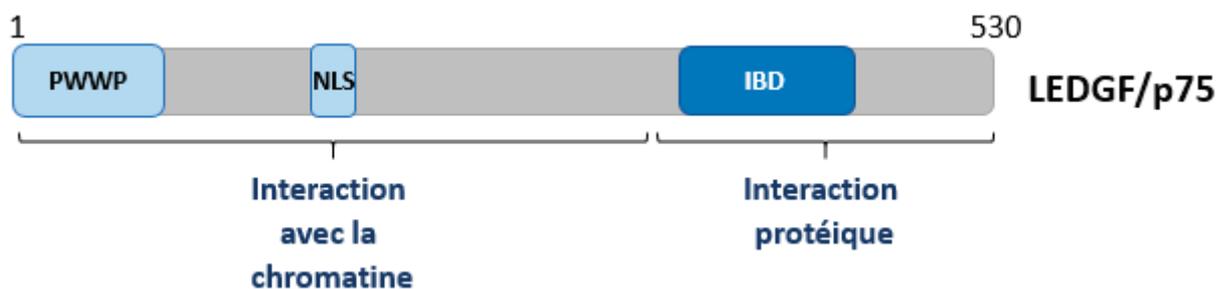


Figure 25 | Organisation schématique de LEDGF/p75 et ses régions d'interaction avec la chromatine (dont le domaine PWWP) et d'interaction protéique (dont le domaine de liaison à l'intégrase IBD) (adaptée de (Debyser *et al.*, 2015)).

Un signal de localisation nucléaire (NLS, pour *Nuclear Localization Signal*) est observé entre les domaines PWWP et IBD. Grâce à ce signal, LEDGF/p75 est localisé principalement dans le noyau des cellules. Son domaine PWWP lui permet d'interagir avec les nucléosomes triméthylés sur la lysine en position 36 de la queue de l'histone H3 (H3K36me3). Cette modification est un marqueur épigénétique associé à l'élongation de la transcription et enrichi dans les régions transcriptionnellement actives (van Nuland *et al.*, 2013). Ainsi, LEDGF/p75 contrôle la localisation de l'intégration dans le génome cellulaire en favorisant les unités de transcription et les zones riches en introns.

D'autres rôles sont attribués à LEDGF/p75 dans le cadre du cycle réplcatif du VIH-1. En effet, cette protéine protège l'IN d'une dégradation protéolytique (Llano *et al.*, 2004) et a

également été décrite comme stimulateur de l'activité enzymatique de l'IN *in vitro* et *in vivo* (Busschots *et al.*, 2005).

II. 3. 2. Les protéines BET

Chez les *Gammaretrovirus*, l'IN du virus MLV interagit avec les protéines Brd2, Brd3, Brd4 et Brdt de la famille BET (Studamire & Goff, 2008). Des observations similaires ont été rapportées pour les IN de FeLV (pour *Feline Leukemia Virus*) (Gupta *et al.*, 2013) et REV (pour *avian Reticuloendotheliosis Virus*) (Serrao *et al.*, 2015).

Les protéines Brd2, Brd3 et Brd4 sont ubiquitaires tandis que Brdt se retrouve spécifiquement dans le tissu testiculaire. Ces protéines existent chez les mammifères, la levure et la drosophile et sont impliquées dans le contrôle du cycle cellulaire, de la transcription et de la réplication de l'ADN (Wu & Chiang, 2007). De plus, elles sont engagées dans des interactions hôte-pathogène (Weidner-Glunde *et al.*, 2010).

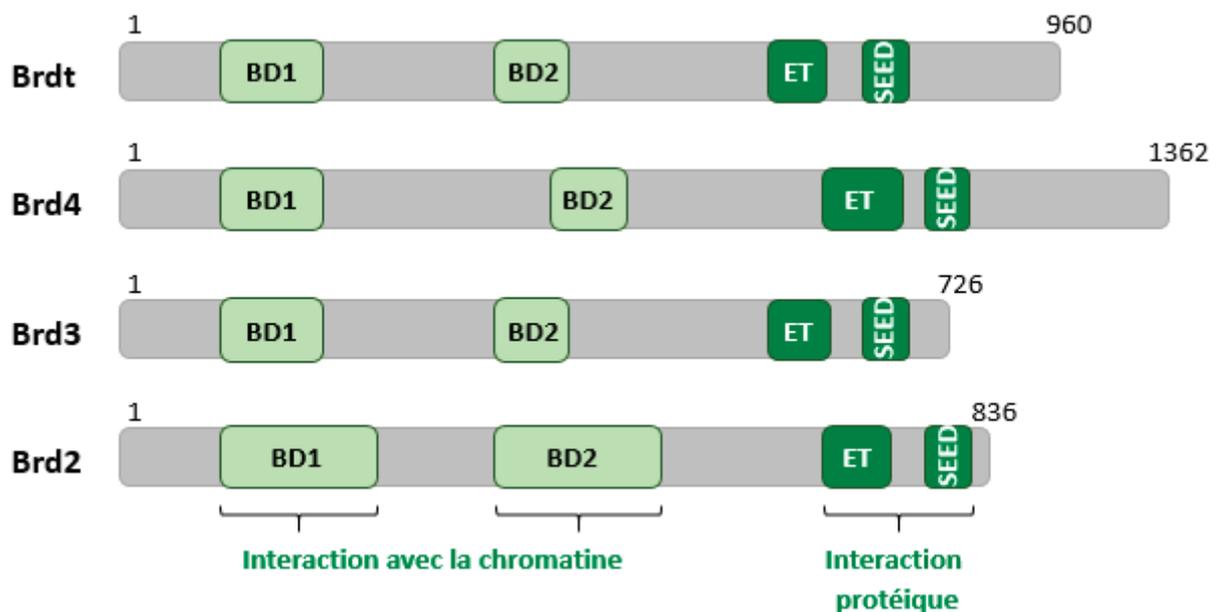


Figure 26 | Organisation schématique des protéines de la famille BET : Brdt, Brd2, Brd3 et Brd4. Leurs domaines d'interaction avec la chromatine (bromodomaines BD1 et BD2) et d'interaction protéique (domaine extra-terminal ET et SEED) sont également représentés (adaptée de (Debyser *et al.*, 2015)).

Ces protéines possèdent deux bromodomaines (BD) au niveau amino-terminal, et un domaine extra-terminal (ET) proche de l'extrémité carboxy-terminale (**Figure 26**). Les BD sont des sous-unités d'environ 110 résidus d'acides aminés, ayant la capacité de fixer des queues

d'histones acétylées (Dey *et al.*, 2003). Le domaine ET est impliqué dans des interactions avec de nombreuses protéines cellulaires et virales (Belkina & Denis, 2012). Les protéines à bromodomains détectent des caractéristiques spécifiques sur la chromatine et recrutent des protéines régulant cette dernière, tels que des facteurs de transcription, des co-activateurs ou des co-répresseurs. Chacune de ces protéines reconnaît des histones particulières, indiquant des différences d'affinité pour la chromatine et des interactions spécifiques avec certaines régions du génome (Dey *et al.*, 2003; Umehara *et al.*, 2010). Enfin, elles possèdent un domaine SEED riche en serine, acide glutamique et acide aspartique, qui est également impliqué dans des interactions protéiques (Belkina & Denis, 2012).

La zone d'interaction entre les IN et les protéines BET a été particulièrement étudiée pour le complexe IN MLV:Brd4 (**Figure 24B**). Cette interaction implique un motif de 17 résidus appelé EBM (pour *ET Binding Motif*) qui est conservé dans la queue carboxy-terminale des IN des *Gammarrétrovirus* (**Figure 27**) et le domaine ET de Brd4 (**Figure 26**) (Sharma *et al.*, 2013). Au sein de ce complexe, l'IN de MLV peut cibler les sites d'initiation de la transcription *via* Brd4. En effet, les sites d'intégration de MLV correspondent aux sites de liaison des protéines BET sur la chromatine et plus précisément les histones H4 di- et tétra-acétylées et les histones H3 di-acétylées (De Rijck *et al.*, 2013).

Des hybrides contenant le domaine PWWP de LEDGF/p75 et le domaine ET de Brd4 permettent, par ailleurs, de rediriger l'intégration de MLV vers les unités actives de transcription, comme chez les *Lentivirus*, loin des sites d'initiation de la transcription. Ce résultat confirme l'importance des protéines BET pour la sélection des sites d'intégration des *Gammarrétrovirus* (De Rijck *et al.*, 2013).

De même, l'utilisation d'inhibiteur des domaines BD inhibe la réplication de MLV à l'étape d'intégration (De Rijck *et al.*, 2013). La petite molécule inhibitrice JQ1 cible ainsi spécifiquement les protéines de la famille BET. Elle agit en se positionnant à l'interface de liaison entre l'un des bromodomains des protéines BET et les queues d'histones acétylées, bloquant ainsi l'interaction avec la chromatine (Filippakopoulos *et al.*, 2010). L'inhibition des protéines BET par JQ1 réduit l'intégration de MLV *in vitro* et a également un effet sur le ciblage des sites d'intégration, le redirigeant loin de ses sites privilégiés (Gupta *et al.*, 2013).



Figure 27 | Organisation schématique des domaines de l'IN de MLV : NED (en jaune), NTD (en vert), CCD (en bleu) et CTD (en orange). Le motif EBM du CTD est représenté en rouge.

Des études RMN ont également permis de mieux comprendre les déterminants moléculaires et structuraux de l'interaction entre le segment carboxy-terminal EBM de MLV IN et le domaine ET de Brd4 (Crowe *et al.*, 2016). Le domaine EBM n'est pas structuré dans la structure isolée du CTD (Lin *et al.*, 2008) alors qu'il adopte une conformation en épingle à cheveux β lorsqu'il interagit avec ET. La structure du complexe EBM IN MLV:ET Brd4 montre en effet que l'interaction se fait par la formation d'un feuillet β antiparallèle intermoléculaire (**Figure 28**) (Crowe *et al.*, 2016).

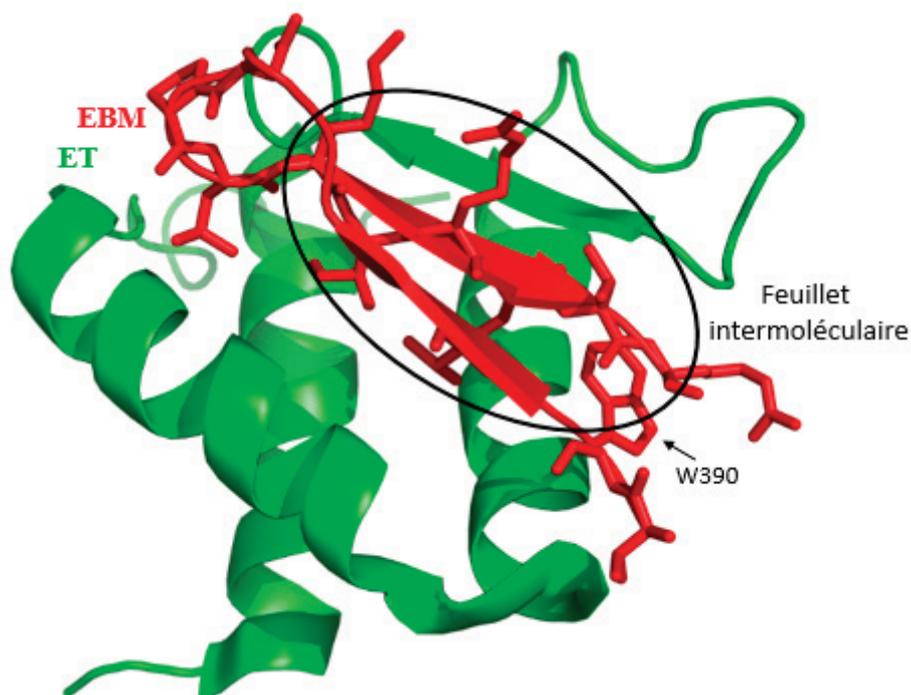


Figure 28 | Structure du domaine ET de Brd4 (en vert) lié à l'EBM de l'IN de MLV (en rouge) résolue par RMN (code PDB : 2N3K). Les résidus de l'IN situés à moins de 4 Å de Brd4 sont représentés en bâton.

Ce feuillet est composé d'un brin β venant du domaine ET de Brd4 et de deux brins β apportés par l'IN, qui viennent se positionner dans une crevasse hydrophobe formée par deux hélices α et le brin β associé de ET. Des acides aminés issus de l'IN de MLV et impliqués dans

cette interaction ont également été mis en évidence. C'est notamment le cas du tryptophane W390 qui est strictement conservé chez les *Gammaretrovirus* et dont la mutation en alanine change le profil d'intégration de MLV (El Ashkar *et al.*, 2014).

II. 4. Inhibiteurs de l'intégrase

Les infections rétrovirales sont à l'origine de nombreuses pathologies pouvant conduire jusqu'au décès. Il est donc crucial de développer des stratégies antirétrovirales afin de bloquer l'infection. Plusieurs étapes spécifiques du cycle de réplication des rétrovirus font ainsi l'objet d'études visant la conception ou l'optimisation de composés antirétroviraux. Ainsi, les mécanismes de fusion cellulaire, de transcription inverse et de maturation sont ciblés par des molécules inhibitrices, toutes approuvées par la FDA (*Food and Drug Administration*) des États-Unis. En raison de son rôle essentiel dans la réplication virale et de l'absence d'équivalent fonctionnel chez l'homme, l'IN est également une cible idéale pour le développement de médicaments contre les infections rétrovirales. Des composés de type INSTI (*Integrase Strand Transfer Inhibitor*) qui ciblent l'intasome sont ainsi utilisés en multithérapie anti-VIH-1 et des recherches sont en cours pour développer des inhibiteurs d'interaction entre l'IN du VIH-1 et son cofacteur LEDGF/p75. Ces deux types d'inhibiteurs seront détaillés dans ce chapitre.

II. 4. 1. Les INSTI – Inhibiteur du transfert de brin

La première description d'une classe de molécules avec la capacité d'inhiber la réplication du VIH-1 remonte à 2000 (Hazuda *et al.*, 2000). Il s'agit d'inhibiteur du transfert de brin (INSTI) qui vient se lier au site actif de l'IN quand la protéine est en complexe avec l'ADN viral sous forme d'intasome (Hare *et al.*, 2010). De par la grande conservation du site actif des IN, ces molécules sont efficaces sur une large variété de rétrovirus. Les INSTI inhibent principalement la réaction de transfert de brin (**Figure 10**) mais ils ont également un effet sur le clivage des extrémités 3' s'ils sont utilisés à forte concentration. Afin d'être efficace, un INSTI doit pouvoir rester lié à l'intasome jusqu'à la destruction complète du PIC par la cellule. C'est le cas des molécules approuvées par la FDA, qui ont un temps de dissociation allant de plusieurs heures à plusieurs jours (Hightower *et al.*, 2011).

L'optimisation d'un pharmacophore de type acide dicétonique a mené à la découverte du Raltégravir (RAL), première molécule à être validée pour le traitement du SIDA en 2007 (Murray *et al.*, 2007). Deux autres molécules au mode d'action identique, Elvitégravir (EVG) et

Dolutégravir (DTG), ont reçu leur autorisation de mise sur le marché (Sato *et al.*, 2006; Johns *et al.*, 2013). Les INSTI partagent deux caractéristiques fonctionnelles : ils possèdent tous un cœur chélatant le Mg^{2+} (généralement une triade d'atomes d'oxygène) ainsi qu'une chaîne latérale aromatique flexible (**Figure 29**).

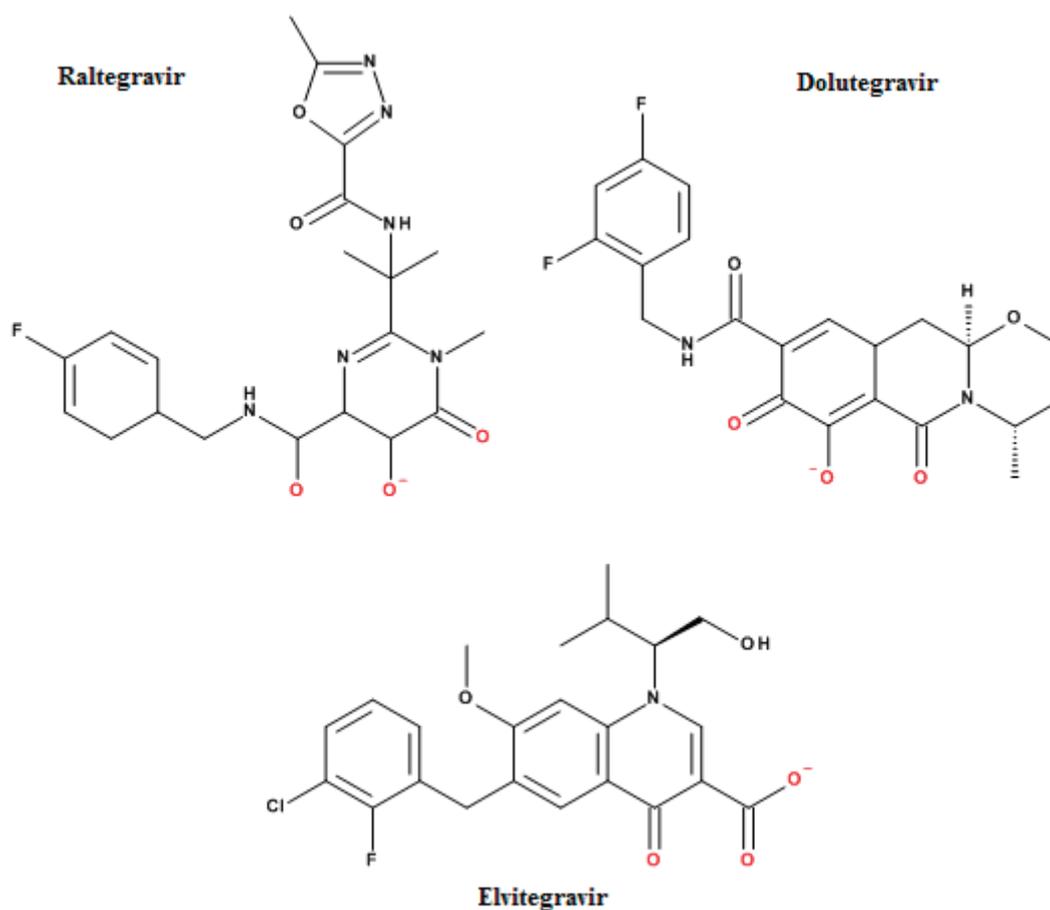


Figure 29 | Structures chimiques des inhibiteurs du transfert de brin (INSTI) du VIH-1 : Raltégravir, Dolutégravir et Elvitégravir. Les atomes chélatant l'ion Mg^{2+} sont présentés en rouge.

La structure 3D de l'intasome de PFV a permis de comprendre plus précisément le mode d'action de ces trois molécules. Ainsi, elles se positionnent dans le site actif de l'IN de façon à ce que leur triade d'oxygène interagisse avec les cations catalytiques et déplacent l'extrémité 3' de l'ADN viral *via* des interactions entre groupes aromatiques (**Figure 30**), ceci empêchant la réaction de transfert de brin d'être catalysée (Hare *et al.*, 2011).

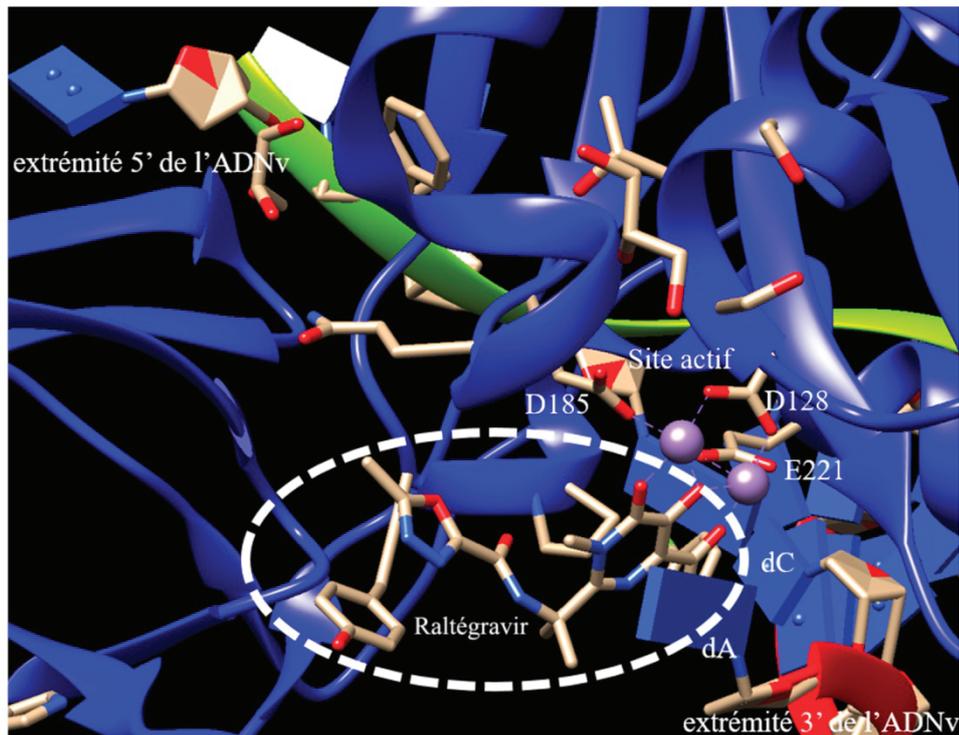


Figure 30 | Vue rapprochée du site actif de l'un des deux monomères internes de l'intasome de PFV (représentation en ruban) en complexe avec une molécule de Raltégravir (représentation en bâton) (code PDB : 3OYA). Le Raltégravir chélate les cations divalents (boules violettes) du site actif où la triade catalytique DDE est représentée en bâton. Cette association déplace l'extrémité 3' de l'ADN viral *via* des interactions de type aromatique entre le Raltégravir et les groupes invariants adénine (dA) et cytosine (dC).

Toutefois, l'apparition d'une résistance aux inhibiteurs de l'IN a été observée depuis la mise sur le marché de ces molécules (Metifiot *et al.*, 2013). La grande variabilité du génome du VIH-1 est à l'origine de l'émergence de variants pour lesquels l'IN porte des mutations lui conférant une résistance aux INSTI (Hare *et al.*, 2010). Cette variabilité est également la raison d'une émergence de résistance aux antirétroviraux ciblant les protéines virales autres que l'IN. Il est donc nécessaire de développer des molécules de nouvelle génération, dont la barrière génétique est meilleure dans le cadre d'une multithérapie antirétrovirale hautement active (HAART, pour *Highly Active AntiRetroviral Therapy*) (Lu *et al.*, 2018). Des résultats encourageants ont été obtenus sur un nouvel INSTI, le Carbotégravir (Jacobson & Flexner, 2017), qui est en phase d'approbation par la FDA et pourrait être utilisé en injection mensuelle avec un deuxième antirétroviral, la Rilpivirine, un analogue non-nucléosidique inhibant la RT.

II. 4. 2. Les ALLINI – Inhibiteurs allostériques

En raison des risques de résistances inhérents à toute classe de molécules ciblant une région précise, ici le site actif de l'IN, les recherches se sont également tournées vers des molécules visant des sites non catalytiques importants pour l'intégration. Ainsi, des inhibiteurs allostériques de l'IN appelées LEDGIN puis ALLINI ont été envisagés (Engelman *et al.*, 2013). Ces molécules, en phase de développement, ciblent la poche de liaison à LEDGF/p75 au niveau de l'interface de dimérisation du CCD de l'IN de VIH-1.

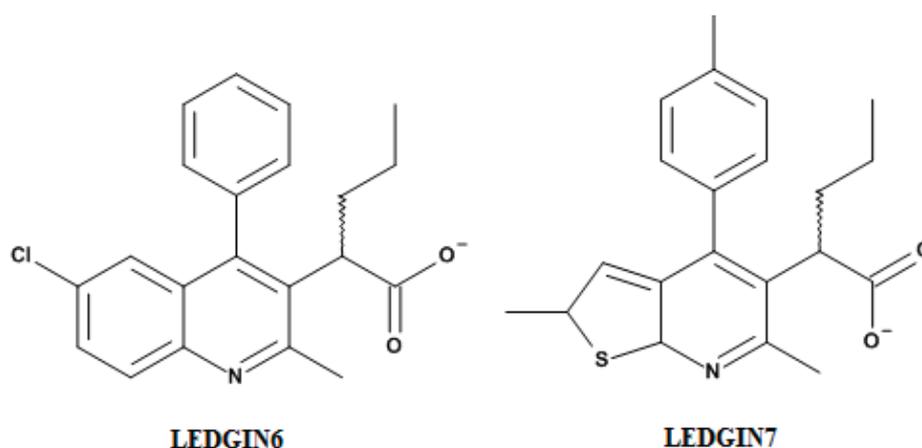


Figure 31 | Structures chimiques des inhibiteurs allostériques de l'IN (ALLINI) du VIH-1 : LEDGIN6 et LEDGIN7.

Les ALLINI possèdent deux caractéristiques qui leur sont propres : une queue carboxylique portant deux oxygènes liant l'IN et mimant l'interaction entre celle-ci et LEDGF/p75 ; et une chaîne latérale aromatique permettant d'imiter l'interaction hydrophobe de l'isoleucine en position 365 de LEDGF/p75 (**Figure 31**).

Des études en cellules infectées ont montré que ces molécules n'ont qu'un effet inhibiteur limité dans la phase pré-intégrative du cycle de réplication du VIH-1. Elles ont cependant un effet inhibiteur important et inattendu dans la phase post-intégrative (Gupta *et al.*, 2014) en provoquant une multimérisation aberrante de l'IN *via* des interactions CCD-CTD. Cette multimérisation empêche l'interaction entre l'IN et l'ARN viral dans la particule virale lors de sa maturation et engendre la production de particules non-infectieuses caractérisées par un complexe ribonucléoprotéique localisé en dehors de l'assemblage capsidique CA (voir **Figure 11**) (Gupta *et al.*, 2016; Koneru *et al.*, 2019). Des inhibiteurs pensés de manière rationnelle pour bloquer une étape d'un cycle viral peuvent ainsi avoir des effets surprenants sur une autre étape.

PROJET DOCTORAL

Le sujet. Mon travail doctoral s'est orienté à la fois sur l'étude de complexes intégrase / cofacteurs cellulaires - afin de mieux comprendre leur mode d'assemblage - et la recherche de nouvelles molécules antirétrovirales bloquant leur interaction. Ma thèse a plus spécifiquement porté sur l'étude moléculaire et structurale de l'IN du *Gammarrétrovirus* endogène porcine PERV-A/C en complexe avec son cofacteur, la protéine humaine Brd2, dans le cadre de la prévention de zoonoses rétrovirales.

Comme expliqué dans la première partie du manuscrit (voir **Introduction §I.2.2.**), le génome du porc contient des séquences rétrovirales endogènes actives, PERV, intégrées à leur génome et transmissibles à leur descendance. Des études ont mis en évidence la capacité de ces séquences rétrovirales à infecter des lignées cellulaires d'origine humaine (Patience *et al.*, 1997). Ce risque de transmission a limité le développement de l'utilisation de tissus porcins en xénotransplantation, bien que le porc soit une source de prélèvement prometteuse. Son élevage dans des sites strictement confinés permet en effet une production à grande échelle avec un haut niveau sanitaire (Moalic *et al.*, 2011).

Face à la pénurie de greffons humains utilisés en allogreffe, des greffons porcins pourraient être utilisés comme peau temporaire pour les patients souffrants de plaies sévères (Scobie *et al.*, 2013). Ils pourraient également être utilisés dans les traitements de la maladie de Parkinson, du diabète ou encore de l'opacité de la cornée (Denner, 2016; Cooper *et al.*, 2018).

Le rétrovirus PERV comporte trois sous-types infectieux A, B et C. PERV-A et PERV-B sont présents dans la lignée germinale de tous les porcs et présentent un tropisme pour les cellules porcines et humaines (Denner, 2008). PERV-C se trouve dans les cellules germinales de certains porcs et infecte seulement les cellules porcines (Takeuchi *et al.*, 1998). Le recombinant PERV-A/C présente un tropisme pour les cellules porcines et humaines et un titre infectieux 500 fois plus élevé que celui de PERV-A pour les cellules humaines (Bartosch *et al.*, 2004). Le provirus PERV-A/C représente donc un risque de maladie infectieuse pour l'Homme lors d'une xénotransplantation. Ce risque est appelé zoonose rétrovirale (Denner, 2008).

Il est crucial de bien comprendre la réplication de PERV-A/C, afin de proposer des stratégies médicales permettant de développer l'utilisation du porc pour des xéngreffes. L'IN de PERV-A/C est une cible thérapeutique de choix pour le développement d'inhibiteurs spécifiques antirétroviraux, en raison de l'absence d'homologue cellulaire chez l'Homme.

Étude des IN de PERV-A/C au sein de l'équipe. Avant le début de ma thèse, mon équipe d'accueil avait collaboré avec les équipes du Dr. Ronfort (Infections Virales et Pathologie Comparée, INRA Lyon) et du Dr. Blanchard (Génétique Virale et Bio-sécurité, ANSES Ploufragan) pour caractériser *in vitro* et *in cellulo* l'inhibition de l'IN de PERV-A/C par le Raltégravir. Cette étude avait permis de construire un modèle de l'intasome de PERV-A/C en complexe avec cet INSTI (**Figure 32**) (Demange *et al.*, 2015) en prenant comme guide la structure cristalline de l'intasome de PFV en complexe avec le Raltégravir (Hare *et al.*, 2011).

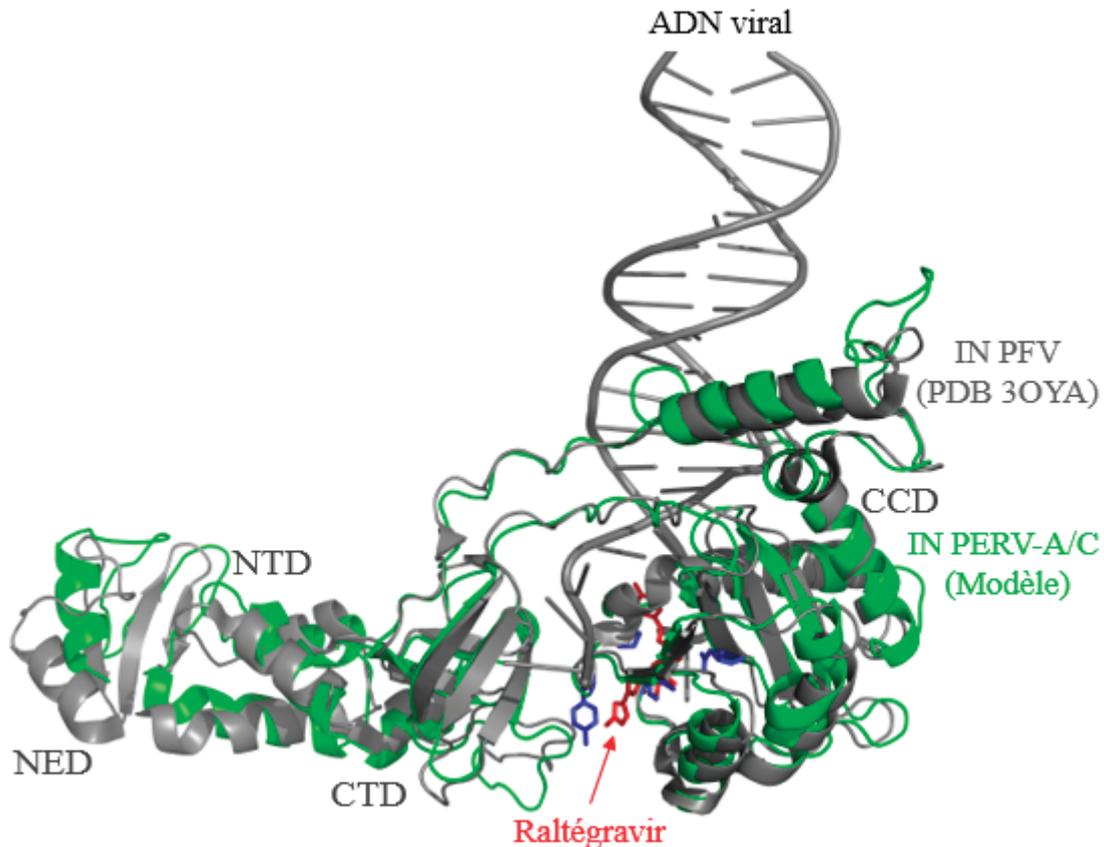


Figure 32 | Superposition des structures des intasomes de PFV (en gris, code PDB : 3OYA) et de PERV-A/C (modèle en vert (Demange *et al.*, 2015)) en complexe avec le Raltégravir, L'ADN viral est représenté en gris.

Ce modèle, construit par le Dr. Yajjou durant son stage doctoral de 2014 à 2017 sous la supervision du Pr. Gouet, suggérait que les déterminants structuraux de l'interaction intasome / Raltégravir étaient conservés entre PFV et PERV-A/C (Demange *et al.*, 2015). De fait, les expériences *in vitro* et *in cellulo* ont montré que le Raltégravir était un puissant inhibiteur de l'IN de PERV-A/C avec une concentration inhibitrice médiane (IC₅₀) de l'ordre du nanomolaire.

Par ailleurs, l'équipe du Dr. Blanchard avait montré que l'intégration du génome de PERV-A/C dans les cellules humaines était fortement favorisée sur les sites d'initiation de la transcription et les îlots CpG, comme observé pour les autres *Gammarétrovirus* (Moalic *et al.*, 2006; Moalic *et al.*, 2009). Elle avait également observé que l'IN de PERV-A/C interagissait comme celle de MLV avec les protéines de la famille BET (voir **Introduction §II.3.2.**) et que la protéine humaine Brd2 se liait spécifiquement à l'IN de PERV-A/C pour être co-localisée dans le noyau. Des expériences de double-hybride menées par l'équipe du Dr. Ronfort montraient que le domaine CTD de l'IN de PERV A/C, et plus particulièrement le segment EBM (378-394), interagissait avec le domaine ET de Brd2 (**Figure 33**).

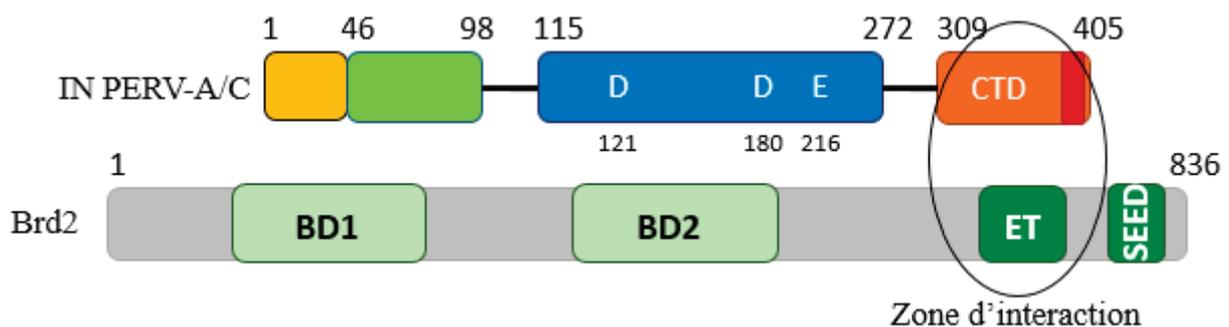


Figure 33 | Interaction entre le domaine CTD de l'IN de PERV-A/C et le domaine ET (667-735) du cofacteur cellulaire Brd2. Surlignée en rouge, la séquence EBM (378-394) du CTD qui est conservée chez les *Gammarétrovirus* et semble essentielle à l'interaction. L'IN comprend également les domaines NED (jaune), NTD (vert) et CCD (bleu) tandis que Brd2 comprend les domaines BD d'interaction avec la chromatine et SEED d'interaction protéique.

Ces résultats d'interaction entre l'IN de PERV-A/C et Brd2 n'avaient pas été publiés lorsque j'ai commencé ma thèse. J'ai participé à cette étude en apportant les aspects biochimiques et structuraux, afin de démontrer par une approche multidisciplinaire que les *Gammarétrovirus* (MLV, PERV, etc.) avaient un schéma d'intégration commun et que leurs IN utilisaient les protéines de la famille BET comme cofacteur cellulaire. Les chapitres suivants portent ainsi sur : 1) La mise au point d'un protocole original d'expression et de purification de deux complexes d'interaction, un complexe CTD:ET-SEED(667-808) et un complexe minimal CTD:ET (667-735) (**Figure 33**) ; 2) Des études pour caractériser *in vitro* la dissociation du complexe ; 3) L'étude structurale du complexe d'interaction entre le CTD d'IN PERV-A/C et le domaine ET de Brd2 (*i.e.* segment 667-735), qui a mené à la détermination de la première structure SAXS d'un complexe CTD:ET. L'ensemble de ces résultats a donné lieu à un article publié dans *Virology*, que je signe en co-premier auteur (Gallay *et al.*, 2019).

J'ai également, durant ma thèse, travaillé dans l'équipe avec le Dr. Robert, IR CNRS, sur la conception rationnelle d'inhibiteurs dirigés contre cette interface hétérodimérique et avec le Dr. Fiorini, CR CNRS, sur la mise au point d'un protocole d'expression et de purification de l'intasome de PERV-A/C en vue de son étude structurale par cristallographie aux rayons X et/ou cryo-microscopie électronique.

MATÉRIELS ET MÉTHODES

I. Techniques de biologie moléculaire

I. 1. Constructions plasmidiques

Plusieurs constructions plasmidiques ont été nécessaires à la réalisation de mon projet de thèse. Pour cela, trois plasmides distincts ont été utilisés : pET-28a, pPROEX-Hta et pHL8. Ces plasmides ont en commun une origine de répllication bactérienne, un site multiple de clonage (MCS, pour *Multiple Cloning Site*), un gène de résistance à un antibiotique et un gène codant le répresseur LacI. Ils possèdent également un promoteur T7 (Studier & Moffatt, 1986) placé sous le contrôle d'un opéron lactose et un terminateur T7, à l'exception du plasmide pPROEX-Hta qui possède un promoteur *trc*. Ce plasmide permet d'obtenir un fort taux d'expression basale du gène placé sous la dépendance du promoteur *trc* (Tegel *et al.*, 2011).

Le gène codant la protéine recombinante d'intérêt est inséré dans le MCS entre deux sites de restriction et permet l'ajout d'étiquettes de purification variables à une extrémité ou l'autre de la protéine (Terpe, 2003). Les étiquettes utilisées au cours de ce projet sont i) l'étiquette poly-histidine, composée de six résidus histidyl, permettant la purification de la protéine par chromatographie d'affinité pour le nickel (Porath *et al.*, 1975) ; et ii) la Glutathion-S-Transférase (GST) qui permet la purification par chromatographie d'affinité pour le glutathion (Smith & Johnson, 1988). Ces étiquettes sont éliminables grâce à la présence d'un site de clivage spécifique d'une protéase entre la protéine d'intérêt et l'étiquette de purification. Il s'agit de la protéase du virus de la gravure du tabac (TEV, pour *Tobacco Etch Virus*) ou de la thrombine.

Les différentes constructions utilisées au cours de ma thèse sont détaillées dans le tableau ci-après ainsi que dans les sous-parties suivantes (**Tableau 1**).

Constructions		Vecteurs d'expression	Étiquette de purification	Sites de clivages	Antibiotique
1	CTD IN PERV-A/C – Brd2 (667-808)	pET-28a	Strep-tag® II et poly-histidine	Protéase du TEV	Kanamycine
2	CTD IN PERV-A/C W382A – Brd2 (667-808)	pET-28a	Strep-tag® II et poly-histidine	Protéase du TEV	Kanamycine
3	CTD IN PERV-A/C R391A/L392A – Brd2 (667-808)	pET-28a	Strep-tag® II et poly-histidine	Protéase du TEV	Kanamycine
4	CTD IN PERV-A/C R397A – Brd2 (667-808)	pET-28a	Strep-tag® II et poly-histidine	Protéase du TEV	Kanamycine
5	CTD IN PERV-A/C – Brd2 (667-735)	pET-28a	Strep-tag® II et poly-histidine	Protéase du TEV	Kanamycine
6	Brd2 (667-808)	pPROEX-Hta	Poly-histidine	Protéase du TEV	Ampicilline
7	IN PERV-A/C	pET-28a	Poly-histidine	Thrombine	Kanamycine
8	IN PERV-A/C	pHL8	GST et poly-histidine	Protéase du TEV	Kanamycine
9	IN PERV-A/C	pPROEX-Hta	Poly-histidine	Protéase du TEV	Ampicilline

Tableau 1 | Constructions plasmidiques utilisées pour la surproduction des protéines d'intérêt.

I. 1. 1. Les constructions plasmidiques MultiTEV

Avant mon arrivée au laboratoire, une construction permettant l'expression simultanée des gènes codant le CTD de l'IN de PERV-A/C et le fragment Brd2 (667-808) (région ET-SEED, voir **Figure 33**) avait été réalisée par le Dr. Yajjou pendant son stage doctoral. Cette construction, basée sur les régions interagissant des deux protéines, avait été clonée dans le plasmide pET-28a. Elle contient une séquence polycistronique composée des trois gènes suivants :

- le gène codant la protéase du TEV ;
- le gène codant le CTD de l'IN de PERV-A/C, avec une étiquette de purification Strep-tag® II à son extrémité amino-terminale ;

- le gène codant le fragment Brd2 (667-808), avec une étiquette poly-histidine à son extrémité amino-terminale.

Ces trois gènes sont séparés par des séquences codant le site de clivage à la protéase du TEV (**Figure 34A**). Cette construction sera nommée MultiTEV dans la suite de ce manuscrit.

A. MultiTEV



B. MultiTEV minimal



Figure 34 | Schématisation des inserts contenant **A**, les gènes codant la protéase du TEV, le CTD de l'IN de PERV-A/C et Brd2 (667-808), appelé MultiTEV ; et **B**, les gènes codant la protéase du TEV, le CTD de l'IN de PERV-A/C et Brd2 (667-735), appelé MultiTEV Minimal.

Cette stratégie MultiTEV avait été mise en place pour la première fois dans l'équipe par le Dr. Yajjou, car elle avait observé la précipitation des deux protéines lors de leur mise en contact après leurs surproductions et purifications isolées. Cette stratégie est basée sur les travaux de l'équipe du Dr. Ruff (Biologie Structurale Intégrative, IGBMC Illkirch) avec laquelle nous collaborons sur l'étude de l'IN de VIH-1 (Levy *et al.*, 2016).

Cette construction originale permet la surproduction d'une polyprotéine où la protéase du TEV, exprimée en premier, reconnaît ses deux sites spécifiques et libère les deux protéines d'intérêt. Ces dernières, qui ont pu former un complexe avant ou après l'action de la protéase, pourront être purifiées grâce à leurs étiquettes de clivage respectives.

Deux mutants ont également été clonés à partir de la construction MultiTEV. La réalisation de ceux-ci est détaillée ci-après (voir **Matériels et Méthodes §I.2.**).

Par la suite, la première construction MultiTEV (**Figure 34A**) a été modifiée afin de contenir un domaine réduit de Brd2 correspondant à ET (**Figure 33**), de façon à représenter un domaine d'interaction minimal avec l'IN de PERV-A/C. Cette seconde construction, nommée MultiTEV Minimal (**Figure 34B**) dans la suite du manuscrit, est ainsi identique à la construction MultiTEV à l'exception du gène codant Brd2 (667-808) qui est substitué par celui codant Brd2 (667-735).

Elle a été réalisée par l'équipe du Dr. Blanchard (Génétique Virale et Bio-sécurité, ANSES Ploufragan).

I. 1. 2. Construction plasmidique de la protéine Brd2 (667-808)

Afin de procéder à des tests d'interaction protéine-ligand avec le fragment Brd2 (667-808) isolé, la séquence codant ce fragment a été clonée dans un plasmide d'expression pPROEX-Hta où elle se trouve sous le contrôle d'un promoteur *trc*.

I. 1. 3. Les constructions plasmidiques de l'IN de PERV-A/C

Trois constructions plasmidiques de l'IN de PERV-A/C entière ont été réalisées au cours de mon projet de thèse, notamment en vue d'une étude structurale de l'intasome. Les vecteurs d'expression pET-28a et pPROEX-Hta ont été choisis de façon à générer des protéines de fusion portant une étiquette poly-histidine alors que le vecteur pHL8 a été choisi pour l'ajout d'une étiquette de solubilité GST (**Tableau 1**).

L'amplification du gène codant l'IN de PERV-A/C a été réalisée par PCR (pour *Polymerase Chain Reaction*) à partir du plasmide p3xFLAG-CMV contenant l'ADN d'intérêt et fourni par l'équipe du Dr. Ronfort. Les amorces utilisées (**Tableau 2**) servent également à l'ajout de sites de restriction aux deux extrémités du gène. Ces sites de restriction permettront le clonage du gène dans les différents vecteurs d'expression. Les amorces F1 et R1 sont utilisées pour le clonage dans le vecteur pET-28a, les amorces F2 et R2 pour le clonage dans le vecteur pPROEX-Hta et les amorces F3 et R3 pour le clonage dans le vecteur pHL8.

Les amplifications par PCR se font dans un mélange réactionnel composé de Q5[®] Reaction Buffer 1X final (*New England Biolabs*), 400 μ M de désoxyribonucléotides triphosphates (dNTP), 10 μ M d'amorces (**Tableau 2**), DMSO 3% final, 10 ng de plasmide portant le gène d'intérêt et 0,02 U/ μ L de Q5[®] High Fidelity ADN Polymerase (*New England Biolabs*). Les réactions ont été réalisées dans un thermocycleur GeneAmp[®] PCR System 9700 (*Applied Biosystems*) et se décomposent en trois étapes.

L'ADN est d'abord dénaturé à 98°C pendant 5 minutes afin d'obtenir de l'ADN monocaténaire à partir de la matrice bicaténaire. Il est ensuite soumis à 25 cycles d'amplification [i) dénaturation à 98°C pendant 30 secondes, ii) hybridation à 55°C pendant 30 secondes puis iii) élongation à 72°C pendant 1 minutes et 30 secondes] permettant une augmentation

exponentielle de la quantité d'ADN. Enfin, une élongation finale de 20 minutes est réalisée à 72°C.

Noms des amorces		Séquences (5' → 3')
1	F1	TAAGCA CATATG ATGATTATACACTGTCCTGGACATCAGAAAGCTA
	R1	TGCTTA GCGGCCGC TTACTTA ACTT GAGGAGTTATTGACAGAGTAAGGAACCA
2	F2	CG CCATGG CGATTATACACTGTCC
	R2	TT GAATTC TTA ACTT GAGGAGTTATTGACAG
3	F3	TAAGCA CATATG ATGATTATACACTGTCCTGGAC
	R3	TGCTTA GCGGCCGC CTA ACTT GAGGAGTTATTGACAGAGT

Tableau 2 | Séquences des amorces utilisées pour l'amplification de l'ADN codant l'IN PERV-A/C. La construction 1 correspond au vecteur pET-28a, la construction 2 au vecteur pPROEX-Hta et la 3 au vecteur PHL8. Les séquences annotées F sont les amorces sens et les séquences annotées R sont les amorces anti-sens. Les sites de restriction (NdeI : CATATG, NotI : GCGGCCGC, NcoI : CCATGG et EcoRI : GAATTC) sont en rouge et les codons stop en gras.

Les produits de PCR résultants sont déposés sur un gel d'agarose 1% (m/v) – TAE (Tris-Acétate/EDTA (pour *EthyleneDiamineTetraAcetic Acid*), 40 mM Tris-acétate pH 7,8, 1 mM EDTA), supplémenté en bromure d'éthidium, agent intercalant de l'ADN. La migration se fait dans ce même tampon TAE et les bandes d'ADN sont révélées aux ultraviolets. Celles-ci sont extraites et purifiées à l'aide du kit NucleoSpin® Gel and PCR Clean-up (*Macherey-Nagel*) suivant les indications du fournisseur.

L'ADN ainsi amplifié et purifié est dosé grâce à un spectrophotomètre NanoDrop (ND-100, *Thermo Scientific*), par mesure de l'absorbance à 260 nm, afin de déterminer sa concentration (voir **Matériels et Méthodes §I.5**).

Cet ADN bicaténaire néo-synthétisé est entouré de sites de restriction apportés par les amorces utilisées et correspondants à des sites uniques présents dans le MCS des plasmides d'expression dans lesquels se feront les clonages. L'IN de PERV-A/C est cloné entre les sites *NdeI* et *NotI* dans les plasmides pET-28a et PHL8 et entre les sites *NcoI* et *EcoRI* dans le plasmide pPROEX-Hta. Pour ce faire, l'insert amplifié par PCR et le vecteur sont chacun mis en

contact avec les deux enzymes de restriction adéquates. En pratique, pour réaliser cette double digestion enzymatique, 1 µg d'ADN sont placés dans un milieu réactionnel contenant 0,02 U/µL de chacune des deux enzymes de restriction (*New England Biolabs*) et le tampon de réaction optimal de celles-ci (ici CutSmart[®] Buffer 1X final, ou NEBuffer[™] 3.1 1X final, *New England Biolabs*). La digestion se fait à 37°C pendant une heure.

Les échantillons digérés sont à nouveau déposés sur gel agarose 1%, et la migration et purification des bandes d'ADN digérés se font comme décrit précédemment.

Enfin, la dernière étape consiste en la ligation de l'insert dans le vecteur plasmidique. Cette réaction est rendue possible grâce à la complémentarité des extrémités sortantes de l'insert d'ADN et du vecteur, créées par la digestion enzymatique. En pratique, l'insert et le vecteur sont mélangés à différents ratios (1:3, 1:1 et 3:1) et mélangé à 0,1 U/µL de l'ADN ligase du phage T4 (*New England Biolabs*) et à son tampon à une concentration de 1X final. L'ADN ligase permet la jonction de deux fragments d'ADN par création de liaisons phosphodiester covalentes. La réaction a lieu à 16°C pendant 16 heures. Le mélange réactionnel est utilisé pour une transformation en *Escherichia coli* (*E. coli*) TOP10 (voir **Matériels et Méthodes §I.3.**) permettant d'amplifier la nouvelle construction et d'isoler les clones la contenant.

I. 2. Mutagenèse dirigée par PCR

Deux mutants ont été réalisés à partir de la construction MultiTEV. Ces deux mutations concernent des acides aminés présents dans le CTD de l'IN de PERV-A/C qui semblent avoir un rôle dans l'interaction avec Brd2. Un mutant simple (W379A) et un double mutant (R391A/L392A) ont été construits (**Tableau 1**).

Ces mutations ont été introduites par mutagenèse dirigée par PCR, en se basant sur la méthode du QuickChange[™] (*Stratagene*) (Kirsch & Joly, 1998). Le principe de cette méthode repose sur le fait qu'une amorce nucléotidique ne nécessite pas une complémentarité absolue pour s'apparier à un ADN cible. Ainsi, il est possible de concevoir deux amorces, chacune portant les mutations désirées et complémentaires au brin opposé du vecteur.

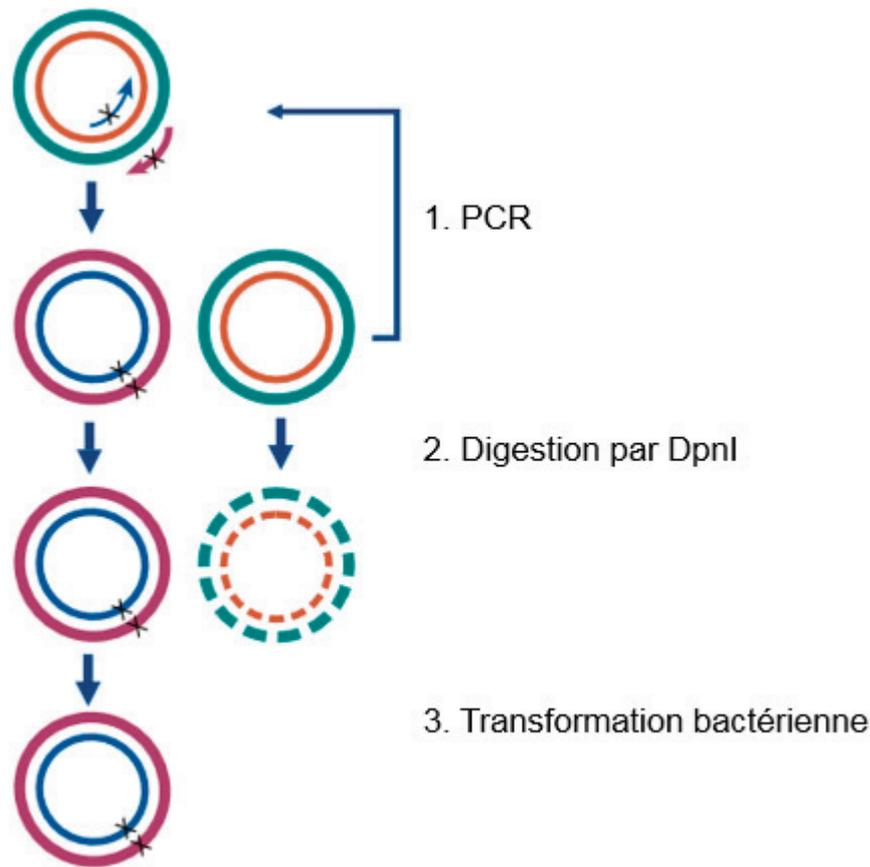


Figure 35 | Principe de la mutagenèse dirigée par PCR. L'ADN plasmidique parental est représenté en vert et orange, l'ADN synthétique en bleu et violet et la mutation insérée est représentée par une croix (d'après www.agilent.com).

La mutagenèse dirigée se déroule en trois étapes (**Figure 35**). La première étape consiste à appairer les deux amorces sur le plasmide cible et à amplifier celui-ci par PCR. L'extension des amorces entraîne l'apparition de brèches où la liaison phosphodiester entre le dernier nucléotide ajouté et l'extrémité 5' de l'amorce n'a pas été catalysée. Lors de la deuxième étape, l'ADN parental est digéré par l'endonucléase DpnI. Cette enzyme de restriction clive spécifiquement l'ADN méthylé, spécifique des plasmides isolés depuis une souche *E. coli*, permettant ainsi la conservation du plasmide synthétique dans lequel la mutation a été insérée. Enfin, le mélange résultant est utilisé pour une transformation bactérienne afin d'amplifier le plasmide muté.

En pratique, deux amorces sont conçues pour chaque construction mutée (**Tableau 3**). Sur chacune de ces amorces, deux nucléotides sont modifiés afin de changer le codon d'origine pour un codon qui sera traduit en alanine. Les amorces F4 et R4 sont utilisées pour construire le mutant W379A, où un tryptophane est muté en alanine et les amorces F5 et R5 sont utilisées

pour le double mutant R391A/L392, où une arginine et une leucine sont toutes deux mutées en alanine. Dans les trois cas, 200 ng du plasmide MultiTEV sont mélangés à 2 µM des amorces correspondantes ainsi que 200 µM de dNTP, le Q5® Reaction Buffer 1X final (*New England Biolabs*) et 0,02 U/µL de Q5® High Fidelity ADN Polymerase (*New England Biolabs*). La PCR est réalisée dans un thermocycleur GeneAmp® PCR System 9700 (*Applied Biosystems*) selon le protocole décrit précédemment (voir **Matériels et Méthodes §I.1.3.**).

Mutants	Noms des amorces	Séquences (5'→3')
W379A	F4	CCACCTCCCGATTCGGGGG GCG AAGCCGAAAAGACTGAAAA
	R4	TTTCAGTCTTTTCGGCTTT CGCC CGAATCGGGAGGTGG
R391A/L392A	F5	CTGAAAATCCCCTTAAGCTT GCCG CCATCGCGTGGTTCC
	R5	GGAACCACGCGATGG GCGGCA AGCTTAAGGGGATTTTCAG

Tableau 3 | Séquences des amorces utilisées pour la mutagenèse dirigée par PCR de la construction MultiTEV : le mutant simple W379A et le double mutant R391A/L392A. Les séquences annotées F sont les amorces sens et les séquences annotées R sont les amorces anti-sens. Les nucléotides mutés sont représentés en rouge.

Au produit de PCR sont ajoutés 0,02 U/µL de l'enzyme DpnI (*New England Biolabs*) et le mélange est incubé à 37°C pendant deux heures, pour une digestion optimale du plasmide d'origine.

La solution résultante ne contient plus que les plasmides mutés néo synthétisés et 10 µL sont amplifiés par transformation de 50 µL de bactérie *E. coli* DH5α (voir **Matériels et Méthodes §I.3.**). Chaque construction est vérifiée par séquençage de l'ADN pour confirmer la présence des mutations introduites (voir **Matériels et Méthodes §I.6.**).

I. 3. Transformation bactérienne

Les plasmides d'intérêt sont introduits dans différentes bactéries, préalablement rendues compétentes par traitement chimique au sein du laboratoire. Plusieurs souches ont été utilisées, toutes issues d'*E. coli*. Il s'agit des souches T7 Express, Rosetta, BL21(DE3)-pLysS et

C41(DE3) utilisées pour l'expression des protéines recombinantes et des souches DH5 α et TOP10 utilisées pour le clonage et la propagation plasmidique.

En pratique, 10 ng d'ADN plasmidique sont ajoutés à 50 μ L de bactéries compétentes. Après une incubation sur glace de 30 minutes, un choc thermique est réalisé en plaçant les bactéries à 42°C pendant 45 secondes puis à nouveau sur glace pendant 2 minutes. C'est ce changement de température qui permet de fragiliser la membrane des bactéries et d'incorporer le plasmide en son sein. À cette solution sont ajoutés 600 μ L de milieu *Lysogeny Broth* (LB) et le mélange est incubé à 37°C pendant une heure, sous agitation à 700 rpm. Les bactéries sont ensuite sédimentées par centrifugation et une partie du surnageant est retiré pour laisser un volume final de 200 μ L. Les cellules sont re-suspendues et la solution est étalée stérilement sur milieu solide LB-agar (*Sigma*) contenant l'antibiotique nécessaire à la sélection des clones transformés. Ces derniers sont visibles sur le milieu sélectif après une nuit d'incubation à 37°C.

I. 4. Extraction de l'ADN plasmidique

L'extraction de l'ADN plasmidique d'intérêt à partir des bactéries transformées est possible avec le kit commercial NucleoSpin[®] Plasmid (*Macherey Nagel*) qui permet une lyse alcaline (Birnboim & Doly, 1979).

Les bactéries DH5 α transformées par le plasmide d'intérêt sont mises en culture dans un milieu contenant l'antibiotique approprié pour sa sélection. Après une nuit d'incubation à 37°C, un volume dépendant du taux de réplication du plasmide et compris entre 1 et 10 mL est prélevé et utilisé pour l'extraction. L'échantillon est centrifugé pendant 30 secondes à 11 000 \times g pour sédimenter les bactéries et l'ADN plasmidique est extrait selon le protocole du fournisseur.

I. 5. Dosage de l'ADN

La quantité d'ADN peut être dosée par mesure d'absorbance à 260 nm (A_{260}) avec un spectrophotomètre Nanodrop2000 (*Thermo Fisher Scientific*). En effet, l'ADN a la capacité d'absorber la lumière ultraviolette entre 240 et 280 nm grâce à ses bases possédant un noyau aromatique. La concentration d'ADN peut être déterminée par la loi de Beer-Lambert :

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

Avec A_{260} , l'absorbance à 260 nm ; ε , le coefficient d'extinction molaire ; L , la longueur du trajet optique et c , la concentration.

De plus, la qualité de l'échantillon d'ADN peut être estimée en mesurant le rapport d'absorbance à 260 nm et 280 nm (A_{260}/A_{280}). Pour un ADN double brin pur, ce rapport à une valeur proche de 1,8. Si le rapport A_{260}/A_{280} est supérieur à 1,8, l'échantillon est contaminé par de l'ARN et, au contraire, s'il est inférieur à 1,8, l'échantillon est contaminé par des protéines.

I. 6. Séquençage de l'ADN

La méthode de Sanger est utilisée pour déterminer la séquence d'un ADN donné (Sanger *et al.*, 1977). Celle-ci utilise des didésoxyribonucléotides triphosphates (ddNTP) portant un fluorophore spécifique à chacun des quatre nucléotides.

Un mélange réactionnel contenant l'ADN d'intérêt, une amorce spécifique à celui-ci, une polymérase, des dNTP et des ddNTP est soumis à une PCR. La polymérase amplifie alors l'ADN en utilisant de façon aléatoire les dNTP, présents en excès dans le milieu, et les ddNTP, plus rares. Lors de l'incorporation de l'un de ces derniers, l'élongation du brin est interrompue et l'ADN est marqué par le fluorophore porté par le ddNTP. Des fragments de différentes tailles sont ainsi synthétisés par les nombreuses réactions, chacun portant le marquage spécifique à son dernier nucléotide. Les chaînes d'ADN sont alors séparées par électrophorèse en fonction de leur taille et un détecteur automatique permet de repérer les fluorophores et, par extension, l'ordre des nucléotides.

L'ensemble des séquençages d'ADN réalisés au cours de ma thèse ont été fait par la société *Eurofins Genomics*.

II. Surproduction et purification des protéines d'intérêt

II. 1. MultiTEV / MultiTEV Minimal

II. 1. 1. Surproduction en système procaryote

L'expression des gènes des constructions MultiTEV et MultiTEV Minimal se fait en système procaryote, dans la souche bactérienne *E. coli* T7 Express (**Figure 36**). Les gènes sont

sous le contrôle du promoteur T7 qui est reconnu spécifiquement par l'ARN polymérase ADN-dépendante du bactériophage T7 codée par le lysogène DE3, contenu dans cette souche bactérienne (Studier & Moffatt, 1986). Ce dernier, placé sous le contrôle du promoteur de l'opéron lactose *LacUV5* est inhibé par le répresseur LacI. L'ajout d'IsoPropylThioGalactoside (IPTG, *Euromedex*), un analogue non hydrolysable du lactose, dans le milieu de culture, permet la levée de l'inhibition par LacI, la synthèse de l'ARN polymérase T7 et la transcription du gène d'intérêt.

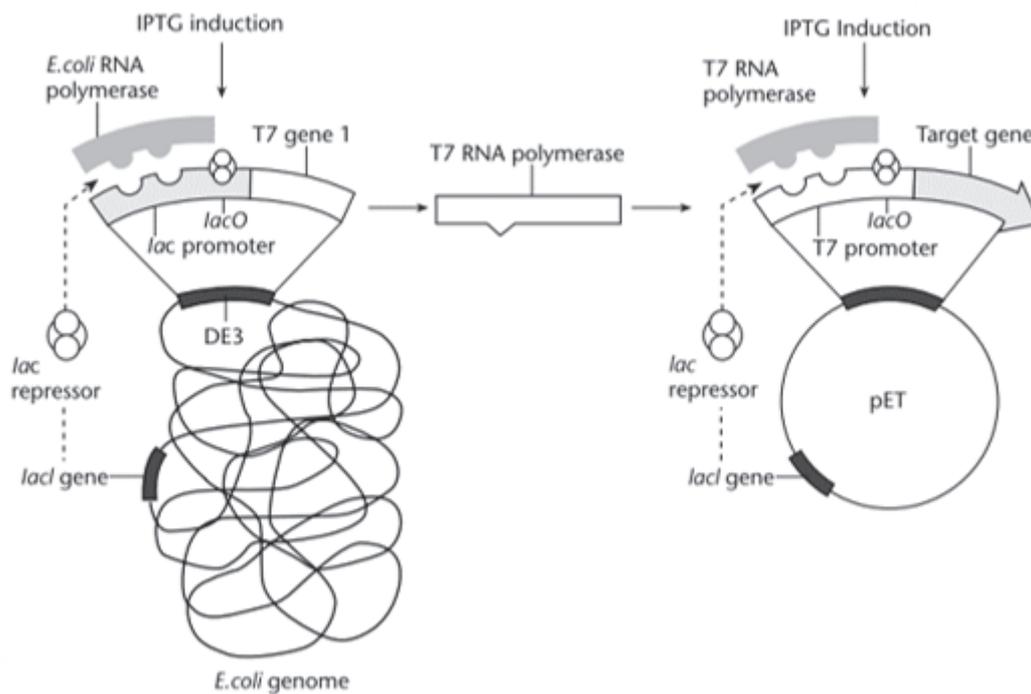


Figure 36 | Régulation des gènes clonés dans un vecteur pET (d'après (Primrose & Twyman, 2006)).

En pratique, après transformation des bactéries par le plasmide d'intérêt, un clone est ensemencé dans une préculture de milieu LB en présence de kanamycine (50 µg/mL final). Celle-ci est incubée à 37°C, sous agitation à 180 rpm, pendant une nuit. Après incubation, la préculture est diluée au 100^{ème} dans du milieu LB contenant l'antibiotique et est placée à 37°C, sous agitation à 180 rpm. La croissance bactérienne est suivie par spectrophotométrie à une longueur d'onde de 600 nm, permettant une mesure de la turbidité du milieu. Lorsque la densité optique atteint une valeur comprise entre 0,5 et 0,9, signifiant que la phase exponentielle de croissance a été atteinte, la surproduction protéique est induite par ajout d'IPTG (1 mM final, *Euromedex*). Après trois heures d'incubation à 37°C, sous agitation à 180 rpm, les bactéries sont

sédimentées par centrifugation pendant 10 minutes à $6\,000 \times g$, à 4°C . Le surnageant est retiré et le culot cellulaire est stocké à -20°C jusqu'à l'étape de lyse bactérienne.

II. 1. 2. Lyse bactérienne

Le culot bactérien issu d'un litre de culture est re-suspendu dans 20 mL de tampon de lyse (20 mM Tris HCl pH 8, 500 mM NaCl, 25 mM imidazole). Du lysozyme (0,2 mg/mL, *Sigma*) facilitant la lyse, et un cocktail d'inhibiteurs de protéases 1X (Complete ULTRA EDTA Free, *Roche*), protégeant les protéines d'intérêt d'une protéolyse par les enzymes bactériennes et donc leur dégradation, sont ajoutés au tampon de lyse. Ce mélange est placé à 4°C pendant une heure et des brassages fréquents sont réalisés.

La lyse mécanique des bactéries re-suspendues se fait par sonication à l'aide de l'appareil Sonifier 250 (*Branson*) émettant des ultrasons. Huit cycles de pulsations d'une minute à une puissance de 40% suivie d'une minute d'incubation sur glace sont nécessaires à la lyse complète de l'échantillon.

Le lysat bactérien est ensuite centrifugé à $14\,000 \times g$ pendant 45 minutes, à 4°C . Les débris cellulaires (fraction insoluble) sont éliminés dans le culot alors que le surnageant correspondant à la fraction soluble contient les protéines d'intérêt. Ce dernier est filtré avec un filtre de porosité de $0,45\ \mu\text{m}$ (*VWR Internationals*). Le lysat et la fraction insoluble sont conservés à 4°C .

II. 1. 3. Purification des protéines d'intérêt

La stratégie adoptée pour purifier les complexes d'intérêt repose sur la différence des étiquettes de purification présentes aux extrémités amino-terminales des fragments de l'IN PERV-A/C et de Brd2, portant respectivement des étiquettes Strep-tag[®] et poly-histidine. Cette stratégie consiste, dans un premier temps, à purifier Brd2 par chromatographie d'affinité pour le nickel en conditions natives. Cette première étape permet de conserver Brd2 ainsi que tous les complexes protéiques auxquels il participe. Dans un second temps, ce mélange protéique est soumis à une chromatographie d'affinité pour la streptavidine de façon à purifier les complexes protéiques portant l'étiquette Strep-tag[®] et donc l'IN PERV-A/C.

Ces deux systèmes de chromatographie sont couplés à un système de chromatographie liquide des protéines à basse pression (FPLC, pour *Fast Protein Liquid Chromatography*) (Aktä Purifier, *GE Healthcare*).

II. 1. 3. 1. Chromatographie d'affinité pour le nickel

La chromatographie d'affinité pour le nickel se base sur la capacité qu'ont les résidus d'histidine de l'étiquette de purification à établir des liaisons de coordination avec les ions nickel. Ces ions sont immobilisés sur une résine Ni-NTA, composée de bras espaceurs constitués d'acide nitrilotriacétique (NTA, pour *NitriloTriacetic Acid*) (Protino[®] Ni-NTA 5 mL, *Macherey-Nagel*). Chaque ion Ni²⁺ établit six liaisons de coordination, deux le reliant aux atomes d'azote du cycle imidazole de la chaîne latérale de deux résidus d'histidine et quatre assurant la liaison au bras NTA (**Figure 37**). Les protéines d'intérêt liées par leur étiquette poly-histidine sont ensuite éluées par l'ajout d'un compétiteur, l'imidazole, présent en excès dans le tampon d'éluion.

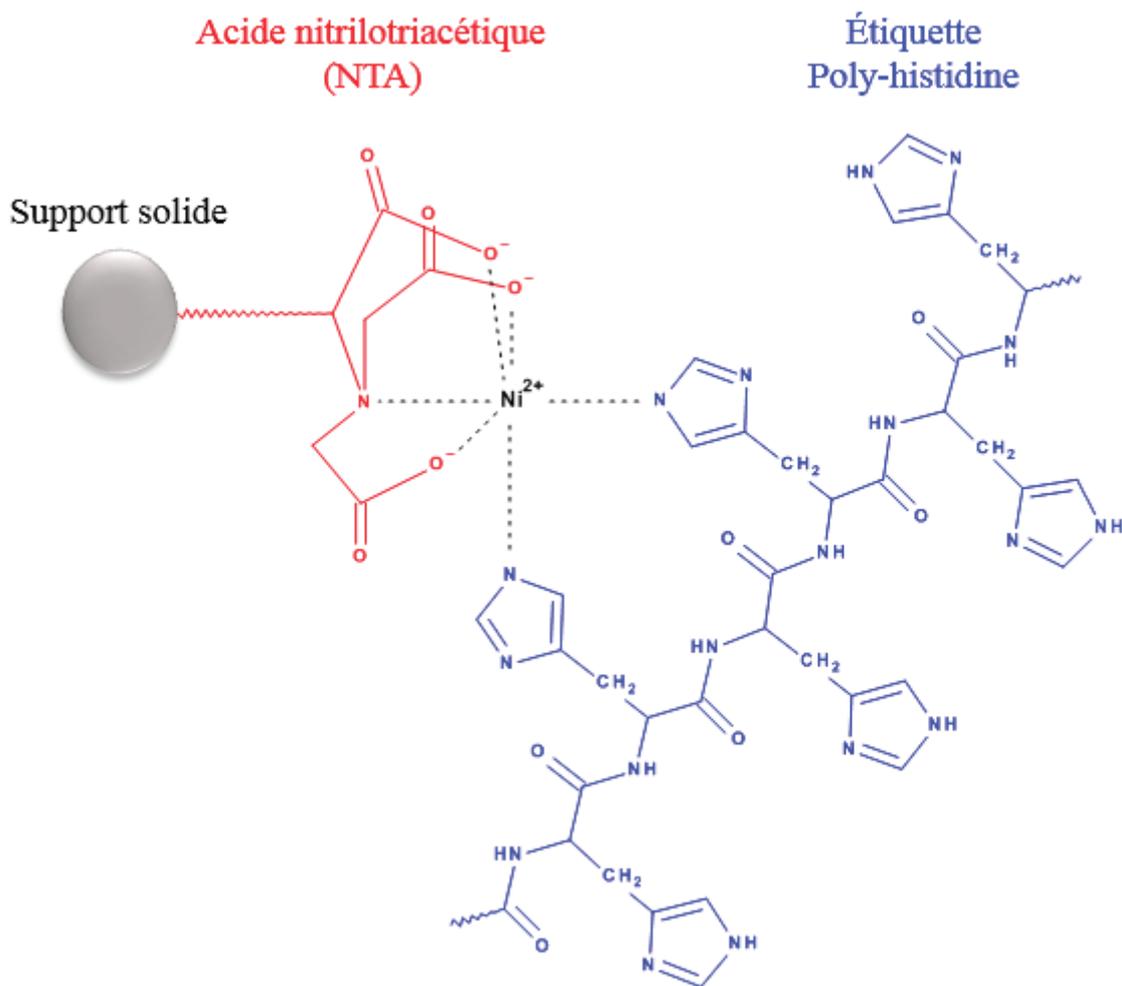


Figure 37 | Liaisons de coordination établies entre les résidus histidyl de l'étiquette poly-histidine, l'acide nitrilotriacétique et l'ion nickel, lors d'une chromatographie d'affinité pour le nickel (d'après (Magdeldin & Moser, 2012)).

En pratique, la purification du complexe protéique est réalisée à partir du lysat bactérien filtré. Celui-ci est chargé sur la colonne Ni-NTA, préalablement équilibrée avec le tampon de lyse. Afin de limiter les fixations aspécifiques de protéines, ce tampon contient de l'imidazole en faible quantité de façon à éliminer les protéines non fixées à la colonne. Cette dernière est ensuite lavée avec un tampon de lavage ayant une force ionique importante (20 mM Tris HCl pH 8, 1 M NaCl, 25 mM imidazole), afin d'éluer les protéines et les acides nucléiques fixés de façon aspécifique ou avec une très faible affinité à la colonne. Après un second lavage avec le tampon de lyse, l'élution est réalisée. Celle-ci se fait avec un gradient croissant d'imidazole (0,025 à 1 M imidazole), grâce à un mélange progressif des tampons de lyse et d'élution (20 mM Tris HCl pH 8, 500 mM NaCl, 1 M imidazole). La sortie de colonne des protéines éluées est détectée par un suivi de l'absorbance à 280 nm. Les différentes fractions d'élution sont ensuite récupérées et analysées par électrophorèse en gel discontinu de polyacrylamide en conditions dénaturantes (SDS-PAGE, pour *Sodium Dodecyl Sulfate - PolyAcrylamide Gel Electrophoresis*).

II. 1. 3. 2. Chromatographie d'affinité pour la streptavidine

La chromatographie d'affinité pour la streptavidine se base sur l'affinité de l'étiquette Strep-tag[®] pour le composé Strep-Tactin[®], une streptavidine modifiée, fixé à la résine de la colonne (Strep-Tactin[®] Superflow[®] cartridge 5 mL, IBA) (**Figure 38**). L'étiquette Strep-tag[®] est un peptide synthétique composé de huit résidus d'acides aminés, qui a la capacité de lier la streptavidine modifiée sur le site de fixation de la biotine, son substrat naturel. Les modifications de la streptavidine pour former le composé Strep-Tactin[®] lui confèrent une grande stabilité et affinité pour cette étiquette. L'interaction entre le Strep-Tactin[®] et l'étiquette Strep-tag[®] est réversible et peut être rompue par ajout de desthiobiotine, un dérivé de la biotine.

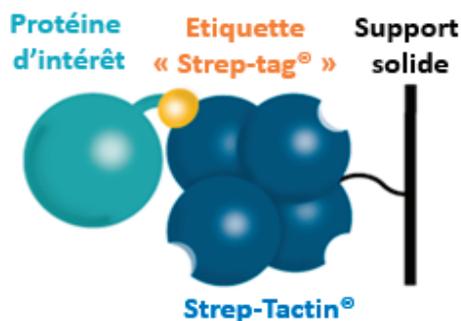


Figure 38 | Principe général de la chromatographie d'affinité pour la streptavidine, avec la colonne Strep-Tactin[®] (d'après <http://www.iba-lifesciences.com/strep-tag.html>).

En pratique, l'échantillon est injecté sur la colonne qui est ensuite lavée avec un tampon W (100 mM Tris HCl pH 8, 150 mM NaCl, 1 mM EDTA). L'élution des protéines est ensuite réalisée par compétition avec la desthiobiotine grâce au tampon E (100 mM Tris HCl pH 8, 150 mM NaCl, 1 mM EDTA, 2,5 mM desthiobiotine).

II. 1. 3. 3. Chromatographie d'exclusion de taille

La chromatographie d'exclusion de taille (SEC, pour *Size Exclusion Chromatography*) est une technique permettant la séparation de macromolécules en fonction de leur poids moléculaire et de leur forme. Elle est couplée à un système FPLC et deux colonnes Superdex ont été utilisées pour ces travaux : les colonnes Superdex 200 Increase 10/300 GL et Superdex S200 5/150 GL (*GE Healthcare*), adaptées pour une séparation optimale dans la gamme 10-600 kDa. La colonne est constituée de granules de gel poreux, formant la phase solide du système. De cette façon, les grosses molécules dont le diamètre est supérieur à celui des pores sont exclues et éluées en premier. Au contraire, les molécules de moyenne et petite tailles voient leur migration ralentie par leur passage dans les pores. La relation linéaire entre le volume d'élution et le logarithme du poids moléculaire de macromolécules connues permet d'établir une courbe de calibration spécifique à la colonne. Celle-ci peut alors être utilisée pour estimer le poids moléculaire et l'état oligomérique des protéines analysées.

La colonne est préalablement équilibrée dans tampon permettant d'éluer les protéines d'intérêt dans leur tampon de stockage définitif. Deux tampons sont utilisés pour les complexes MultiTEV et MultiTEV Minimal : un tampon final (20 mM Tris HCl pH 8, 500 mM NaCl) et un tampon BicAm (200 mM bicarbonate d'ammonium pH 8). La solution protéique concentrée est injectée dans un faible volume (400 μ L). L'élution des macromolécules est suivie par mesure de l'absorbance à 280 nm (A_{280}).

II. 1. 4. Concentration protéique

Les échantillons protéiques sont concentrés par centrifugation à l'aide d'un système d'ultrafiltration. Celui-ci est un concentrateur contenant une membrane semi-perméable de PolyEther Sulfone (PES, *Sartorius*). Cette membrane permet d'obtenir des seuils de coupure différents, séparant les protéines selon leur poids moléculaire. Les particules de plus faible poids moléculaire et le tampon sont filtrés, de façon à augmenter la concentration des protéines

d'intérêt. Des cycles de centrifugation à $5\,000 \times g$ et à 4°C sont réalisés, jusqu'à l'obtention de la concentration protéique attendue.

II. 2. Brd2 (667-808)

II. 2. 1. Surproduction en système procaryote

L'expression du fragment Brd2 (667-808) se fait dans la souche *E. coli* Rosetta, modifiée pour faciliter l'expression des protéines eucaryotes. Cette souche possède des caractéristiques similaires à celles de la souche *E. coli* T7 Express et est également inductible à l'IPTG (voir **Matériels et Méthode §II.1.1.**).

En pratique, après transformation des bactéries par le plasmide d'intérêt, une dizaine de clones sontensemencés dans une préculture de milieu LB en présence d'ampicilline ($50\ \mu\text{g}/\text{mL}$ final). Celle-ci est incubée pendant quatre heures, à 37°C sous agitation à 180 rpm. La préculture est ensuite diluée au $40^{\text{ème}}$ dans du milieu LB contenant l'antibiotique approprié et du glucose (1% final (w/v)), et est placée à 37°C , sous agitation à 180 rpm. Des mesures par spectrophotométrie sont réalisées pour suivre la croissance bactérienne et la surproduction protéique est induite par ajout d'IPTG (1 mM final) lorsque la valeur de la densité optique atteint 0,6. Après deux heures d'incubation à 37°C , sous agitation à 180 rpm, les bactéries sont sédimentées par centrifugation pendant 10 minutes à $6\,000 \times g$, à 4°C . Le culot cellulaire est stocké à -20°C en attente de l'étape de lysé bactérienne.

II. 2. 2. Lyse bactérienne

Le culot bactérien résultant d'un litre de culture est re-suspendu dans 10 mL de tampon A (50 mM Tris HCl pH 7,5, 1 M NaCl, 20 mM imidazole, 0,1 mM EDTA, 2 mM β -mercapto-éthanol, 1% glycérol) par gramme de culot. Après ajout de lysozyme (0,2 mg/mL, *Sigma*) et de cocktail d'inhibiteurs de protéases 1X (Complete ULTRA EDTA Free, *Roche*), le mélange est incubé à 4°C .

Comme décrit dans le paragraphe **Matériels et Méthodes §II.1.2.**, la lyse mécanique est réalisée par sonication. Deux cycles de pulsations à une puissance de 50% d'une durée de deux minutes, séparés par une incubation sur glace de deux minutes, sont appliqués sur l'échantillon. Le lysat bactérien est centrifugé à $11\,000 \times g$ pendant 45 minutes, à 4°C . La fraction soluble contenant la protéine d'intérêt est conservée à 4°C .

II. 2. 3. Purification de Brd2 (667-808)

La purification de Brd2 (667-808) est possible grâce à l'étiquette poly-histidine portée par cette protéine à son extrémité amino-terminale. Un système de chromatographie d'affinité pour le nickel (voir **Matériels et Méthodes §II.1.3.1.**), couplé à un système FPLC est utilisé.

En pratique, le lysat bactérien filtré est chargé sur la colonne Ni-NTA (Protino[®] Ni-NTA 5 mL, *Macherey-Nagel*) préalablement équilibrée avec le tampon A et toutes les protéines non fixées à la colonne sont éluées avec ce même tampon. L'éluion est réalisée avec un gradient croissant d'imidazole (0,02 à 0,5 M imidazole), à l'aide d'un mélange des tampons A et B (50 mM Tris HCl pH 7,5, 0,5 M NaCl, 0,5 M imidazole, 0,1 mM EDTA, 2 mM β -mercapto-éthanol, 1% glycérol). L'éluion des protéines est détectée par un suivi de l'absorbance à 280 nm et les fractions d'éluion sont analysées par SDS-PAGE.

II. 2. 4. Clivage de l'étiquette d'affinité par la protéase du TEV

Afin d'éviter que l'étiquette de purification poly-histidine portée par Brd2 (667-808) n'interfère dans l'interaction avec l'IN de PERV-A/C, celle-ci est clivée par digestion enzymatique. Un site de clivage à la protéase du TEV se situe entre la protéine et son étiquette pour permettre la digestion. Cette protéase reconnaît spécifiquement les séquences ENLYFQG et ENLYFQS et clive entre les résidus Q et G ou Q et S (Carrington & Dougherty, 1988). La protéase est surproduite en fusion avec une étiquette poly-histidine au sein du laboratoire.

En pratique, la solution protéique est incubée avec la protéase du TEV dans un boudin de dialyse (SnakeSkin[™] Dialyses Tubing 7 000 MWCO, *Thermo Scientific*) avec un rapport massique de 1:20 (TEV:Brd2). Le boudin de dialyse est placé dans un litre de tampon 50 mM Tris HCl pH 7,5, 500 mM NaCl et 5 mM β -mercapto-éthanol, de façon à s'affranchir de l'imidazole, et mis à 4°C sous agitation douce pendant une nuit. Après digestion, l'étiquette poly-histidine clivée et la protéase du TEV sont éliminés par chromatographie d'affinité pour le nickel et la digestion est contrôlée par SDS-PAGE.

II. 3. L'intégrase entière de PERV-A/C

II. 3. 1. Tests de surproduction en système bactérien

Des tests d'expression ont été réalisés de façon à déterminer la meilleure condition de surproduction de l'IN de PERV-A/C soluble. Pour cela, plusieurs facteurs ont été variés afin de voir leurs effets sur l'expression du gène.

Le protocole utilisé pour chacun de ces tests est le même : 50 µL de bactéries compétentes sont transformées avec 100 ng de plasmide recombinant comme décrit dans le **Matériels et Méthodes §I.3.**, avant d'être étalées sur une boîte de LB-agar contenant l'antibiotique adapté. Les colonies ayant poussées sont repiquées et utilisées pour réaliser une préculture de 5 mL de milieu LB supplémenté en antibiotique, puis incubées à 37°C pendant 4h. Cette préculture sert ensuite à ensemencer au 40^{ème} du milieu de culture LB, en présence de l'antibiotique approprié. L'expression du gène est induite par ajout d'IPTG lorsque la valeur de la densité optique de la culture atteint 0,6. Enfin, les bactéries sont sédimentées par centrifugation à 6 000 × g pendant 10 minutes à 4°C. Le culot bactérien est conservé à -20°C en attendant l'étape de lyse cellulaire.

À partir de ce protocole, plusieurs conditions d'expression ont été testées. Tout d'abord, différentes bactéries d'expression ont été utilisées : *E. coli* T7 Express et B121(DE3)-pLysS, *E. coli* Rosetta utilisée pour l'expression des codons rares et *E. coli* C41(DE3) utilisée pour l'expression des protéines toxiques (Dumon-Seignovert *et al.*, 2004).

La concentration finale en IPTG (de 0,1 mM à 1 mM), la température (de 18°C à 37°C) et le temps d'incubation (de 1h à 16h) sont également des facteurs qui ont été variés afin d'optimiser le protocole. La variation de ces facteurs peut permettre la diminution de la formation de corps d'inclusion ou encore une production plus lente et contrôlée de la protéine d'intérêt

Différentes ajouts de molécules dans le milieu de culture ont également été testés. Tout d'abord, la transcription peut être contrôlée avec l'ajout de 100 µg/mL de rifampicine une heure après le début de l'induction. Cet antibiotique inhibe la croissance bactérienne en se fixant à l'ARN polymérase bactérienne, ceci, bloquant la transcription. L'ARN polymérase T7 qui assure la transcription du gène recombinant n'est pas affectée et l'expression de ce dernier peut continuer (Fiorini *et al.*, 2008). Ainsi, si la protéine est nocive, celle-ci n'aura pas d'effet sur la bactérie inhibée.

Un second essai concerne l'ajout d'ions Zn^{2+} dans le milieu de culture en raison du motif HHCC présent dans le NTD de l'IN qui est proche d'un domaine en doigt de zinc (Matsuda *et al.*, 2006). Celui-ci aide au repliement de la protéine.

Enfin, l'ajout d'1% de glucose dans le milieu de culture permet de diminuer l'expression basale de la protéine d'intérêt dans la bactérie (Rosano & Ceccarelli, 2014) afin de limiter la fuite d'expression et de contrôler la surproduction lors de l'induction.

II. 3. 2. Lyse bactérienne

Les culots bactérien précédemment sédimentés sont décongelés et soumis à une lyse cellulaire pour permettre l'extraction de la protéine d'intérêt.

Dans l'optique d'améliorer la solubilité de la protéine, différents tampons de lyse ont été testés (**Tableau 4**). Pour chacun d'eux, le pH, la concentration en sels, les détergents et les additifs ont été variés de façon à trouver la condition idéale pour l'extraction des protéines produites.

Les culots sont re-suspendus dans 10 μ L de tampon de lyse par microgramme d'échantillon. Après ajout d'un mélange d'inhibiteurs de protéases 1X (Complete ULTRA EDTA Free, Roche), le mélange est incubé 30 minutes à 4°C.

Les cellules sont lysées par sonication, comme décrit dans **Matériels et Méthodes §II.1.2.**, avec quatre cycles de pulsations à 40% de puissance pendant 30 secondes, séparés par une incubation sur glace. Le lysat bactérien obtenu est centrifugé à $11\ 000 \times g$ pendant 30 minutes à 4°C. Le surnageant, correspondant à la fraction soluble, est conservé et le culot, correspondant à la fraction insoluble, est re-suspendu dans 200 μ L de tampon de lyse. Ces échantillons sont réservés afin d'être analysés par SDS-PAGE.

Tampons de lyse	Composition	Tampons de lyse	Composition
L1	Tris HCl pH 8 20 mM NaCl 500 mM Imidazole 25 mM	L2	Tris HCl pH 8 50 mM NaCl 1 M Imidazole 20 mM β -mercaptoéthanol 2 mM Glycérol 10%
L3	HEPES pH 7,5 20 mM NaCl 1 M Imidazole 5 mM NP-40 0,05% Glycérol 10% β -mercaptoéthanol 2 mM	L4	HEPES pH 7,5 20 mM NaCl 1 M Imidazole 5 mM NP-40 0,05% Glycérol 10% β -mercaptoéthanol 2 mM $MgCl_2$ 2,5 mM $CaCl_2$ 0,5 mM
L5	HEPES pH 7,5 20 mM NaCl 1 M Imidazole 5 mM NP-40 0,1% Glycérol 10% β -mercaptoéthanol 5 mM $MgCl_2$ 2,5 mM $CaCl_2$ 0,5 mM	L6	Tris HCl pH 7,5 50 mM NaCl 1 M Imidazole 20 mM β -mercaptoéthanol 2 mM EDTA 0,1 mM Glycérol 10%
L7	Tris HCl pH 7,5 50 mM NaCl 1 M Imidazole 20 mM β -mercaptoéthanol 2 mM EDTA 0,1 mM Glycérol 10% CHAPS 1 mM	L8	Tris HCl pH 7,5 50 mM NaCl 1 M Imidazole 20 mM β -mercaptoéthanol 2 mM EDTA 0,1 mM Glycérol 10% CHAPS 1 mM $MgCl_2$ 1 mM $ZnSO_4$ 1 μ M ATP 2 mM
L9	Tris HCl pH 8,5 50 mM NaCl 1 M Imidazole 20 mM β -mercaptoéthanol 2 mM EDTA 0,1 mM Glycérol 10% CHAPS 1 mM $MgCl_2$ 1 mM $ZnSO_4$ 1 μ M ATP 2 mM	L10	Tris HCl pH 6,5 50 mM NaCl 1 M Imidazole 20 mM β -mercaptoéthanol 2 mM EDTA 0,1 mM Glycérol 10% CHAPS 1 mM $MgCl_2$ 1 mM $ZnSO_4$ 1 μ M ATP 2 mM

Tableau 4 | Compositions des tampons testés pour la lyse bactérienne. HEPES : acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique ; EDTA : *ethylenediaminetetraacetic acid* ; CHAPS : 3-[(3-cholamidopropyl)diméthylammonio]-1-propanesulfonate ; ATP : adénosine triphosphate.

II. 3. 3. Solubilisation de l'IN de PERV-A/C

Dans certains cas, la surproduction de protéines en système procaryote peut mener à leur agrégation sous la forme de corps d'inclusion. Ce phénomène est dû au repliement partiel ou incorrect de la protéine et entraîne la production de protéines insolubles. Plusieurs protocoles de solubilisation par dénaturation par l'urée ont été testés.

Le premier se base sur la méthode du *Single freeze-thawing cycle* mise au point pour réduire l'agrégation tout en facilitant le repliement protéique (Qi *et al.*, 2015). Le culot bactérien est repris dans 20 mL de tampon L5 en présence du mélange d'inhibiteurs de protéase 1X avant d'être soniqué quatre fois à une puissance de 50% pendant 30 secondes. L'échantillon est centrifugé 20 minutes à $12\,000 \times g$, à 4°C. Le culot contenant les corps d'inclusion est re-suspendu deux fois dans un tampon identique contenant 1 M d'urée et centrifugé à chaque cycle. Après chaque étape, une aliquote du surnageant est conservée. Enfin, l'échantillon est repris dans un tampon L5 contenant 2 M d'urée et placé pendant une nuit à -20°C. Le lendemain, ce dernier est décongelé et centrifugé, puis les différentes fractions sont analysées par SDS-PAGE.

La deuxième méthode utilisée consiste à lyser les bactéries dans un tampon contenant de l'urée. Les culots sont re-suspendus dans un tampon L7 contenant 2 à 8 M d'urée, en présence du cocktail d'inhibiteurs de protéases 1X. Après une heure d'incubation à 4°C, l'échantillon est soniqué à une puissance de 50% pendant deux cycles de 2 minutes. Enfin, le lysat est centrifugé 45 minutes à $11\,000 \times g$, à 4°C et le surnageant est conservé en vue de l'étape de purification.

II. 3. 4. Purification de l'IN de PERV-A/C

L'IN de PERV-A/C porte une étiquette poly-histidine à son extrémité amino-terminale qui permet sa purification. Pour cela, un système de chromatographie d'affinité pour le nickel (voir **Matériels et Méthodes §II.1.3.1.**), couplé à un système FPLC est utilisé.

En pratique, le lysat bactérien filtré est chargé sur la colonne Ni-NTA (Protino[®] Ni-NTA 5 mL, *Macherey-Nagel*) préalablement équilibrée avec le tampon utilisé pour la lyse (**Tableau 4**). Un premier lavage de la colonne est réalisé avec un tampon identique au tampon de lyse et supplémenté avec 2 mM d'ATP. Cette étape permet de décrocher les protéines chaperonnes cellulaires et donc de s'affranchir d'un nombre de contaminants. L'élution est ensuite réalisée avec une première étape isocratique à 0,1 M d'imidazole, suivie d'un gradient de

0,1 à 0,5 M d'imidazole réalisé à l'aide des tampons de lyse et tampons d'éluion correspondants (Tableau 5).

Tampons d'éluion	Composition	Tampons d'éluion	Composition
E2	Tris HCl pH 8 50 mM NaCl 200 mM Imidazole 0,5 M β -mercaptoéthanol 2 mM Glycérol 10%	E3	HEPES pH 7,5 20 mM NaCl 1 M Imidazole 0,5 M NP-40 0,05% Glycérol 10% β -mercaptoéthanol 2 mM
E4	HEPES pH 7,5 20 mM NaCl 1 M Imidazole 0,5 M NP-40 0,05% Glycérol 10% β -mercaptoéthanol 2 mM MgCl ₂ 2,5 mM CaCl ₂ 0,5 mM	E5	HEPES pH 7,5 20 mM NaCl 1 M Imidazole 0,5 M NP-40 0,1% Glycérol 10% β -mercaptoéthanol 5 mM MgCl ₂ 2,5 mM CaCl ₂ 0,5 mM
E7	Tris HCl pH 7,5 50 mM NaCl 0,5 M Imidazole 0,5 M β -mercaptoéthanol 2 mM EDTA 0,1 mM Glycérol 10% CHAPS 1 mM	E8	Tris HCl pH 7,5 50 mM NaCl 1 M Imidazole 0,5 M β -mercaptoéthanol 2 mM EDTA 0,1 mM Glycérol 10% CHAPS 1 mM MgCl ₂ 1 mM ZnSO ₄ 1 μ M
E9	Tris HCl pH 8,5 50 mM NaCl 1 M Imidazole 0,5 M β -mercaptoéthanol 2 mM EDTA 0,1 mM Glycérol 10% CHAPS 1 mM MgCl ₂ 1 mM ZnSO ₄ 1 μ M	E10	Tris HCl pH 6,5 50 mM NaCl 1 M Imidazole 0,5 M β -mercaptoéthanol 2 mM EDTA 0,1 mM Glycérol 10% CHAPS 1 mM MgCl ₂ 1 mM ZnSO ₄ 1 μ M

Tableau 5 | Compositions des tampons d'éluion testés pour la purification de l'IN de PERV-A/C. HEPES : acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique ; EDTA : *ethylenediaminetetraacetic acid* ; CHAPS : 3-[(3-cholamidopropyl)diméthylammonio]-1-propanesulfonate.

Dans les cas où une solubilisation à l'urée a été réalisée pendant la lyse, plusieurs suites sont envisagées lors de la purification. La première consiste à appliquer un gradient d'urée simultanément au gradient d'imidazole permettant l'élution de la protéine d'intérêt. Ce gradient permet ainsi un repliement lent de la protéine qui facilite la formation de structures secondaires au fur et à mesure de la diminution de concentration d'urée.

Une deuxième possibilité consiste à éluer la protéine d'intérêt dans un tampon dont la concentration d'urée n'a pas été modifiée, tout en suivant le protocole détaillé ci-avant. De cette façon, la protéine purifiée est maintenue dans des conditions dénaturantes et il est nécessaire d'éliminer l'urée présente dans le tampon. Cela se fait par dialyse avec l'ajout d'ADN ou de partenaire cellulaire pour favoriser le repliement et stabiliser la structure (voir **Matériels et Méthodes §II.3.6.**).

L'élution des protéines est détectée par un suivi de l'absorbance à 280 nm et les fractions d'élution sont analysées par SDS-PAGE.

II. 3. 5. Préparation du substrat d'ADN de l'intasome

De façon à assembler l'intasome de PERV-A/C, un ADN mimant le produit de l'intégration concertée a été produit. Celui-ci a été réalisé sur la base de l'ADN utilisé pour la résolution de la structure de l'intasome de RSV (Yin *et al.*, 2016) et consiste en l'hybridation de trois oligonucléotides formant deux doubles hélices d'ADN séparées par un court ADN simple brin (**Figure 39**).

La première double hélice consiste en vingt paires de bases représentant l'extrémité du LTR de PERV-A/C et la seconde mime l'ADN cellulaire, ciblé par l'IN.



Figure 39 | Substrat d'ADN utilisé pour l'assemblage de l'intasome de PERV-A/C. Il est formé par l'hybridation de trois oligonucléotides. Deux parties représentant l'ADN viral (en violet) et l'ADN cible (en bleu) sont séparées par six paires de bases.

Les trois oligonucléotides nécessaires à la formation de l'ADN sont re-suspendus dans le tampon dans lequel le test d'assemblage de l'intasome est réalisé (voir **Matériels et Méthodes §II.3.6.**) et mélangés en quantités égales. L'échantillon est chauffé à 95°C pendant 5 minutes puis lentement abaissé à température ambiante afin de favoriser l'hybridation des trois brins. L'ADN est ainsi prêt à être utilisé.

II. 3. 6. Assemblage de l'intasome de PERV-A/C avec Brd2 (667-808)

De nombreux tests ont été réalisés afin d'optimiser l'assemblage de l'intasome de PERV-A/C en complexe avec Brd2 (667-808).

Pour une première série d'essais, l'IN et Brd2 (667-808) purifiées sont dialysées (SnakeSkin™ Dialysis Tubing, *Thermo Scientific*) en présence la protéase du TEV (suivant le protocole en **Matériels et Méthodes §II.2.4.**). Le boudin de dialyse est placé dans un litre de tampon D1 (20 mM Tris HCl pH 7,5, 150 mM NaCl, 1 mM DTT, 1 mM MgCl₂) sous agitation lente pendant 16 heures à 4°C. L'ADN mimant le produit d'intégration concertée est ajouté dans le boudin de dialyse et replacé sous agitation pendant 16 heures supplémentaires. La solution dialysée commençant à former un précipité, sa concentration saline est montée à 250 mM. Après 2 heures, elle est récupérée et supplémentée avec 5 mM de CHAPS (3-[(3-cholamidopropyl)diméthylammonio]-1-propanesulfonate) avant d'être centrifugée 5 minutes à 14 000 × g, à 4°C, pour ne récupérer que le surnageant. Celui-ci est injecté sur une colonne Superdex 200 Increase 10/300 GL (*GE Healthcare*), pré-équilibrée en tampon D1 avec 250 mM NaCl, pour réaliser une chromatographie d'exclusion stérique (voir **Matériels et Méthodes §II.1.3.3.**).

Le deuxième essai réalisé commence par trois dialyses successives de l'IN avec Brd2 (667-808). Chacune d'elle a une durée de 2 heures et se fait en tampon D2 (50 mM Tris HCl pH 7,5, 0,1 mM EDTA, 0,5 M NaCl, 2 mM β -mercaptoéthanol, 1% glycérol, 1 mM $MgCl_2$) avec une concentration en urée décroissante (2 M, 1 M puis sans urée). Elles sont placées sous agitation lente et à 4°C. La préparation d'ADN est ajoutée en excès dans le boudin de dialyse avant de débiter la dernière dialyse. Deux tests sont réalisés à partir de cet échantillon :

- Dans un premier cas, l'échantillon est concentré par centrifugation dans un Amicon[®] Ultra-0,5 (*Merck Millipore*) filtrant les particules de taille inférieure à 10 kDa. La solution est ensuite injectée sur la colonne Superdex 200 Increase 10/300 GL (*GE Healthcare*), pré-équilibrée en tampon D2 pour réaliser une chromatographie d'exclusion stérique.

- Dans un deuxième cas, 5 mM CHAPS et 4 μ M $ZnSO_4$ sont ajoutés à la solution qui est ensuite placée sur roue à 4°C pendant 16 heures. L'échantillon récupéré est injecté sur la colonne Superdex 200 Increase 10/300 GL (*GE Healthcare*), pré-équilibrée en tampon D2 pour réaliser une chromatographie d'exclusion stérique.

Un troisième test débute avec la dilution de l'IN purifiée de façon à diminuer la concentration en NaCl de 1 M à 0,3 M. Dans un même temps, l'échantillon est incubé avec 10 μ L de thrombine par milligramme de protéine, de façon à cliver l'étiquette de l'IN surproduite à partir du plasmide d'expression pET-28a. La thrombine coupe spécifiquement entre l'arginine et la glycine dans la séquence d'acides aminés Leu-Val-Pro-Arg-Gly-Ser. La solution diluée est injectée sur une colonne HiTrap[®] Q HP (*Sigma*) afin de réaliser une chromatographie à échange d'ions et, plus précisément, d'anions. Pour ce faire, on utilise une résine cationique permettant la retenue spécifique de molécules chargées négativement par des interactions ioniques. L'échantillon non retenu sur la colonne est récupéré et supplémenté avec du CHAPS 5 mM. Il est ensuite concentré par centrifugation et l'ADN est ajouté en excès. La solution est placée dans un boudin de dialyse plongé dans un tampon D3 (50 mM Tris HCl pH 7,5, 0,1 mM EDTA, 100 mM NaCl, 2 mM β -mercaptoéthanol, 1% glycérol, 1 mM $MgCl_2$, 1 μ M $ZnSO_4$) et agité lentement pendant 2 heures à 4°C. L'échantillon est à nouveau concentré par centrifugation avant d'être injecté sur la colonne Superdex 200 Increase 10/300 GL (*GE Healthcare*), pré-équilibrée en tampon D3.

Lors du quatrième essai où un test de pH est réalisé, l'IN purifiée est dialysée avec l'ADN en excès, soit dans tampon D4 (50 mM Tris HCl pH 8,5, 0,1 mM EDTA, 0,5 M NaCl, 2 mM β -

mercaptoéthanol, 1% glycérol, 1 mM MgCl₂, 1 μM ZnSO₄, 5 mM CHAPS), soit dans un tampon D5 (50 mM Tris HCl pH 6,5, 0,1 mM EDTA, 0,5 M NaCl, 2 mM β-mercaptoéthanol, 1% glycérol, 1 mM MgCl₂, 1 μM ZnSO₄, 5 mM CHAPS), pendant une nuit à 4°C. L'échantillon dialysé est ensuite dilué de façon à ne diminuer que la concentration saline (de 500 mM à 250 mM NaCl) avant d'être chargé sur la colonne HiTrap[®] Heparin HP (*Sigma*). Celle-ci est composée d'héparine immobilisée sur des billes de sépharose et permet, entre autres, la fixation de protéines liant l'ADN. L'élution est réalisée avec un gradient salin allant jusqu'à 2 M NaCl.

Pour une dernière série d'essais, l'IN et Brd2 (667-808) purifiées sont dialysées dans un litre de tampon D6 (20 mM HEPES pH 7,5, 300 mM NaCl, 0,35 mM DTT, 1 mM MgCl₂, 1% glycérol) sous agitation lente pendant 16 heures à 4°C. Le substrat d'ADN est ajouté dans le boudin de dialyse avant d'être remplacé sous agitation à 4°C pendant 16 heures supplémentaires. Trois tests indépendants sont ensuite réalisés à partir de l'échantillon dialysé :

- Il est chargé sur une colonne Ni-NTA (Protino[®] Ni-NTA 5 mL, *Macherey-Nagel*), préalablement équilibrée en tampon D6. L'élution de l'échantillon est réalisée grâce à un gradient (0 – 1 M imidazole) avec le tampon D6 ayant une plus forte concentration d'imidazole.
- Une partie de l'échantillon protéique est concentrée par centrifugation (voir **Matériels et Méthodes §II.1.4.**) dans un Amicon[®] Ultra-0,5 (*Merck Millipore*) permettant de filtrer les particules de taille inférieure à 10 kDa. Il est ensuite chargé sur une colonne Superdex 200 Increase 10/300 GL (*GE Healthcare*), pré-équilibrée en tampon D6 pour réaliser une chromatographie d'exclusion stérique (voir **Matériels et Méthodes §II.1.3.3.**)
- Dans un troisième cas, la solution nucléoprotéique est chargée sur une colonne HiTrap[®] Heparin HP (*Sigma*) pré-équilibrée en tampon D6. L'élution est réalisée avec un gradient de sel en utilisant le tampon D6 contenant 2 M NaCl.

Les différentes fractions issues de ces essais sont analysées par SDS-PAGE.

III. Techniques analytiques

III. 1. Électrophorèse sur gel discontinu de polyacrylamide en conditions dénaturantes

L'électrophorèse sur gel discontinu de polyacrylamide en conditions dénaturantes, ou SDS-PAGE, permet la séparation de protéines selon leurs poids moléculaires.

Le gel est dit discontinu car il est constitué de deux parties, un gel de concentration et un gel de séparation. Ceux-ci diffèrent par leur pH et leur concentration en acrylamide/bis-acrylamide 37,5/1 (v/v) (*Euromedex*). Le taux d'acrylamide détermine la densité du gel et ainsi la gamme de séparation des protéines (**Tableau 6**). De ce fait, plus la concentration du gel de séparation est élevée, plus les protéines de petites tailles seront séparées. La mise en contact de persulfate d'ammonium (APS, pour *Ammonium PerSulfate*) (*Sigma*) et de N,N,N',N'-tétraméthyléthylènediamine (TEMED) (*Euromedex*) permet la polymérisation de l'acrylamide. Quant à lui, le gel de concentration, en partie supérieure, permet l'obtention de bandes fines, chargées en protéines, qui migreront ensuite dans le gel de séparation où elles seront séparées.

	Gel de concentration (4%)		Gel de séparation		
			10%	12%	15%
Acrylamide/bis-Acrylamide 30%	330 µL		1,65 mL	2 mL	2,5 mL
1 M Tris HCl pH 6,8	312,5 µL	3 M Tris HCl pH 8,8	625 µL	625 µL	625 µL
Eau	1,74 mL		2,5 mL	2,12 mL	1,62 mL
SDS 10%	100 µL		200 µL	200 µL	200 µL
APS 10%	25 µL		50 µL	50 µL	50 µL
TEMED	2,5 µL		5 µL	5 µL	5 µL

Tableau 6 | Composition des gels de concentration 4% et de séparation 10%, 12% et 15% pour un gel discontinu de polyacrylamide.

Les échantillons sont préalablement chauffés 5 minutes à 95°C dans du tampon Laemmli (69,45 mM Tris-HCl pH 6,8, 11,1% (v/v) glycérol, 1,1% SDS, 0,005% bleu de bromophénol, *BioRad*) (Laemmli, 1970), contenant un agent réducteur (0,0125% (v/v) β-mercaptoéthanol, *MP Biomedicals*), puis sont déposés sur le gel. Le SDS est un détergent anionique qui permet la dénaturation des protéines et les charge négativement. Un marqueur de poids moléculaires (Protein Ladder Plus, *Euromedex*) est déposé sur le gel de façon à estimer la taille des protéines d'intérêt.

Après dépôt des échantillons sur le gel, un champ électrique est appliqué et la migration se fait en tampon Tris Glycine-SDS pH 8,5 (25 mM Tris Base, 192 mM Glycine, 0,1% SDS, *Euromedex*) pendant 10 minutes à 90 V, puis à 140 V jusqu'à ce que le front de migration arrive au bas du gel.

III. 1. 1. Révélation au bleu de Coomassie

La coloration au bleu de Coomassie permet une vérification de la présence et de la pureté des protéines d'intérêt après migration complète sur gel. Elle permet également l'estimation de la quantité relative de protéine.

Une fois la migration terminée, le gel est plongé dans une solution commerciale à base de bleu de Coomassie (InstantBlue, *Expedeon*), qui a la capacité de se lier aux résidus d'acides aminés aromatiques et basiques. Le gel reste en contact avec la solution pendant une heure avant d'être rincé à l'eau distillée.

III. 1. 2. Révélation au nitrate d'argent

La coloration au nitrate d'argent permet une révélation des protéines plus sensible que par le bleu de Coomassie. De façon similaire, elle a pour but de vérifier la pureté de l'échantillon. Cette coloration se base sur la réduction des ions argent par les chaînes latérales des résidus d'acides aminés, les rendant sensibles à la lumière et permettant la révélation des protéines.

En pratique, le gel est plongé dans une solution d'éthanol 40% contenant 25 μ L de formaldéhyde 37% pendant 10 minutes. Il est ensuite rincé à l'eau distillée deux fois puis immergé dans une solution de $\text{Na}_2\text{S}_2\text{O}_3$ 0,02% (w/v) pendant 10 minutes. Le gel est à nouveau rincé deux fois à l'eau distillée avant d'être mis en contact avec le nitrate d'argent 0,1% (w/v) durant 10 minutes. La révélation se fait dans une solution de Na_2CO_3 3% (w/v) et $\text{Na}_2\text{S}_2\text{O}_3$ 0,001% (w/v) contenant 25 μ L de formaldéhyde, pendant 10 minutes. La coloration du gel est arrêtée avec l'ajout de 2,3 M d'acide citrique.

III. 2. Immunodétection après électrotransfert sur membrane de nitrocellulose

Cette méthode, aussi appelée *Western-Blot* ou immuno-empreinte, permet la détection et l'identification de protéines par l'utilisation d'anticorps spécifiques. Après séparation des protéines par SDS-PAGE, celles-ci sont transférées sur une membrane de nitrocellulose (Amersham Protran 0,2 NC, *GE Healthcare*) grâce à un courant électrique à ampérage constant de 400 mA. Ce transfert se fait dans un tampon composé de 10 mM CAPS pH 11 (pour acide 3-(CyclohexylAmino)-1-PropaneSulfonique) (*Sigma*) et d'éthanol 10% (v/v) pendant 45 minutes.

Les protéines se fixent à la membrane grâce à des interactions de types ionique et hydrophobe, en conservant les positions relatives qu'elles avaient sur le gel.

Une étape de saturation est ensuite réalisée afin de minimiser les interactions aspécifiques entre la membrane de nitrocellulose et les anticorps qui seront ajoutés ultérieurement. Pour cela, la membrane est plongée dans un tampon PBS 1X (*Euromedex*), Tween20 0,1% (v/v) (*Sigma*), lait écrémé 5% (*Régilait*) pendant une heure. Ce sont les protéines contenues dans le lait qui assurent le blocage des sites d'interactions non spécifiques.

La détection des protéines portant une étiquette poly-histidine débute par la mise en contact de la membrane avec un anticorps primaire monoclonal de souris (1/6 000^{ème} final) (*Sigma*), spécifique de cette étiquette, dans le tampon de saturation, pendant une heure. Trois lavages avec du tampon PBS 1X et Tween20 0,1% (v/v) sont nécessaires pour éliminer les anticorps non liés à la membrane. Cette dernière est ensuite incubée avec un anticorps secondaire de chèvre anti-immunoglobulines G de souris, couplé à la phosphatase alcaline (1/10 000^{ème} final, *Sigma*), dans du tampon PBS 1X, Tween20 0,1% (v/v) pendant une heure. La membrane est enfin rincée plusieurs fois avec une solution de PBS 1X, Tween20 0,1% (v/v), puis une fois avec du PBS 1X avant révélation.

Dans le cas des protéines portant une étiquette Strep-tag[®] II, ce n'est pas un anticorps primaire qui est utilisé, mais un composé Strep-Tactin[®] couplé à la phosphatase alcaline (1/6 000^{ème} final, *IBA*). De façon similaire à l'anticorps primaire anti-histidine, celui-ci est mis en contact avec la membrane de nitrocellulose dans le tampon de saturation. Trois lavages sont ensuite réalisés avec un tampon PBS 1X, Tween20 0,1% (v/v) de façon à enlever le surplus avant révélation.

Dans les deux cas, la révélation des protéines d'intérêt se fait par chimioluminescence, grâce au système de révélation SigmaFast[™] BCIP[®]/NBT (*Sigma*). Il contient un substrat BCIP (5-Bromo-4-Chloro-3-Indolyl Phosphaste) et un chromogène NBT (NitroBlue Tetrazolium). Après quelques minutes d'incubation avec la membrane, le substrat est hydrolysé par la phosphatase alcaline en un produit intermédiaire qui dimérise. Au cours de cette dimérisation, un relargage de protons provoque la réduction du chromogène NBT en un produit final insoluble violet, traduisant la présence des protéines d'intérêt. Enfin, la coloration est stoppée par rinçage de la membrane à l'eau distillée.

III. 3. Pontage covalent

Le pontage covalent chimique permet de figer de façon permanente une interaction protéine-ligand ou protéine-protéine. De telles interactions peuvent ensuite être observées après migration sur SDS-PAGE.

III. 3. 1. Pontage covalent au BS³

Le Bis(SulfoSuccinimidyl) Subérate (BS³) est un agent de pontage covalent présentant deux fonctions esters de N-HydroxySuccinimide (NHS), une à chacune de ses extrémités, espacées par un bras non réactif. Les fonctions NHS interagissent avec les amines primaires des chaînes latérales des résidues de lysine pour former une liaison amide stable.

En pratique, 6 µg de protéine d'intérêt sont incubés avec 2 mM de BS³ (*Sigma*) dans un tampon 100 mM HEPES pH 8, 600 mM NaCl pendant une heure, à température ambiante. La réaction est stoppée par l'ajout de 4 mM Tris HCl pH 7,5 et 0,5 mM glycine et incubation pendant 30 minutes à température ambiante. Les échantillons sont ensuite traités par addition de tampon Laemmli (1X final) et chauffés 5 minutes à 95°C avant d'être analysés par SDS-PAGE.

III. 3. 2. Pontage covalent au GTA

Le glutaraldéhyde (GTA) est une petite molécule portant deux fonctions aldéhydes pouvant réagir avec les fonctions amines des protéines. Cette molécule, plus petite que le BS³, est moins spécifique.

En pratique, 5 mg de protéine et 4 µL de tampon 20 mM Tris HCl pH 8 sont mélangés à 1 µL de GTA 0,1% (w/v) (*Pierce*) et incubés pendant 20 minutes à température ambiante. La réaction est ensuite arrêtée avec l'addition d'1 µL de glycine 200 mM. Après l'ajout de tampon Laemmli (1X final), les échantillons sont chauffés 5 minutes à 95°C et analysés par SDS-PAGE.

III. 4. Diffusion dynamique de la lumière (DLS)

La diffusion dynamique de la lumière (DLS, pour *Dynamic Light Scattering*) est une technique non invasive permettant d'estimer le rayon hydrodynamique (Rh) de particules en solution et également de caractériser la dispersité de l'échantillon analysé. Celle-ci se base sur le mouvement brownien des particules, c'est-à-dire le mouvement dû à leur collision avec les molécules du solvant dans lequel elles se trouvent (Brown & Bennett, 1866). En effet, plus les

particules sont de petite taille, plus leur mouvement est rapide ; à l'inverse des particules de grosses tailles seront plus lentes. Un faisceau laser est appliqué sur l'échantillon et les molécules en suspension le composant rediffusent la lumière, créant des fluctuations d'intensité du signal qui peuvent être mesurées. Ces fluctuations dépendent du coefficient de diffusion transversal D_T des particules. L'équation de Stokes-Einstein relie ce coefficient au rayon hydrodynamique R_H des particules :

$$R_H = \frac{K_B \cdot T}{6 \cdot \pi \cdot \eta \cdot D_T}$$

Avec T , la température ; η , la viscosité du milieu et K_B , la constante de Boltzmann.

La concentration de l'échantillon, la nature de la surface de la particule et le type d'ions dans la solution ont une influence sur le coefficient de diffusion. Le rayon hydrodynamique ainsi mesuré peut être relié à un poids moléculaire théorique. En effet, des protéines globulaires standards, dont les rayons hydrodynamiques sont connus, sont utilisées pour estimer celui de la protéine d'intérêt.

L'indice de polydispersité (PdI, pour *Polydispersity Index*) est calculé automatiquement pour chaque échantillon et permet d'obtenir des informations sur la dispersité de l'échantillon. Un très faible PdI indique une solution monodisperse alors qu'un PdI au-dessus de 0,7 est signe d'une solution polydisperse.

Les expériences de DLS sont réalisées avec un appareil Zetasizer Nano S (*Malvern Instruments*) à une température de 8°C. Pour ce faire, 100 μ L d'échantillon protéique sont filtrés sur filtres 0,02 μ m et placés dans une cuve stérile en plastique (*Brand*). Pour chaque expérience, un triplicata de dix mesures est réalisé.

IV. Mesure d'interaction

IV. 1. Titration calorimétrique isotherme (ITC)

La titration calorimétrique isotherme (ITC, pour *Isothermal Titration Calorimetry*) est une méthode physique permettant la détermination quantitative des paramètres thermodynamiques d'une interaction moléculaire (Velazquez-Campoy *et al.*, 2004). Notamment utilisée pour l'étude de l'interaction protéine/ligand, elle est de plus en plus utilisée pour les interactions protéine/protéine.

Cette méthode repose sur le principe que l'interaction de deux molécules entraîne des changements thermodynamiques qui peuvent être mesurés. Il est alors possible de calculer la constante d'association (K_a) de l'interaction. Dans le cas de cette thèse, les protéines d'intérêt étant purifiées sous forme de complexe, c'est la constante de dissociation (K_D) de celui-ci qui a pu être calculée. En effet, lors de la dilution de l'échantillon dans le tampon, une libération de chaleur peut également être enregistrée. Cette libération de chaleur, représentée par l'énergie de Gibbs d'association (ΔG_a) ou l'énergie de Gibbs de dissociation (ΔG_D), peut être reliée au K_a et K_D par les formules suivantes :

$$\Delta G_a = -RT \cdot \ln K_a$$

$$\Delta G_D = -RT \cdot \ln K_D$$

Avec R, la constante universelle des gaz parfaits ($1,9872 \text{ cal.mol}^{-1}.\text{K}^{-1}$) et T, la température en kelvin.

L'appareil est composé d'une chambre de réaction remplie de tampon dans lequel est plongée la seringue contenant le complexe protéique (**Figure 40**). À chaque intervalle de temps très court, la seringue injecte un petit volume dans la chambre de réaction, déclenchant la dissociation du complexe protéique. Le dégagement de chaleur résultant est alors mesuré par la cellule.

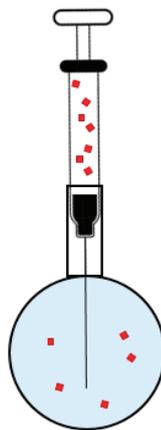


Figure 40 | Schéma de la cellule de réaction d'ITC (d'après (Velazquez-Campoy *et al.*, 2004)). La seringue contient la solution protéique (rouge) et l'injecte dans la chambre de réaction contenant le tampon (bleu).

En pratique, les mesures sont réalisées avec l'appareil MicroCal PEAQ-ITC (*Malvern*) à 25°C. Vingt injections de 2 µL d'échantillon sont réalisées toutes les 210 secondes et le milieu est mélangé grâce à une agitation de la chambre de réaction à 750 rpm.

IV. 2. Essais de type pull down

Le *pull down* est une technique de co-précipitation permettant d'isoler une ou plusieurs protéines grâce à l'immobilisation d'une protéine ou molécule appât. Cette dernière est liée à un substrat en phase solide et est utilisée pour l'extraction des protéines interagissant avec elle.

En pratique, l'IN de PERV-A/C portant son étiquette poly-histidine et Brd2 (667-808) dont l'étiquette a été clivée sont mélangées de façon à avoir un rapport molaire 1:1. Le milieu de réaction est porté à un volume final de 30 µL afin que celui soit composé de 20 mM HEPES pH 7,5, 300 mM NaCl, 0,35 mM DTT, 1 mM MgCl₂ et 10% glycérol. Afin de contrôler l'échantillon initial, 5 µL du mélange sont conservés. Afin de tester les composés potentiellement compétiteurs, 5 µL de ceux-ci sont ajoutés au mélange afin de porter leur concentration finale à 1 µM, à l'exception du témoin dans lequel 5 µL de tampon de réaction est ajouté. Chaque échantillon est incubé à 4°C pendant 16 heures.

Après incubation, 5 µL de billes magnétiques avec une affinité pour l'étiquette poly-histidine (HIS-Select® Nickel Magnetic Agarose Beads, *Millipore*) sont ajoutées au milieu, de même que 150 µL de tampon de lavage (20 mM HEPES pH 7,5, 300 mM NaCl, 0,35 mM DTT, 1 mM MgCl₂, 10% glycérol et 20 mM imidazole). Les échantillons sont à nouveau incubés pendant 2 heures à 4°C avant d'être rincés deux fois avec le tampon de lavage. Un support magnétique permet lors de chaque étape de séparer les billes de la fraction soluble. Enfin, l'éluion est réalisée par ajout de 5 µL de tampon Laemmli et les différentes fractions sont analysées par SDS-PAGE.

V. Étude structurale

V. 1. Diffusion des rayons X aux petits angles

La diffusion des rayons X aux petits angles (« *Small Angle X-ray Scattering* » ou SAXS) est une technique permettant l'analyse structurale des macromolécules à partir d'une solution monodisperse.

L'échantillon est soumis à un faisceau monochromatique de rayons X, qui interagit de manière élastique avec les couches électroniques des atomes. L'intensité des ondes réémises est enregistrée jusqu'à un angle 2θ maximum et permet de sonder la structure de l'échantillon (**Figure 41**). La résolution de l'information structurale s'améliore avec l'augmentation de 2θ , tandis que la valeur de l'intensité réémise à un angle nul est inconnue. En effet, la majorité des photons du faisceau incident traverse l'échantillon sans interaction et le détecteur est protégé de ce dernier par une cible de plomb (*beamstop*), opaque aux rayons X.

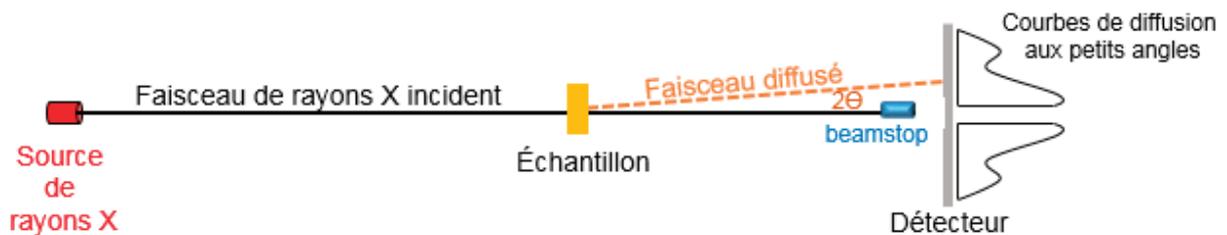


Figure 41 | Schématisation de la diffusion des rayons X aux petits angles.

Cette méthode permet généralement d'obtenir en premier la valeur du rayon de giration R_g de la macromolécule d'intérêt qui est défini par l'expression :

$$R_g^2 = \frac{\sum m_i \cdot r_i^2}{\sum m_i}$$

Avec m_i , la masse du $i^{\text{ème}}$ atome dans la macromolécule et r_i , la distance au centre de masse de la $i^{\text{ème}}$ macromolécule.

En effet, la courbe de diffusion $I(q)$, mesurée aux très petits angles avec $q=4\pi \cdot \sin\theta/\lambda$, respecte l'équation de Guinier :

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

La représentation de Guinier (**Figure 42A**) est ainsi une droite dont la pente est proportionnelle à R_g et dont l'extrapolation à $q=0$ permet d'obtenir $I(0)$. Il est alors possible de déterminer la valeur du poids moléculaire de la particule en solution (et son état d'oligomérisation) selon l'équation :

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

Avec V_{sol} , le volume de la solution protéique ; c , la concentration protéique ; M_w , le poids moléculaire ; N , le nombre d'Avogadro ; m_p , le nombre d'électron dans la particule ; ρ_s , la densité du solvant et V_p , le volume spécifique partiel de la particule.

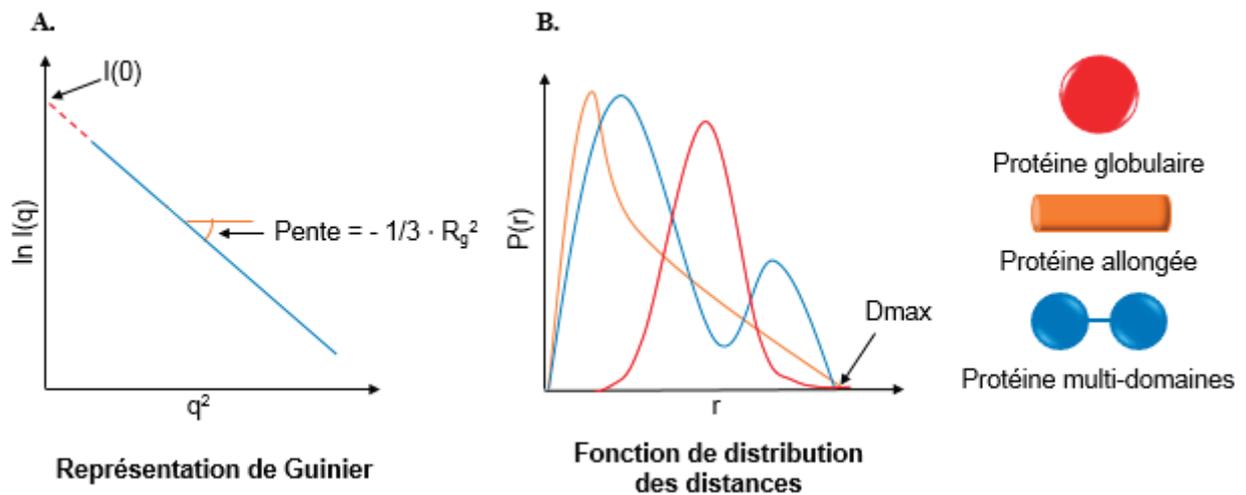


Figure 42 | **A.** Représentation de Guinier, $\ln I(q) = f(q^2)$ et **B.** fonctions de distribution des distances $P(r)$ associées aux modèles *ab initio* correspondants (adaptée de (Svergun & Koch, 2003)).

Une analyse des données SAXS selon une représentation de Kratky, avec $q^2 I(q)$ en fonction de q , permet de caractériser la flexibilité de la macromolécule en solution. Un tracé en forme de cloche est caractéristique d'une protéine globulaire, tandis qu'un tracé avec une pente ascendante aux grandes valeurs de q est caractéristique d'une protéine non-structurée. Les tracés intermédiaires caractérisent des protéines partiellement flexibles (**Figure 43**).

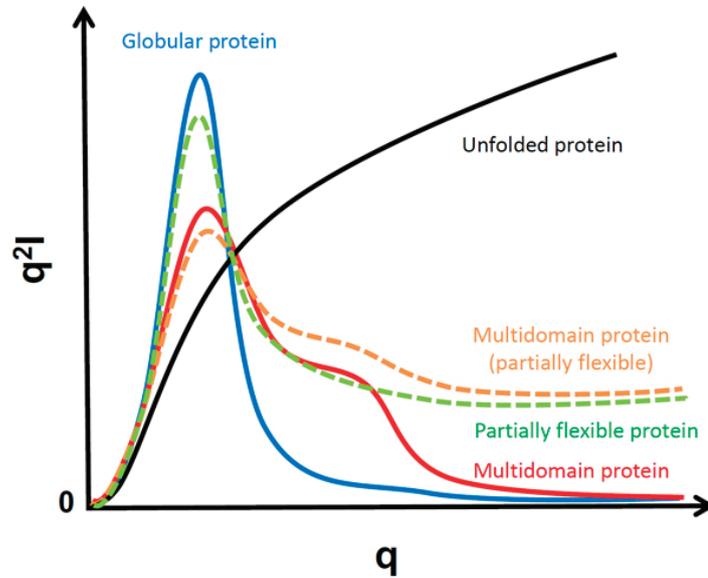


Figure 43 | Représentation de tracés de Kratky caractéristiques en fonction de la forme de la macromolécule biologique étudiée et de son degré de flexibilité.

De plus, un calcul de la transformée de Fourier inverse des intensités mesurées $I(q)$ jusqu'à l'angle 2θ maximal permet de remonter à la fonction de distribution des distances $P(r)$ (**Figure 42B**). Le D_{\max} correspond à la valeur interatomique de distance maximale de la particule. La valeur du R_g peut également être calculée à partir de la $P(r)$ selon l'équation.

$$R_g^2 = \int_0^{D_{\max}} r^2 P(r) dr / \int_0^{D_{\max}} P(r) dr$$

Il est enfin possible de calculer un modèle *ab initio* de l'enveloppe SAXS de la protéine d'intérêt, sous la forme d'un assemblage compact de billes représentant chacune le volume moyen occupé par un résidu (Putnam *et al.*, 2007). La résolution de cette enveloppe SAXS est généralement comprise entre 4 et 2 nm. La courbe de diffusion calculée par transformée de Fourier à partir de ce modèle 3D doit être en accord avec la courbe de diffusion expérimentale $I(q)$. La différence entre les deux courbes est estimée par un test statistique de χ^2 .

$$\chi^2 = \frac{1}{m-1} \sum_{k=1}^m \left(\frac{I_{fit}(q) - I_{exp}(q)}{SE_{I_{exp}(q)}} \right)^2$$

Où m représente le nombre d'intensités expérimentales $I(q)$ prises en compte pour le test et SE l'erreur sur ces mesures.

En pratique, nos analyses SAXS ont été réalisées sur la ligne BioSAXS BM29 du synchrotron ESRF de Grenoble, avec un faisceau de rayon X incident de longueur d'onde $\lambda = 0,991 \text{ \AA}$. En raison de la nécessité d'analyser un échantillon monodisperse, l'enregistrement se fait à la suite directe d'une chromatographie d'exclusion stérique (voir **Matériels et méthodes II.1.3.3.**) afin de séparer les différents états oligomériques de la protéine (méthode de SEC-SAXS). Pour cela, une colonne Superdex S200 5/150 GL (*GE Healthcare*) est utilisée.

Les mesures sont réalisées à une température de 4°C , dans un tampon 20 mM Tris HCl pH 8 et 0,5 M NaCl. Des échantillons dont la concentration varie de 6 à 15 mg/mL sont injectés et analysés.

Les données mesurées sont analysées avec les programmes CHROMIXS et SAS-DATA de la suite ATSAS (Franke *et al.*, 2017). Ces programmes permettent de déterminer les paramètres décrits précédemment, tels que le R_g , la $P(r)$ et le D_{\max} , qui sont utilisés comme contraintes pour le calcul de l'enveloppe SAXS. Cette dernière est réalisée à l'aide du programme DAMMIF à partir de la superposition de cinquante modèles calculés. Enfin, le logiciel Chimera (Pettersen *et al.*, 2004) est utilisé pour visualiser l'enveloppe SAXS.

V. 2. Cristallographie aux rayons X

L'échantillon biologique est cette fois-ci sous forme cristalline. Tout comme en SAXS, cet échantillon est soumis à un faisceau monochromatique de rayons X. Les faisceaux diffractés sont cependant plus intenses qu'en SAXS, en raison de la présence d'interférences constructives dues à la périodicité du réseau cristallin. Les intensités peuvent ainsi être mesurées jusqu'à un angle 2θ plus grand permettant d'obtenir une meilleure résolution de la structure finale, idéalement de l'ordre de l'angström.

V. 2. 1. Cristallogenèse

La cristallogenèse permet le passage d'une molécule d'un état soluble à un état solide ordonné, aussi appelé cristal. Un cristal est caractérisé par la répétition tri périodique (dans les trois dimensions de l'espace) d'une maille élémentaire contenant lesdites molécules.

La croissance cristalline nécessite la présence de sels, de tampon et d'un agent précipitant dont les natures et concentrations influent sur la solubilité protéique (**Figure 44**). Une augmentation progressive de la concentration en protéine conduit à une solution saturée en

protéine, nécessaire à la formation de germes cristallins. C'est l'étape de nucléation. L'étape de croissance cristalline entraîne alors la diminution de la concentration protéique jusqu'à atteindre un équilibre, correspondant à la zone métastable. Lors de cette étape, les germes cristallins augmentent en taille.

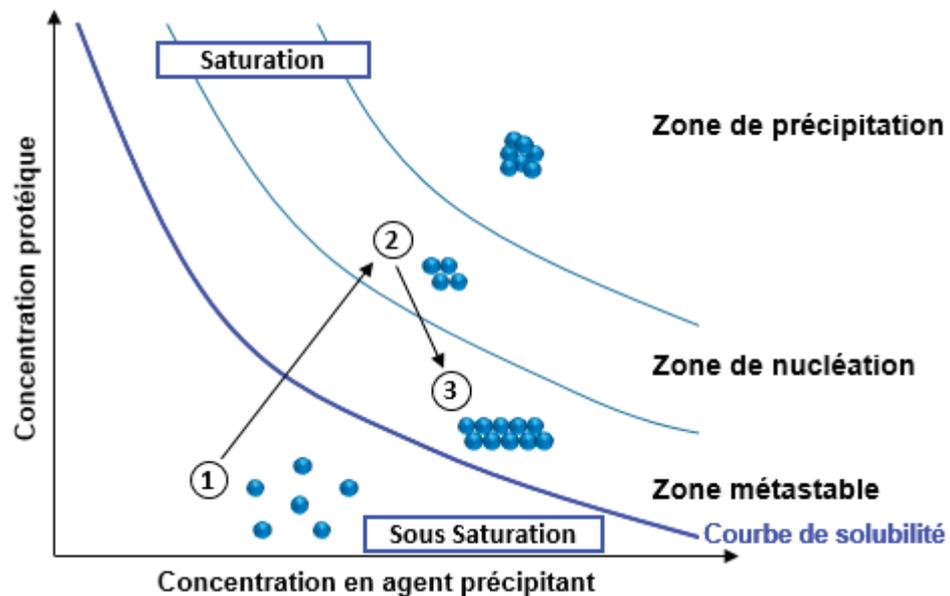


Figure 44 | Diagramme de solubilité en fonction des concentrations en agent précipitant et en protéine. La courbe de solubilité sépare la zone de sous saturation (1) des trois zones de sursaturation (zone de nucléation (2), zone métastable (3) et zone de précipitation) (d'après (Kupitz *et al.*, 2014)).

Les conditions permettant la cristallisation sont propres à chaque protéine, ainsi la première étape de la cristallogénèse consiste à tester un maximum de paramètres afin de déterminer la combinaison optimale pour la protéine d'intérêt. Ainsi, des variations de la concentration et de la nature de l'agent précipitant, de la concentration en protéines, du pH, de la température de croissance ou encore de la présence d'additifs sont réalisées. Parmi les agents précipitants, on retrouve trois classes chimiques : les polymères organiques (et plus particulièrement les polyéthylènes glycol, PEG), les sels et les solvants organiques.

La méthode la plus utilisée en cristallogénèse afin de porter la solution protéique de la zone de sous saturation à la zone de nucléation est la méthode de la diffusion de vapeur. Elle permet d'augmenter de façon douce et progressive la concentration en protéine. Elle utilise un milieu hermétiquement clos dans lequel se trouve un réservoir contenant l'agent précipitant et une goutte composée d'un mélange du même agent précipitant et de la protéine d'intérêt. La différence de concentration en agent précipitant dans la goutte et dans le réservoir provoque un

transfert d'eau de la goutte (moins concentrée en agent précipitant) vers le réservoir (plus concentré en agent précipitant). Cette diffusion, qui est due à la tension de vapeur entre les deux milieux, s'arrête lorsqu'un état d'équilibre est atteint entre les deux parties. Au cours de ce processus lent, le volume de la goutte est diminué, et les concentrations en agent cristallisant et en protéine augmentent progressivement. Ce phénomène conduit la protéine à un état de sursaturation, pouvant aboutir à la formation de cristaux.

Deux procédés peuvent être utilisés pour appliquer la méthode de la diffusion de vapeur : les techniques de la goutte suspendue et de la goutte assise (**Figure 45A**).

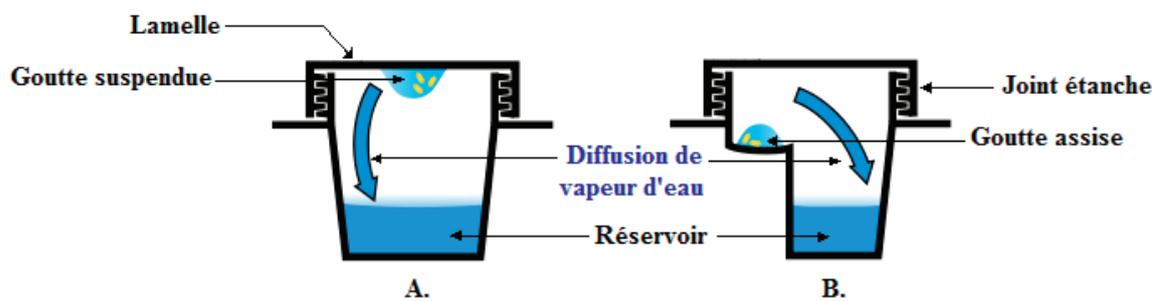


Figure 45 | Schéma représentant les techniques de la goutte suspendue (A) et de la goutte assise (B).

La technique de la goutte assise se réalise dans des plaques 96 puits et permet un criblage des conditions de cristallisation. À l'inverse, la technique de la goutte suspendue, réalisée dans des plaques de 24 puits, est utilisée pour l'optimisation des conditions de cristallisation.

V. 2. 2. Criblage des conditions de cristallisation

Le criblage des conditions de cristallisation est réalisé dans des plaques de 96 puits (MRC 2 Well Crystallization Plate, *Hampton Research*). Celles-ci contiennent des réservoirs de 90 μL de solution de cristallisation, chacun associés à deux plateformes sur lesquelles sont déposées des gouttes de 400 nL. Ces gouttes sont formées d'un mélange de 200 nL de solution protéique à des concentrations variables et de 200 nL d'agent précipitant identique à celui contenu dans le réservoir.

La réalisation de ces nanogouttes est automatisée à l'aide d'un robot de nano-pipetage *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 PSF de Lyon.

Afin de déterminer les paramètres conduisant à la formation de cristaux de protéine, de nombreuses conditions de cristallisation sont testées. Pour cela, différents kits commerciaux de criblage sont utilisés. Ils permettent de faire varier la valeur du pH (de 4 à 9,5) mais aussi de tester plusieurs agents précipitants (sels et PEG). La concentration de la protéine d'intérêt ainsi que son tampon de stockage peuvent également être variés.

Lorsque des cristaux apparaissent dans certaines nanogouttes, les conditions de cristallisation correspondantes sont reproduites dans des plaques de 24 puits (*Greiner bio-one*). Des essais en gouttes suspendues sont réalisés manuellement sur des lamelles siliconées (**Figure 45B**). Les microgouttes de 2 μL , déposées sur ces lamelles, contiennent un mélange de solution protéique et d'agent précipitant à différents ratios (1:1, 1:2 et 2:1). Chacun des 24 puits est constitué d'un réservoir contenant 500 μL de solution de cristallisation. Les plaques sont ensuite placées à 19°C.

Dans le cas où des cristaux de taille suffisante se forment dans une goutte, ceux-ci sont pêchés avec une boucle en nylon et cryo-protégés par trempage dans une solution d'éthylène glycol (20% (v/v)). La cryo-protection tend à éviter la formation de microcristaux de glace autour du cristal lors de sa congélation dans l'azote liquide. Cette dernière a pour but de protéger le cristal des dégâts radiatifs dus aux rayons X et limite ainsi sa dégradation.

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

V. 2. 3. Diffraction aux rayons X

Le cristal est illuminé par un faisceau, généralement monochromatique, de rayons X pour déterminer la structure des molécules qui le composent à une échelle atomique. Le faisceau incident interagit avec chaque atome composant le cristal qui réémettent des ondes sphériques. Du fait de la périodicité du cristal, les rayons X subissent des interférences constructives et destructives qui déterminent des directions privilégiées de réémission des ondes dans l'espace.

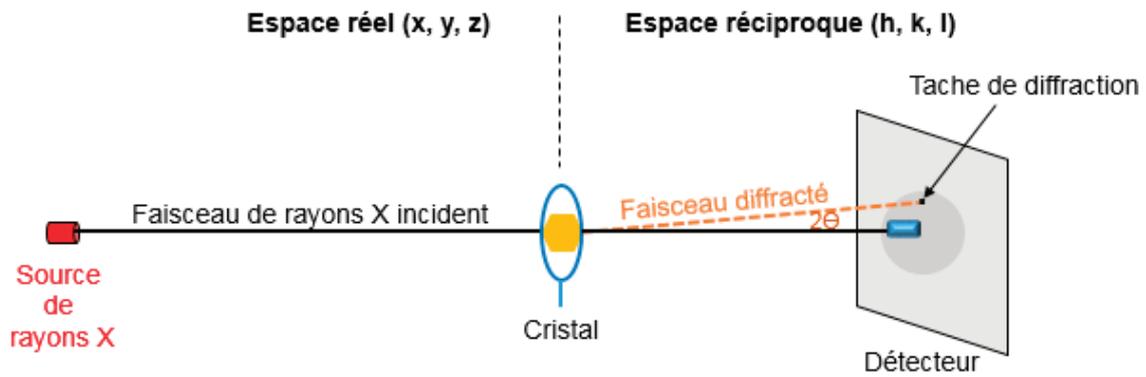


Figure 46 | Schéma représentant la diffractions aux rayons X d'un cristal. Le faisceau de rayons X entre en contact avec le cristal, ce qui le réémet dans des directions privilégiées de l'espace réciproque. Le détecteur permet d'enregistrer les taches de diffraction résultantes. Le *beamstop*, en bleu, préserve le détecteur de la puissance du faisceau incident à l'angle 2θ nul.

Ce phénomène de diffraction permet la réémission des ondes dans des directions précises suivant des indices (h, k, l) de l'espace réciproque, qui est défini par des vecteurs x^*, y^*, z^* dépendant des vecteurs a, b, c de la maille du cristal. Ces ondes sont caractérisées par une phase φ et une amplitude $|F|$ qui est calculée à partir des intensités $I(hkl)$ mesurées par le détecteur (**Figure 46**). Expérimentalement, seules ces intensités peuvent être mesurées et l'équation suivante permet de calculer leurs amplitudes :

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

Cependant, les phases ne peuvent pas être mesurées expérimentalement. Il est donc nécessaire d'utiliser des méthodes de phasage, telles que le remplacement moléculaire, afin d'estimer les phases des ondes diffractées.

L'onde diffractée est caractérisée par un facteur de structure $F(hkl)$ qui englobe son amplitude et sa phase et qui est égal à la somme des facteurs de diffusion atomique des atomes de coordonnées x_j, y_j et z_j contenus dans la maille cristalline. Elle prend en compte les décalages de phase des ondes diffusées par les différents atomes :

$$F(hkl) = |F(hkl)| \cdot e^{i\varphi_{hkl}} \text{ ou } F(hkl) = \sum_{j=1}^{j=n} f_j \cdot e^{2\pi i(hx_j + ky_j + lz_j)}$$

Avec φ_{hkl} , la phase de la réflexion hkl ; f_j , le facteur de diffusion de l'atome j et n , le nombre d'atome dans la maille.

Le facteur de structure $F(hkl)$ permet ensuite de calculer la fonction de densité électronique $\rho(x,y,z)$ en tout point (x,y,z) du cristal. En effet, la densité électronique correspond à la transformée de Fourier inverse du facteur de structure $F(hkl)$:

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

Avec V , le volume du cristal de protéine.

Ainsi, connaissant la fonction de densité électronique, il est possible de représenter des cartes tridimensionnelles de densité dans lesquelles seront placés les atomes de la protéine d'intérêt. Par reconstruction manuelle ou automatisée, il sera possible, au final, de remonter à la structure 3D de la protéine cristallisée.

V. 3. Criblage virtuel de chimiothèque

Le criblage virtuel *in silico* (ou *molecular docking*) est une technique bioinformatique de simulation utilisée pour la conception rationnelle de ligands ou de molécules bioactives et, par extension, de médicaments. Elle consiste à cribler des bases de données informatiques de petites molécules (ou chimiothèques) dans le but d'identifier lesquelles seraient le plus à même d'interagir avec une cible biologique, typiquement un récepteur protéique ou une enzyme.

Cette technique nécessite la connaissance de la structure tridimensionnelle de la cible protéique d'intérêt. Les logiciels de *docking* vont ainsi évaluer, pour chaque composé de la chimiothèque et grâce à une fonction de score, sa propension à interagir avec la cible de structure 3D connue. Chaque composé est modélisé informatiquement et ses propriétés physico-chimiques sont prises en compte (nature des atomes, des liaisons chimiques, flexibilité intramoléculaire, charges globales et partielles, etc.). Il est également possible de donner une certaine flexibilité aux chaînes latérales de la protéine cible, en se restreignant cependant à une zone de petite taille, sans quoi les temps de calcul augmentent de manière colossale. La fonction de score calculée pour chaque ligand permet ensuite d'interclasser les composés et, ainsi, de déterminer les meilleures molécules candidates. Enfin, les logiciels de *docking* permettent d'estimer une différence d'énergie libre de liaison (ΔG en kcal/mol) pour chaque ligand, valeur de laquelle on peut déduire un K_D théorique.

Cette étude a été conduite au sein de l'équipe par le Dr. Robert qui a bénéficié d'un accès au Centre de Calcul de l'IN2P3 (ccIN2P3, USR6402) de Villeurbanne, permettant l'utilisation de 2 500 processeurs en parallèle. Nous avons réalisé le criblage complet de la branche '*Clean drug-like*' de la chimiothèque virtuelle ZINC (<https://zinc.docking.org>) (Sterling & Irwin, 2015) qui comprend plus de 15 millions de composés. L'intérêt de cette branche est que chaque composé obéit à la Règle de Lipinski (Lipinski *et al.*, 2001; Lipinski, 2004). Cette dernière implique qu'un médicament en prise orale satisfait, dans l'immense majorité des cas, quatre propriétés physico-chimiques : 1/ pas plus de 5 groupements donneurs de liaison hydrogène ; 2/ pas plus de 10 groupements accepteurs ; 3/ un poids moléculaire compris entre 150 et 500 Da ; et 4/ un coefficient de partage octanol-eau (log P) inférieur à 5. La branche '*Clean drug-like*' ajoute une dernière condition, celle que le composé doit avoir un nombre de degré de liberté inférieur ou égal à 7.

L'ensemble des calculs ont été réalisés en utilisant le logiciel Autodock VINA (Trott & Olson, 2010) qui, aujourd'hui, fait référence dans le domaine. La puissance de calcul dont nous disposons au ccIN2P3 nous a permis de réaliser cette étude en environ un mois, alors que ce même protocole aurait pris plusieurs décennies sur un ordinateur standard.

V. 4. Microscopie électronique en transmission

La microscopie électronique en transmission (TEM, pour *Transmission Electron Microscopy*) est une technique de microscopie utilisant un faisceau d'électrons traversant une couche très mince d'échantillon afin d'observer son image agrandie grâce à l'utilisation de lentilles magnétiques et d'un détecteur (**Figure 47**).

Des canons à électron à émission de champ (FEG, pour *Field Electron Gun*) ont récemment été développés pour les microscopes électroniques. Ils permettent d'obtenir un faisceau brillant et cohérent après application d'une différence de potentiel électrique pour les extraire d'une pointe de tungstène. Le faisceau accéléré passe alors dans la colonne du microscope, où règne un vide poussé, et est focalisé par un premier système de lentilles sur l'échantillon placé sur une grille. L'image électronique projetée sur le détecteur provient de la déviation élastique des électrons incidents par les noyaux des atomes de l'échantillon. Le contraste de l'image est faible pour les échantillons biologiques. Ainsi, des techniques d'amélioration du contraste sont utilisées. Elles peuvent être invasives (ajout de colorants pour augmenter le contraste d'intensité) ou non-invasives (utilisation des lentilles magnétiques pour augmenter le contraste de phase).

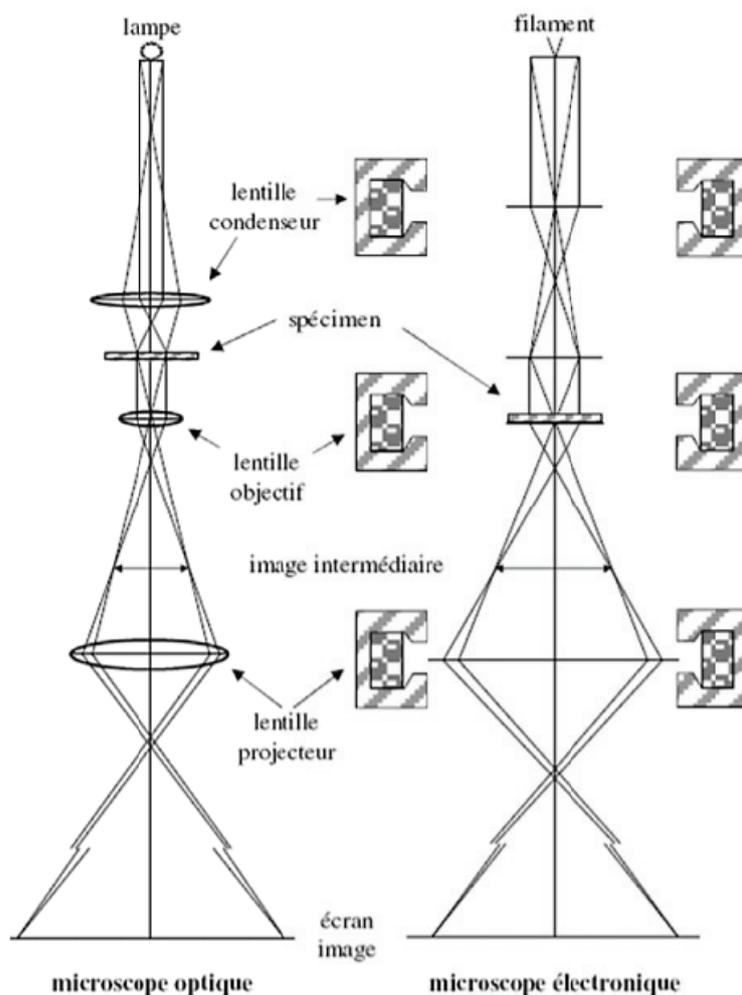


Figure 47 | Schéma comparant les principes des microscopies optique et électronique.

Coloration négative. Les molécules biologiques sont principalement composées d'atomes légers, comme le carbone, l'oxygène ou encore l'hydrogène. De ce fait, elles possèdent un faible pouvoir diffuseur. La coloration négative permet d'augmenter le contraste de l'échantillon par rapport au solvant qui l'entoure, en le recouvrant de molécules de métaux lourds comme des sels d'uranyl possédant un fort pouvoir diffuseur.

Typiquement, une goutte d'échantillon (30 μL), concentré à environ 0,2 mg/ml, est adsorbée sur une grille en carbone mince (10 nm d'épaisseur). Un sel d'atome lourd est alors ajouté avant buvardage pour en éliminer l'excès. La grille est ensuite placée dans un microscope électronique à transmission et une série d'images est enregistrée.

Nos échantillons ont été analysés avec le microscope électronique FEI Tecnai12 120 kV de la plateforme de l'Institut de Biologie Structurale (IBS) de Grenoble. Le sel d'atome lourd utilisé

est du sodium silicotungstate. Les images obtenues permettent l'observation de structures à une résolution maximale d'environ 2 nm.

Cryo-microscopie électronique. Cette technique ne nécessite pas de colorants et réduit les dommages d'irradiation causés par le faisceau d'électrons incident. Elle permet aujourd'hui d'obtenir des structures 3D de macromolécules biologiques à l'échelle quasi-atomique, lorsqu'elle est opérée sur des microscopes FEG de 200-300 kV de très grande stabilité possédant des caméras à détection directe d'électron (Cheng, 2018).

L'échantillon est placé sur un film de carbone troué de manière régulière, lui-même positionné sur une grille métallique. Après buvardage, l'échantillon est congelé rapidement dans de l'éthane liquide à l'aide d'un cryo-plongeur pour être vitrifié. Il est alors mis sur un porte-échantillon qui est inséré dans le microscope électronique. Le porte-échantillon est maintenu à très basse température grâce à une circulation d'azote liquide. L'échantillon est ensuite traversé par le faisceau d'électrons et des images (micrographes) sont enregistrées. Ces images permettent d'observer les projections des particules qui se sont réparties sur le film de support selon différentes orientations. Les échantillons placés dans les trous du film sont plus contrastés, ce contraste pouvant encore être augmenté par un effet de défocalisation du aux lentilles magnétiques.

Les micrographies enregistrées sont finalement traitées pour remonter à la structure 3D de l'échantillon. Pour se faire, un échantillonnage des projections 2D est effectué ; les projections présentant la même orientation sont moyennées afin d'améliorer le rapport signal sur bruit, puis ces classes 2D présentant l'échantillon sous différents angles sont mathématiquement combinées pour calculer l'enveloppe 3D de la particule.

Nos échantillons ont été analysés sur la plateforme de biologie structurale de Milan, Italie, à l'aide d'un cryo-microscope électronique FEI FEG 200 kV Talos Arctica. Les échantillons ont été déposés sur une grille revêtue de carbone troué (QUANTIFOIL R 2/1), épongés avec du papier filtre, congelés dans de l'éthane liquide à l'aide d'un cryo-plongeur FEI Vitrobot avec des temps variant de 1 à 3 secondes puis conservés dans de l'azote liquide.

RÉSULTATS ET DISCUSSION

I. Étude de l'interaction entre l'IN de PERV-A/C et Brd2

Cette première partie porte sur l'étude moléculaire et structurale de la liaison de l'IN de PERV-A/C à la protéine humaine Brd2.

La protéine Brd2, aussi appelée RING3 (pour *Really Interesting New Gene 3*), appartient à la famille BET (voir l'**Introduction §II.3.2**). Elle possède deux domaines BD dans sa région amino-terminale (**Figure 33**) lui permettant de lier les queues d'histones acétylées. À la différence de Brd4, Brd2 interagit principalement avec la lysine acétylée en position 12 de la queue d'histone H4 (Kanno *et al.*, 2004). La structure cristalline du domaine BD1 de Brd2, en complexe avec la queue d'histone H4 acétylée, montre que la majorité des contacts intermoléculaires implique les quinze résidus N-terminaux de la queue d'histone. Le domaine BD1, présent sous forme dimérique, est ainsi capable de lier deux queues d'histones acétylées (Umehara *et al.*, 2010).

Brd2 et Brd4 sont associées à la chromatine durant tout le cycle cellulaire, à l'inverse d'autres protéines à bromodomaine qui s'en dissocient au cours de la mitose. C'est cette caractéristique de Brd2 qui est utilisée par l'herpèsvirus KSHV pour amarrer son génome au chromosome lors de la mitose et se propager lors de la division cellulaire de l'hôte (Viejo-Borbolla *et al.*, 2005).

Brd2 est un co-activateur et corépresseur de la transcription, régulant ce processus en fonction de l'environnement dans lequel elle se trouve. Entre autre, elle joue un rôle dans le recrutement de facteurs de transcription sur la chromatine, permettant l'activation de la transcription du gène codant la cycline A (Denis *et al.*, 2006). Chez la souris, une perturbation de la régulation de la transcription par Brd2 entraîne une apparition d'obésité sévère (Wang *et al.*, 2009).

De plus, comme expliqué dans le chapitre **Sujet Doctoral**, j'ai été impliquée durant ma thèse dans un projet collaboratif visant à montrer que Brd2 humaine était un cofacteur cellulaire de l'IN de PERV-A/C et à caractériser le complexe d'interaction. Nos collaborateurs, les Dr. Blanchard et Ronfort, avaient montré, par des approches combinées de double-hybride en levure et de mutagenèse dirigée, que le domaine ET de la région carboxy-terminale de Brd2 interagissait avec la séquence EBM du CTD de l'IN PERV-A/C qui est conservée chez les *Gammarrétrovirus* (**Figure 48**). Ce résultat d'interaction était en accord avec ceux obtenus par

comprenant le CTD de l'IN PERV-A/C et le fragment carboxy-terminal ET-SEED (667-808) de Brd2 (**Figure 33**) en vue de son étude biochimique et structurale.

Les limites 667-808 de Brd2 ont été choisies sur la base d'une prédiction du désordre intrinsèque de cette protéine réalisée à l'aide du serveur GlobPlot (Linding *et al.*, 2003). Ce dernier prédisait que l'extrémité carboxy-terminale 809-836 de la protéine était fortement désordonnée (**Figure 49**), tandis que les résultats en double-hybride obtenus par nos collaborateurs montraient que cette zone n'était pas nécessaire pour l'interaction avec l'IN de PERV-A/C.

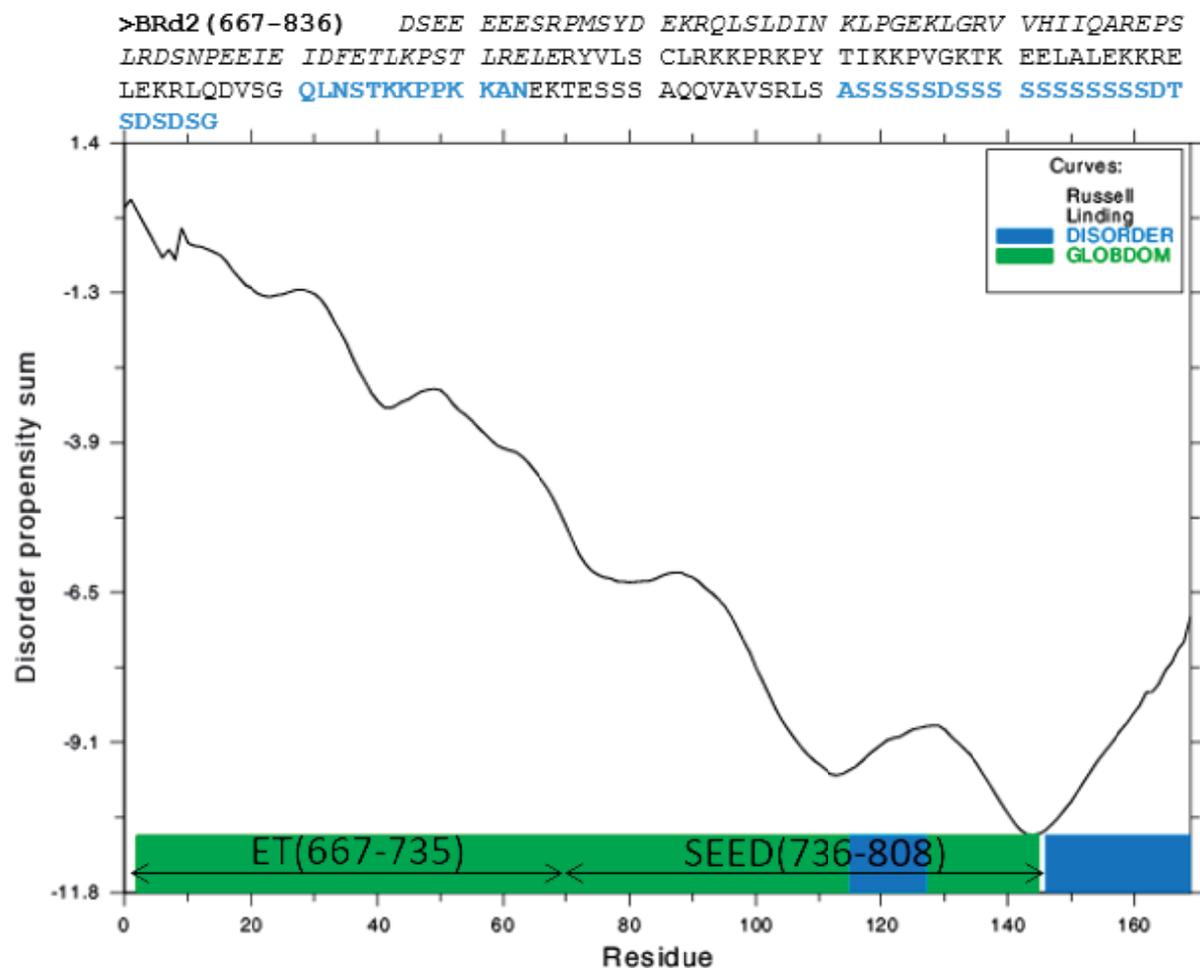


Figure 49 | Prédiction par le serveur Globprot (Linding *et al.*, 2003) du désordre intrinsèque de la région carboxy-terminale 667-836 de la protéine humaine Brd2. Les résidus de la région ET (667-735) sont en italique. Les segments prédits comme désordonnés sont en bleus.

Les zones de désordre intrinsèque sont un obstacle à la production de cristaux de protéine et il a été décidé de se restreindre à la région 667-808 de Brd2 afin de faciliter l'étude structurale.

Cette étude pouvait également permettre de déterminer la première structure atomique du domaine SEED des protéines BET, aucune séquence homologue de structure connue n'étant détectée dans la PDB avec le logiciel de modélisation SWISS-MODEL (Waterhouse *et al.*, 2018).

Le double fragment CTD IN PERV A/C:Brd2 (667-808) a été exprimé à partir d'une construction originale appelée MultiTEV, qui avait été mise au point dans l'équipe par le Dr. Yajjou afin d'obtenir un complexe stable (voir **Matériels et Méthodes §I.1.1.**). J'ai optimisé le protocole de surproduction et de purification du complexe sous la direction du Dr. Yajjou, comme décrit ci-dessous, afin de le caractériser.

I. 1. Le CTD de l'IN PERV-A/C et Brd2 (667-808) – Stratégie MultiTEV

I. 1. 1. Surproduction et purification du complexe MultiTEV

Dans un premier temps, le plasmide pET-28a portant la construction MultiTEV a été utilisé pour la transformation des bactéries *E. coli* T7 Express. Un clone est repiqué et mis en culture dans du milieu LB en présence de kanamycine, puis l'expression a été induite par ajout d'1 mM d'IPTG et le milieu a été incubé pendant trois heures à 37°C.

Les bactéries sont lysées par sonication dans le tampon de lyse et deux chromatographies sont utilisées pour purifier le complexe protéique. En effet, les protéines d'intérêt portent toutes deux une étiquette de purification qui leurs sont propres (le CTD de l'IN portant une étiquette Strep-tag[®] et Brd2 (667-808) portant une étiquette poly-histidine), permettant de purifier le complexe MultiTEV en deux étapes. Dans un premier temps, Brd2 (667-808) a été purifiée par chromatographie d'affinité pour le nickel afin de récupérer la protéine seule et en complexe avec le CTD de l'IN. L'échantillon résultant a ensuite été purifié par chromatographie d'affinité pour la streptavidine permettant de retenir uniquement le CTD de l'IN de PERV-A/C.

I. 1. 1. 1. Purification par chromatographie d'affinité pour le nickel

Le lysat bactérien a tout d'abord été injecté sur colonne Ni-NTA. L'injection de l'échantillon et son élution ont été suivies par mesure de l'absorbance à 280 nm en sortie de

colonne, de même que la conductance du tampon, indiquant les changements de concentration saline (**Figure 50A**).

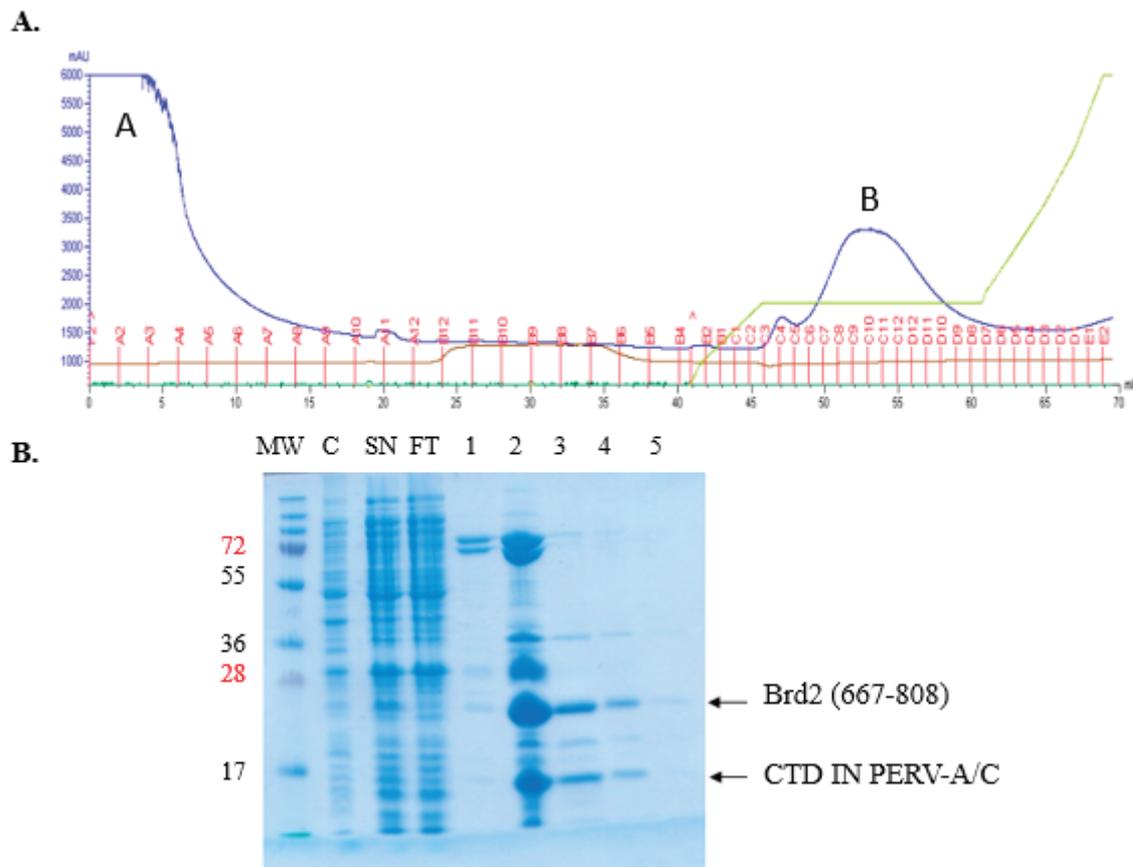


Figure 50 | Purification par chromatographie d'affinité pour le nickel du complexe MultiTEV. **A.** Chromatogramme de purification. En bleu, l'absorbance à 280 nm ; en marron, la conductance et en vert, le gradient d'imidazole. Le pic A correspond à l'échantillon non retenu et le pic B correspond aux fractions d'élution. **B.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; piste C, fraction non soluble ; piste SN, fraction soluble ; piste FT, fraction non retenue et pistes 1 à 5, fractions d'élution.

Le chromatogramme de purification montre un premier pic A correspondant à la sortie des protéines non retenues par la colonne, alors que le pic B représente l'élution des protéines lors du gradient d'imidazole. Les protéines sont éluées à 26% de la concentration finale, soit 250 mM d'imidazole. Le saut de conductance entre ces deux pics correspond, quant à lui, à l'augmentation de la concentration en NaCl (de 500 mM à 1 M) lors du passage du tampon de lavage. Cependant, celui-ci n'a décroché aucune protéine.

Les différentes fractions ont ensuite été analysées par SDS-PAGE (**Figure 50B**). La fraction d'élution correspondant à l'épaulement présent sur le pic B, et déposée en piste 1, ne contient

aucune des deux protéines d'intérêt et seulement des contaminants bactériens. Sur les pistes suivantes (pistes 2 à 5), correspondant au pic d'élution B, on peut observer deux bandes migrant entre les marqueurs de poids moléculaires de 17 kDa et 28 kDa. Une analyse par spectrométrie de masse a confirmé qu'il s'agissait respectivement du CTD de l'IN de PERV-A/C (12 kDa) et de Brd2 (667-808) (18 kDa). En revanche, ces deux protéines ne sont pas pures dans l'échantillon élué et plusieurs bandes correspondant à des contaminants bactériens sont également visibles. Néanmoins, la purification se faisant en plusieurs étapes, il sera possible de s'affranchir de ces contaminants par la suite.

On peut également remarquer qu'une bande correspondant à Brd2 (667-808) est visible dans la fraction soluble (piste SN) et n'est plus présente pour la fraction non retenue (piste FT), cela indiquant que l'intégralité de cette protéine a été retenue par la colonne. Au contraire, une bande correspondant au CTD de l'IN est visible pour les pistes correspondant aux fractions solubles et non retenues (pistes SN et FT, respectivement). Ainsi, il est possible qu'une partie du CTD de l'IN n'ait pas été retenue sur la colonne en raison de son étiquette non spécifique et que la fraction de CTD éluée avec Brd2 (667-808) le fut grâce à son interaction avec cette dernière.

1. 1. 1. 2. Purification par chromatographie d'affinité pour la streptavidine

Les fractions d'élution contenant ces protéines ont ensuite été regroupées et injectées sur une colonne Strep-Tactin[®] pour réaliser une chromatographie d'affinité pour la streptavidine. À nouveau, l'injection de l'échantillon et son élution ont été suivies par mesure de l'absorbance à 280 nm en sortie de colonne (**Figure 51A**). La conductance des tampons a également été mesurée lors de leurs passages dans la colonne.

On peut tout d'abord observer une augmentation de la conductance lors de l'injection de l'échantillon (pic A, correspondant à la fraction non retenue). Celle-ci est due à la différence de concentration saline entre le tampon dans lequel se trouve l'échantillon (500 mM NaCl), et le tampon W avec lequel a été pré-équilibrée la colonne (150 mM NaCl). De plus, une augmentation de l'absorbance à 280 nm au niveau du pic A correspond à l'élution des protéines non retenues sur la colonne.

Le pic B, quant à lui, correspond aux fractions d'élution, décrochées de la colonne grâce au tampon E contenant de la desthiobiotine.

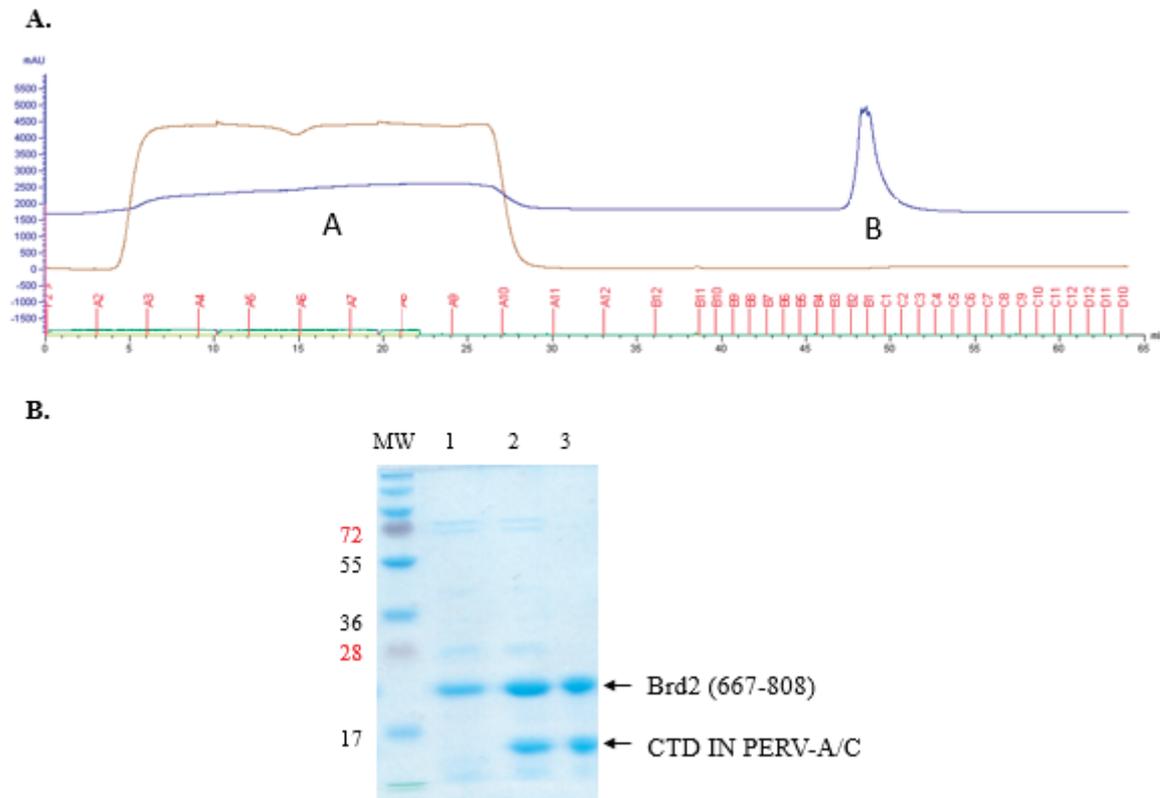


Figure 51 | Purification par chromatographie d'affinité pour la streptavidine du complexe MultiTEV. **A.** Chromatogramme de purification. En bleu, l'absorbance à 280 nm et en marron, la conductance. Le pic A correspond à l'échantillon non retenu et le pic B aux fractions d'élution. **B.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; piste 1, fraction non retenue et pistes 2 et 3, fractions d'élution.

L'analyse par SDS-PAGE (**Figure 51B**) montre que dans la fraction non retenue (piste 1), seule une bande correspondant à Brd2 (667-808) est visible. Cela signifie que l'intégralité du CTD de l'IN a été retenue par la colonne et qu'il est possible que seule la fraction de Brd2 non complexée à l'IN sorte au moment de l'injection. Les deux fractions d'élution (pistes 2 et 3) contiennent deux bandes qui correspondent à Brd2 (667-808) et au CTD de l'IN de PERV-A/C. Cependant, quelques contaminants sont encore présents dans la fraction d'élution et une étape de purification supplémentaire, par chromatographie d'exclusion de taille, sera donc nécessaire.

Une analyse par immunodétection après électrotransfert sur membrane de nitrocellulose a été réalisée avant cette dernière étape afin de confirmer la nature des deux protéines purifiées (**Figure 52**).

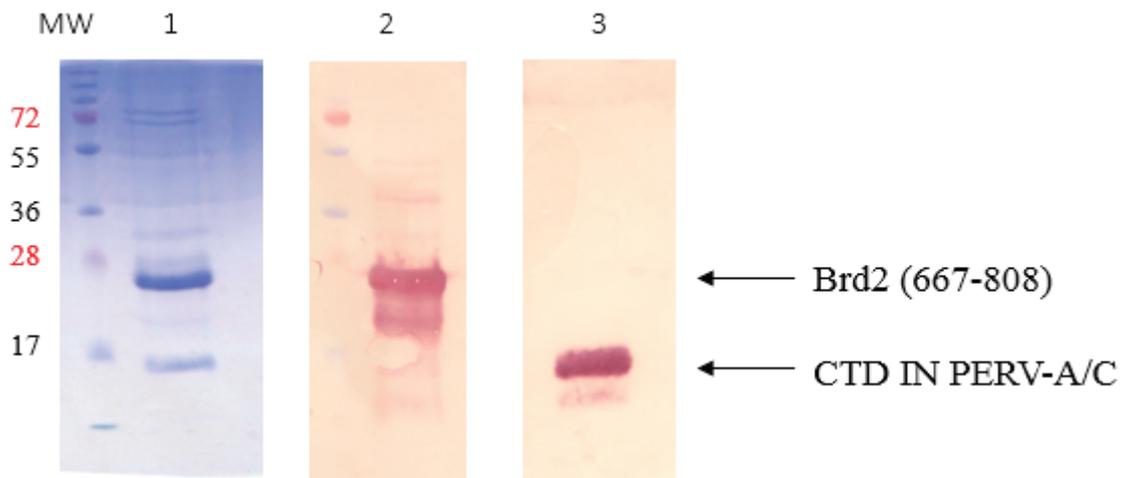


Figure 52 | Analyse par immunodétection après électrotransfert sur membrane de nitrocellulose du complexe MultiTEV purifié. **1.** Analyse SDS-PAGE de la fraction d'éluion après chromatographie d'affinité pour la streptavidine, colorée au bleu de Coomassie. **2.** Révélation de l'étiquette poly-histidine. **3.** Révélation de l'étiquette Strep-tag[®]. Avec MW, les marqueurs de poids moléculaires.

La piste 1 représente la fraction d'éluion après la purification par chromatographie d'affinité pour la streptavidine et l'échantillon qui a été analysé par immunodétection après électrotransfert sur membrane de nitrocellulose. On voit bien deux bandes représentant Brd2 (667-808), migrant sous le marqueur de poids moléculaire de 28 kDa, et le CTD de l'IN de PERV-A/C, migrant sous le marqueur de 17 kDa.

Après révélation de l'étiquette poly-histidine, une bande est identifiée sous le marqueur de poids moléculaire de 28 kDa (piste 2). Cette bande correspond ainsi à Brd2 (667-808), porteur d'une étiquette poly-histidine. La révélation de l'étiquette Strep-tag[®] provoque l'apparition d'une bande sous le marqueur de poids moléculaire de 17 kDa (piste 3). Celle-ci représente le CTD de l'IN de PERV-A/C, porteur d'une étiquette Strep-tag[®].

I. 1. 1. 3. Chromatographie d'exclusion de taille

L'échantillon, élué après la chromatographie d'affinité pour la streptavidine, a ensuite été injecté sur une colonne Superdex 200 Increase 10/300 GL de façon à séparer les différents composants en fonction de leur poids moléculaire. Le tampon dans lequel se trouve l'échantillon a été modifié au moment de l'élution dans le but d'éliminer la desthiobiotine présente dans le tampon E et pouvant potentiellement affecter la croissance cristalline. Ce changement de tampon est réalisé avec le tampon BicAm dédié à la caractérisation biochimique du complexe ainsi qu'à une partie de l'étude structurale, ou avec le tampon final pour l'étude structurale.

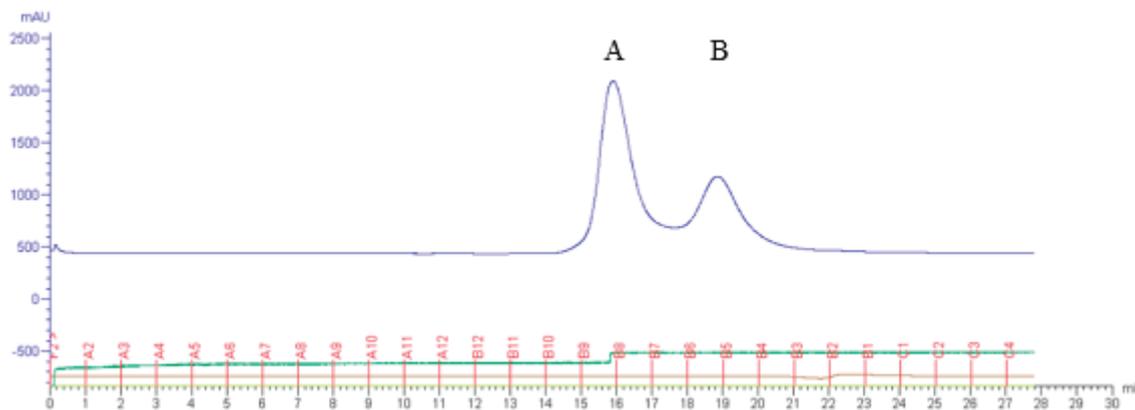


Figure 53 | Chromatographie d'exclusion de taille du complexe MultiTEV. En bleu, le suivi de l'absorbance à 280 nm.

Le suivi de l'absorbance à 280 nm permet d'identifier deux pics d'élution distincts (**Figure 53**). Le pic A est élué à un volume de 16 mL, correspondant approximativement à un poids moléculaire de 29 kDa, alors que le pic B est élué à 19 mL, soit environ 13 kDa. Cela peut s'expliquer par le fait que le complexe, sous forme hétérodimérique composé d'un monomère de Brd2 et d'un monomère du CTD de l'IN, soit élué dans le pic A. En revanche, il semblerait qu'une fraction de celui-ci ait été dissociée et que les deux protéines isolées soient éluées dans le pic B, la colonne n'étant pas assez résolutive pour la séparation des protéines de 18 kDa et 12 kDa.

Les fractions correspondant au pic A ont donc été analysées par pontage covalent.

I. 1. 1. 4. Pontage covalent

Une expérience de pontage covalent avec l'agent de pontage BS³ a été réalisée afin de vérifier l'état oligomérique de l'échantillon A. Les résultats ont été analysés par SDS-PAGE (Figure 54).

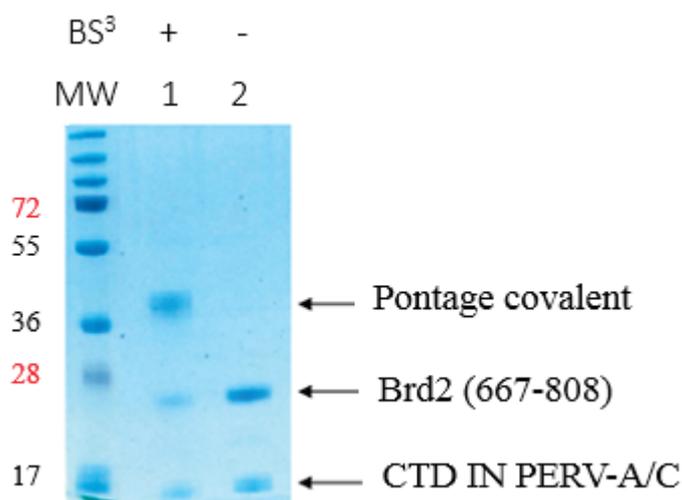


Figure 54 | Analyse par SDS-PAGE de l'état oligomérique du complexe MultiTEV après pontage covalent au BS³. Piste 1 : échantillon incubé avec l'agent de pontage. Piste 2 : échantillon sans agent de pontage. Avec MW, les marqueurs de poids moléculaires.

En absence d'agent pontant (piste 2), deux bandes correspondant au CTD de l'IN de PERV-A/C et à Brd2 (667-808) sont bien observées.

En présence de BS³ (piste 1), une bande de poids moléculaire plus haut est visible à environ 36 kDa. Cette bande est cohérente avec la taille attendue pour le complexe d'intérêt, qui est de 30 kDa pour l'hétérodimère composé d'une molécule de Brd2 (667-808) et d'une molécule de CTD de l'IN de PERV-A/C. Après l'extraction de cette bande, les deux protéines d'intérêt ont été identifiées en spectrométrie de masse, confirmant leur interaction.

L'échantillon hétérodimérique, purifié après chromatographie d'exclusion de taille et correspondant au complexe formé du CTD de l'IN de PERV-A/C et de Brd2 (667-808), a ensuite été concentré par centrifugation. Son homogénéité a été vérifiée par analyse DLS, à une concentration de 9 mg/mL, où un faible indice de polydispersité de 0,1 a été mesuré.

Ainsi, la stratégie MultiTEV a permis de surproduire et purifier de façon efficace le complexe hétérodimérique formé de Brd2 (667-808) et du CTD de l'IN de PERV-A/C. Le rendement de purification est de 2 à 4 mg de protéines par litre de culture.

I. 1. 2. Étude in vitro de la dissociation du complexe MultiTEV

I. 1. 2. 1. Titration calorimétrique isotherme

Des expériences de titration par calorimétrie isotherme (ITC) ont été réalisées afin de mesurer la constante de dissociation (K_D) du complexe purifié. Une mesure de K_D par ITC est généralement faite en injectant une série de petits aliquots de ligands dans une solution contenant l'interactant protéique et en mesurant la variation de chaleur due à la formation du complexe. Ce protocole ne peut cependant pas être utilisé dans le cas du MultiTEV, puisque nous partons d'un complexe déjà formé. Il nous est seulement possible de le dissocier par dilution.

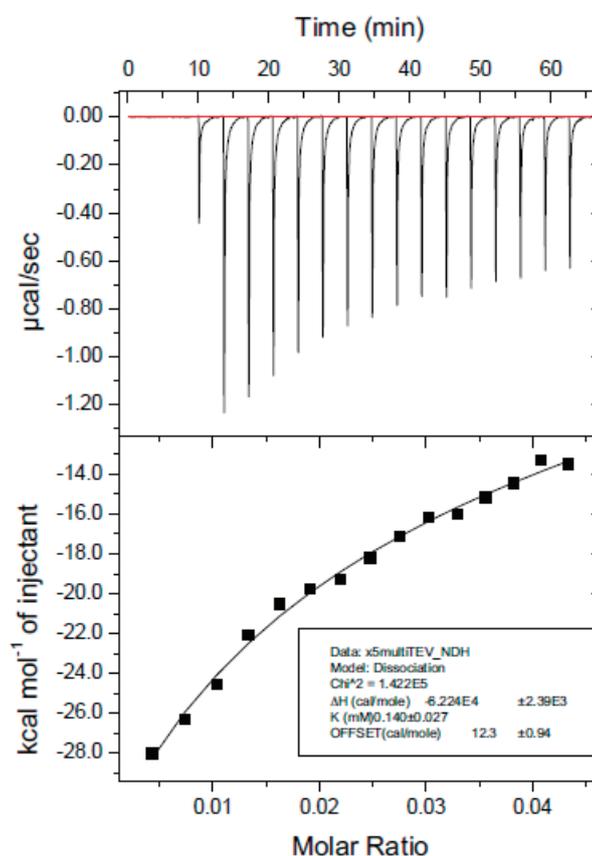


Figure 55 | Titration calorimétrique de la dissociation du CTD de l'IN de PERV-A/C et de Brd2 (667-808). En haut, différence calorimétrique mesurée à chaque injection d'échantillon dans la cellule. En bas, données normalisées par mole de protéines injectées.

En effet, la dissociation de deux protéines entraîne également une variation de chaleur donnant des informations sur la stabilité du complexe et permettant de calculer un coefficient de dissociation K_D (**Figure 55**). Ce protocole de mesure particulier a déjà été utilisé et validé par d'autres groupes (Luke *et al.*, 2005).

Il a ainsi été possible de mesurer par ITC une valeur de K_D de 140 μM pour le complexe formé du CTD de l'IN de PERV-A/C et de Brd2 (667-808), ordre de grandeur déjà observé dans d'autres cas d'interactions protéines-protéines.

1. 1. 2. 2. Mutagenèse dirigée

Nos collaborateurs avaient effectué une analyse systématique du segment EBM (378-394) de l'IN de PERV A/C, montrant que des mutations en alanine abolissaient l'interaction avec Brd2 en double-hybride de levure. Nous avons voulu tester si ces mutations étaient également disruptives pour le complexe MultiTEV et, pour ce faire, deux types de mutants alanine ont été construits : 1/ le simple mutant W379A qui affecte un résidu de tryptophane strictement conservé chez les IN gammarétrovirales (**Figure 48**) ; ce tryptophane est essentiel au profil d'intégration de MLV (El Ashkar *et al.*, 2014). 2/ Le double mutant R391A/L392A, qui affecte des résidus localisés au niveau d'un brin β intermoléculaire dans la structure RMN du complexe EBM MLV IN:ET Brd4 (Crowe *et al.*, 2016).

Ces deux mutants se trouvent sur la partie CTD de la construction MultiTEV et j'ai mis en place un système de double purification identique à celui précédemment décrit pour leur analyse (voir **Matériels et Méthodes §I.2.**).

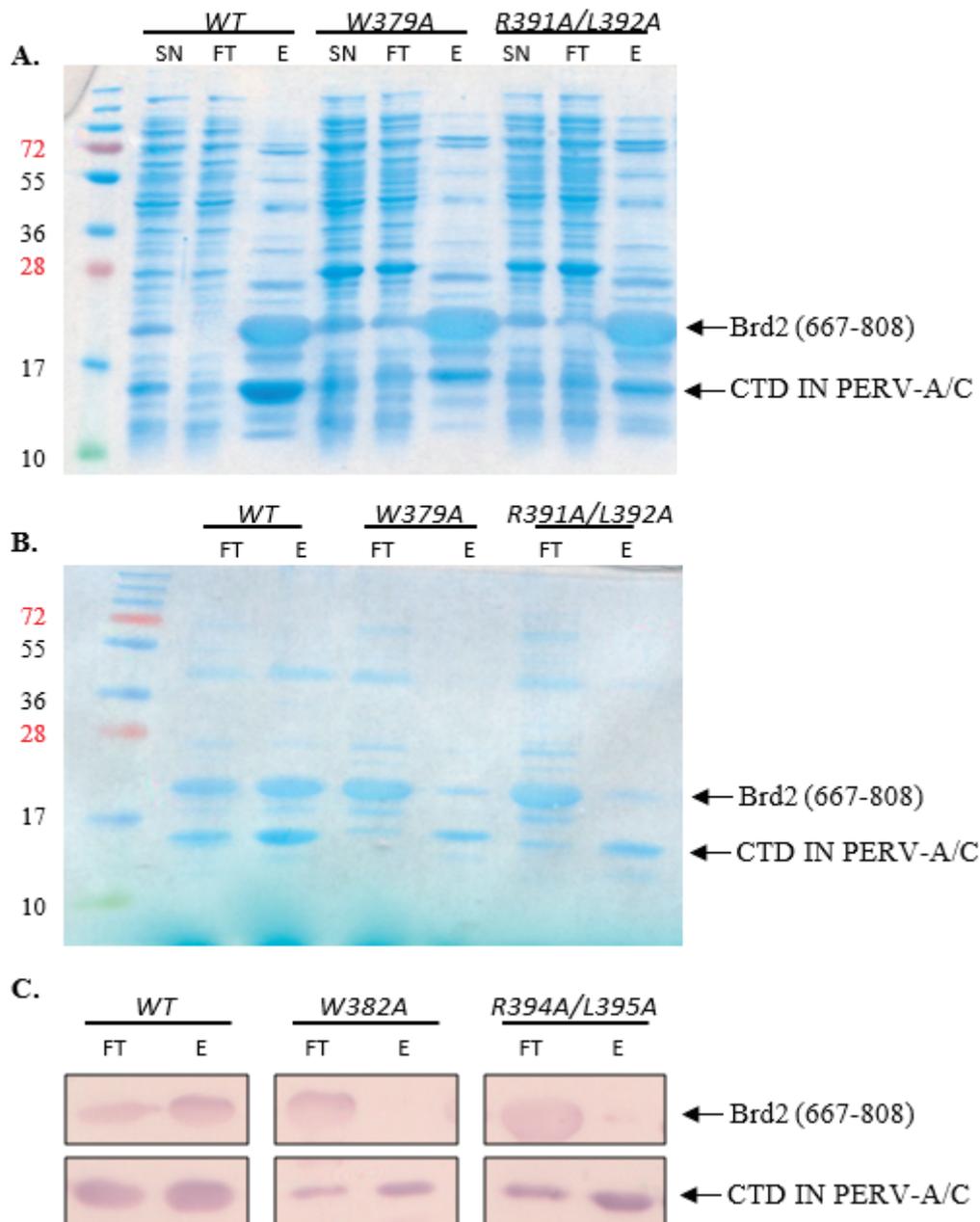


Figure 56 | Analyse par SDS-PAGE des fractions de purification après chromatographie d'affinité pour le nickel (**A**) et pour la streptavidine (**B**) du complexe MultiTEV et de ses deux mutants. Les pistes WT correspondent au complexe MultiTEV non muté formé du CTD de l'IN de PERV-A/C et de Brd2 (667-808) ; les pistes W379A et R391A/L392A aux mutants de l'IN de PERV-A/C. Pistes SN, fractions solubles ; pistes FT, fractions non retenues et pistes E, fractions d'éluion. (**C**) Détection par immunodétection après électrotransfert sur membrane de nitrocellulose de Brd2 (667-808) (détection de l'étiquette poly-histidine) et du CTD de l'IN de PERV-A/C (détection de l'étiquette Strep-tag[®]).

On peut tout d'abord noter que le complexe contrôle non muté est bien surexprimé. Après deux étapes de purification, il est retrouvé dans la fraction d'élution finale (**Figure 56B, piste WT – E**).

En revanche, les deux mutants (W379A et R391A/L392A) ont un profil de purification distinct du profil natif. En effet, après la première étape de purification par chromatographie d'affinité pour le nickel, une bande correspondant à Brd2 (667-808), avec une intensité identique à celle obtenue pour le complexe non muté, est visible dans les fractions d'élution de ces deux mutants (**Figure 56A, pistes W379A et R391A/L392A – E**). Cependant, dans les deux cas, la bande correspondant au CTD de l'IN de PERV-A/C muté est présente avec une intensité plus faible que celle observée pour le profil natif. Cela indique qu'une quantité moindre d'IN a été récupérée en complexe avec Brd2.

Les fractions d'élution obtenues après chromatographie d'affinité pour le nickel sont ensuite à nouveau purifiées par chromatographie d'affinité pour la streptavidine. Une fois encore, les mutants W379A et R391A/L392A présentent un profil d'élution comparable. En effet, la bande majoritaire retrouvée dans la fraction non retenue correspond à Brd2 (667-808) qui, avec son étiquette poly-histidine et sans son interaction avec l'IN, est éluée pendant l'injection (**Figure 56B, pistes W379A et R391A/L392A – FT**). Similairement, la bande majoritaire observée dans les fractions d'élution correspond à l'IN mutée, retenue sur la colonne grâce à son étiquette Strep-tag[®] (**Figure 56B, pistes W379A et R391A/L392A – E**). De plus, seule une quantité négligeable de Brd2 (667-808) semble avoir été éluée en même temps que l'IN.

La présence des deux protéines d'intérêt a été validée dans chacune de ces fractions grâce à une immunodétection des étiquettes poly-histidine et Strep-tag[®] (**Figure 56C**).

Ces résultats confirment l'importance de ces résidus particuliers de tryptophane, d'arginine et de leucine dans l'interaction de l'IN de PERV-A/C avec Brd2.

I. 1. 3. Étude structurale aux rayons X du complexe MultiTEV

I. 1. 3. 1. Étude par cristallographie aux rayons X

Afin de résoudre la structure cristalline du complexe, nous avons mené un large criblage des conditions de cristallisation avec, au total, plus de 3 000 conditions testées.

Les différents kits utilisés sont les suivants : MbClass Suite (*Qiagen*), Pact Ier (*Qiagen*), Protein Complex Suite (*Qiagen*), PEGs Suite (*Qiagen*), PEG/Ion HT™ (*Hampton Research*), PEGs II Suite (*Qiagen*), Crystal Screen (*Hampton Research*), AmSO₄ Suite (*Qiagen*), JCSG+ Suite (*Qiagen*), Morpheus (*Molecular Dimensions*), MPD Suite (*Qiagen*) et 5K (Plateforme *Protein Science Facility*).

Ces essais ont été réalisés à partir d'un échantillon protéique se trouvant dans deux tampons de stockage différents : le tampon final (20 mM Tris HCl pH 8, 0,5 M NaCl) et le tampon BicAm (200 mM bicarbonate d'ammonium pH 8, 100 mM NaCl), et à de nombreuses concentrations différentes. L'ensemble des conditions testées sont résumées en **Tableau 7**.

Tampons de stockage		Kits												
		MbClass Suite	Pact Ier	Protein Complex Suite	PEGs Suite	PEG/Ion HT™	PEGs II Suite	Crystal Screen	AmSO ₄ Suite	JCSG+ Suite	Morpheus	MPD Suite	5K	
Final	4,6													
	5													
	7,5													
	8,2													
	8,6													
	9,6													
	15													
	16													
	26													
	BicAm	4,2												
		8,7												
		10												
		18												

Tableau 7 | Tableau récapitulatif des concentrations protéiques et kits de criblage des conditions de cristallisation testés pour le complexe entre Brd2 (667-808) et le CTD de l'IN de PERV-A/C. En rouge, les kits de criblage testés n'ayant pas mené à l'apparition de cristaux et, en jaune, ceux ayant conduit à l'apparition de cristaux.

Dans l'une des conditions du kit 5K, de nombreux cristaux sont apparus après quelques jours. Il s'agit de la condition 17, composée de 0,1 M HEPES pH 7,5, 0,2 M acétate de calcium

et 20% PEG 5 000 MME. Ces cristaux de petites tailles se sont formés à la suite des essais de cristallisation avec le complexe protéique conservé dans le tampon final et aux concentrations de 7,5 et 8,2 mg/mL.

Cette condition de cristallisation a ensuite été reproduites en plaques de 24 puits, dans un but d'optimisation et afin d'obtenir des cristaux de taille adéquate aux expériences de diffraction des rayons X. Pour cela, plusieurs rapports entre protéine et agent précipitant ont été testés dans la goutte (1:1, 1:2 et 2:1). De plus, plusieurs concentrations de protéine, allant de 7 à 20 mg/mL, ont été utilisées. Il a rapidement été observé que la taille des cristaux augmentait avec la concentration, suggérant l'obtention de cristaux de protéine.

Ces cristaux ont été pêchés et cryo-protégés avant d'être analysés à l'ESRF de Grenoble. Malheureusement, ces cristaux ont diffracté avec un profil correspondant à celui d'un sel.

Au final, aucun des essais initiés au cours de cette étude n'a mené à l'obtention de cristaux de protéines.

1. 1. 3. 2. Étude par SAXS

Des enregistrements de type SEC-SAXS, où l'échantillon protéique est soumis au faisceau de rayons X en sortie de chromatographie d'exclusion de taille afin d'améliorer sa monodispersité, ont été effectués par le Dr. Yajjou sur la ligne de BioSAXS BM29 de l'ESRF de Grenoble. Ces enregistrements ont permis de confirmer que la masse du complexe MultiTEV correspondait bien à un hétérodimère et ont respectivement donné des valeurs de R_g et de D_{max} de 3,1 nm et 11 nm (Yajjou-Hamalian, 2017). Un modèle *ab initio* de forme allongée a été généré, mais en l'absence de modèle atomique pour la partie SEED du complexe MultiTEV, la résolution de l'enveloppe SAXS calculée était insuffisante pour identifier les parties correspondant au CTD de l'IN de PERV-A/C et à Brd2 (667-808) (résolution estimée de 3 nm).

Je me suis donc attachée à mettre au point le protocole d'expression et de purification d'un nouveau complexe de taille réduite appelé MultiTEV Minimal, qui comprend le CTD de l'IN PERV-A/C et le domaine ET de Brd2, soit Brd2 (667-735). Ainsi, même si nous renoncions à la possibilité de résoudre la structure atomique du domaine SEED, nous espérons avoir une meilleure chance d'obtenir de nouvelles informations structurales sur la formation des complexes IN:BET.

I. 2. Le CTD de l'IN PERV-A/C et Brd2 (667-735) – Stratégie MultiTEV Minimal

Les modifications apportées au plasmide ne concernent pas le gène codant le CTD de l'IN de PERV-A/C, mais la partie Brd2 qui a été réduite au seul domaine ET (résidus 667-735 de Brd2).

I. 2. 1. Surproduction et purification du complexe MultiTEV Minimal

La stratégie de surproduction et de purification mise en place pour le complexe MultiTEV Minimal est identique à celle décrite précédemment pour le complexe MultiTEV.

De la même façon, des bactéries *E. coli* T7 Express ont été transformées avec le plasmide pET-28a portant la nouvelle construction. Un clone a été mis en culture dans un milieu LB supplémenté en kanamycine et l'induction de l'expression s'est faite par ajout d'1 mM d'IPTG suivie d'une incubation pendant trois heures à 37°C.

Après lyse mécanique des bactéries dans le tampon de lyse, le même protocole de purification que celui utilisé pour le complexe MultiTEV a été suivi. En effet, le CTD de l'IN de PERV-A/C est toujours porteur d'une étiquette Strep-tag[®] et le nouveau fragment Brd2 (667-735) possède une étiquette poly-histidine à son extrémité amino-terminale, ce qui permet la purification du complexe en deux étapes de chromatographie.

La première étape de purification a consisté en une chromatographie d'affinité pour le nickel qui, en théorie, permet de purifier Brd2 (667-735) seule ou en complexe avec le CTD de l'IN. Le lysat bactérien a été injecté sur une colonne Ni-NTA et suivi par mesure de l'absorbance à 280 nm. La conductance du tampon a également été mesurée tout au long de la chromatographie (**Figure 57A**).

Sur le chromatogramme de purification, deux fortes augmentations de l'absorbance à 280 nm sont visibles, la première correspondant à la sortie de colonne des protéines non retenues et la seconde à l'élution des protéines. En raison d'un dysfonctionnement du système Aktä Purifier (*GE Healthcare*) présent dans le laboratoire à cette période, le gradient d'imidazole pour l'élution des protéines n'a pu être réalisé. Cette dernière a donc été faite avec un tampon

contenant une forte quantité d'imidazole (soit 1 M). Puis l'échantillon a été dilué dans le même tampon sans imidazole, afin d'abaisser la concentration de ce dernier jusqu'à 250 mM, avant injection sur colonne Strep-Tactin®. Une augmentation de la conductance du milieu signale le passage du tampon de lavage, mais celui-ci ne semble pas entraîner l'élution de protéines.

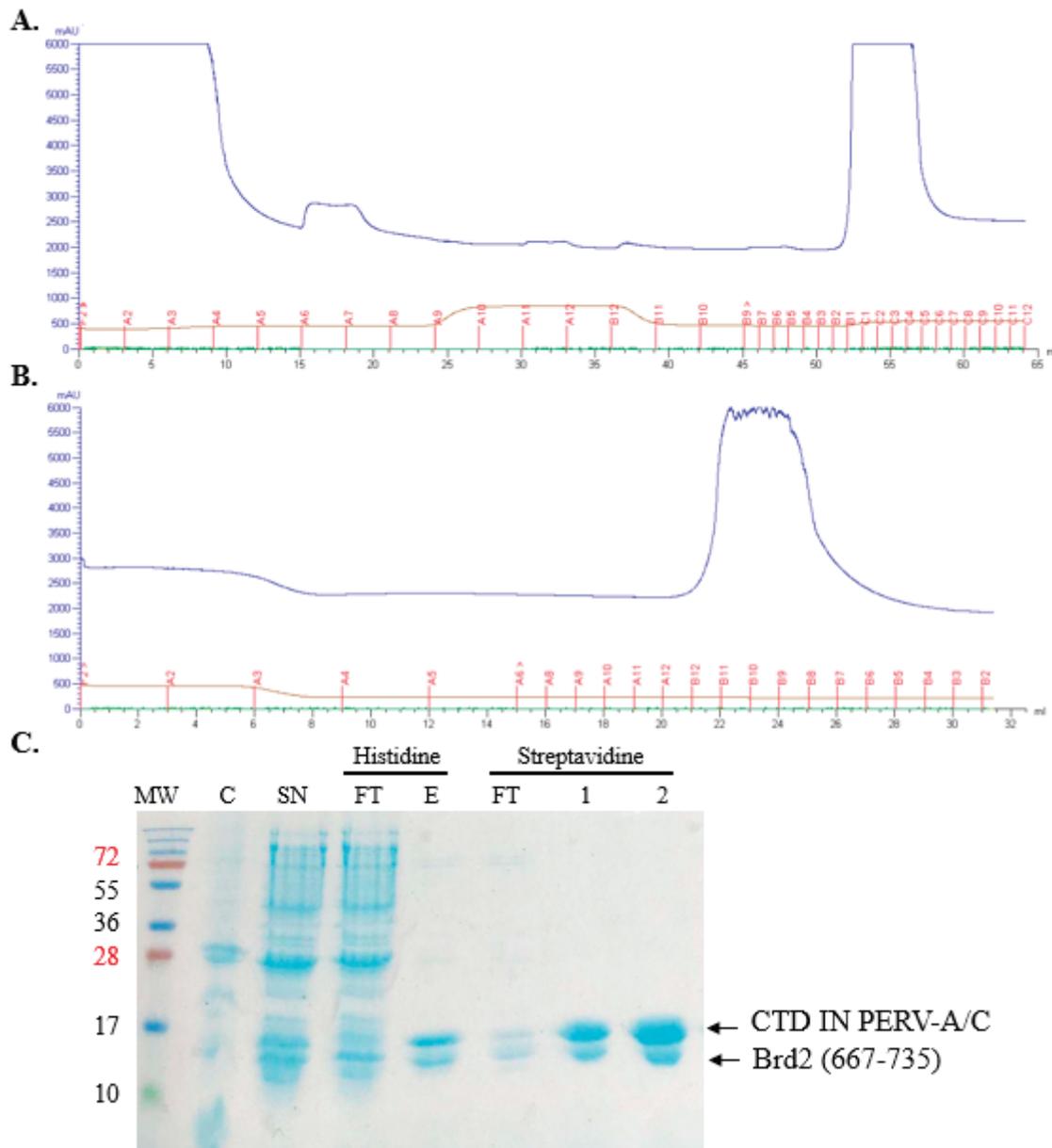


Figure 57 | Purification en deux étapes du complexe MultiTEV Minimal. **A.** Chromatogramme de purification par chromatographie d'affinité pour le nickel. En bleu, l'absorbance à 280 nm et en marron, la conductance. **B.** Chromatogramme de purification par chromatographie d'affinité pour la streptavidine. En bleu, l'absorbance à 280 nm et en marron, la conductance. **C.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; piste C, fraction non soluble ; piste SN, fraction soluble ; pistes FT, fractions non retenues ; piste E, fraction d'élution de la chromatographie d'affinité pour le nickel et pistes 1 et 2, fractions d'élution de la chromatographie d'affinité pour la streptavidine.

Les fractions de purification sont analysées par SDS-PAGE (**Figure 57C**). La fraction soluble résultant de la lyse bactérienne est déposée en piste SN. Deux bandes majoritaires sont visibles entre les marqueurs de poids moléculaires de 10 kDa et 17 kDa, correspondant au CTD de PERV-A/C (12 kDa) et à Brd2 (667-735) (9 kDa). Les fractions d'élution groupées et diluées sont déposées en piste E où, à nouveau, deux bandes correspondant aux protéines d'intérêt peuvent être observées. La rétention du CTD de l'IN sur la colonne malgré l'absence d'étiquette poly-histidine peut être expliquée par son interaction avec Brd2 (667-735). En revanche, quelques bandes indiquant la présence de contaminants de hauts poids moléculaires sont visibles.

Dans un second temps, les fractions d'élution diluées ont été injectées sur colonne Strep-Tactin[®] pour la purification par chromatographie d'affinité pour la streptavidine. L'absorbance à 280 nm et la conductance des tampons sont mesurées en sortie de colonne.

Deux augmentations de l'absorbance à 280 nm sont observées sur le chromatogramme (**Figure 57B**). La première, de faible intensité, indique la sortie de colonne des protéines non retenues et la deuxième, de plus forte intensité, représente la sortie de colonne de l'échantillon lors de l'élution par compétition avec la desthiobiotine.

Les fractions de purification sont analysées par SDS-PAGE (**Figure 57C**). La piste FT, représentant la fraction non retenue sur la colonne, montre des bandes de hauts poids moléculaires indiquant la présence des contaminants protéiques. Les deux bandes correspondant aux protéines d'intérêt sont également observées et cela semble provenir de la saturation en protéines de la colonne. Les pistes 1 et 2, quant à elles, représentent les fractions d'élution. Deux bandes sont observées pour ces fractions ; une bande supérieure correspondant au CTD de l'IN de PERV-A/C et une bande inférieure correspondant à Brd2 (665-735).

De la même façon que pour le complexe MultiTEV, les deux protéines sont bien surproduites et purifiées, témoignant de la formation du complexe MultiTEV Minimal. Le rendement de purification est de 3 à 5 mg de protéines par litre de culture.

I. 2. 2. Caractérisation du complexe MultiTEV Minimal

Les protéines purifiées ont ensuite été analysées afin de déterminer l'état oligomérique du complexe ainsi que sa pureté en solution.

I. 2. 2. 1. Chromatographie d'exclusion de taille

Tout d'abord, l'échantillon composé du complexe MultiTEV Minimal a été analysé par chromatographie d'exclusion de taille. L'échantillon a été injecté sur la colonne Superdex 200 Increase 10/300 GL permettant de séparer les protéines selon leur poids moléculaire et, simultanément, de changer le tampon dans lequel se trouve l'échantillon pour le tampon final ne contenant pas de desthiobiotine.

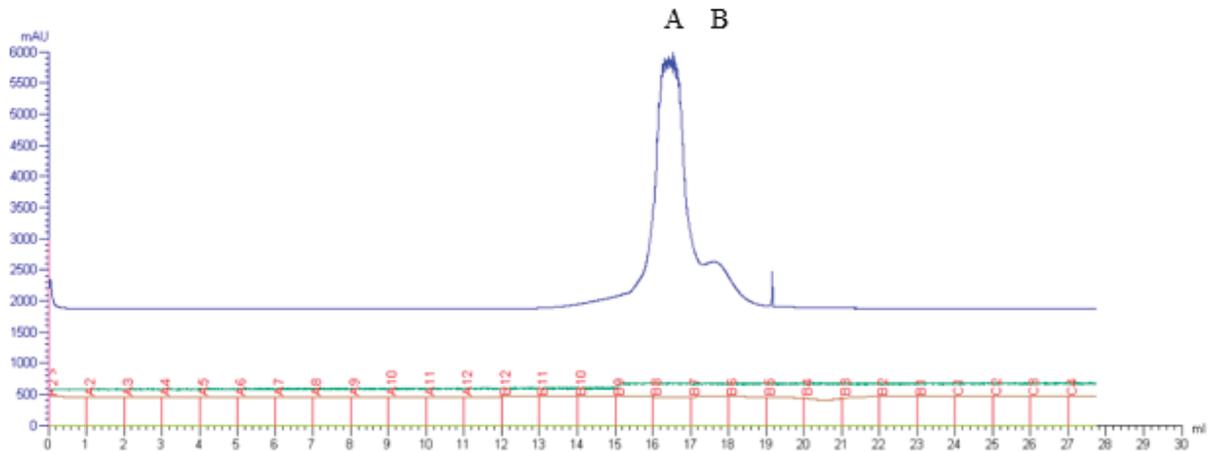


Figure 58 | Chromatographie d'exclusion de taille du complexe MultiTEV Minimal. En bleu, le suivi de l'absorbance à 280 nm.

L'absorbance à 280 nm est suivie tout au long de l'élution et permet d'identifier deux pics représentant la sortie de colonne de protéines (**Figure 58**). Les pics A et B correspondent à des fractions éluées à des volumes de 16,5 mL et 18 mL, respectivement. Le pic A, de plus grande surface, semble de nouveau correspondre à un complexe hétérodimérique formé cette fois d'un monomère de Brd2 (667-735) et d'un monomère du CTD de l'IN de PERV-A/C. Le pic B, quant à lui de plus petite surface, peut correspondre aux deux protéines dissociées, non séparées en raison de la résolution de la colonne.

La fraction correspondant au pic A a été analysée par SDS-PAGE tandis que son état oligomérique était vérifié par pontage covalent, comme décrit ci-dessous.

I. 2. 2. 2. Pontage covalent

L'expérience de pontage covalent a été réalisée avec l'agent pontant GTA à la suite de la chromatographie d'exclusion de taille. Les échantillons ont ensuite été analysés par SDS-PAGE (Figure 59).

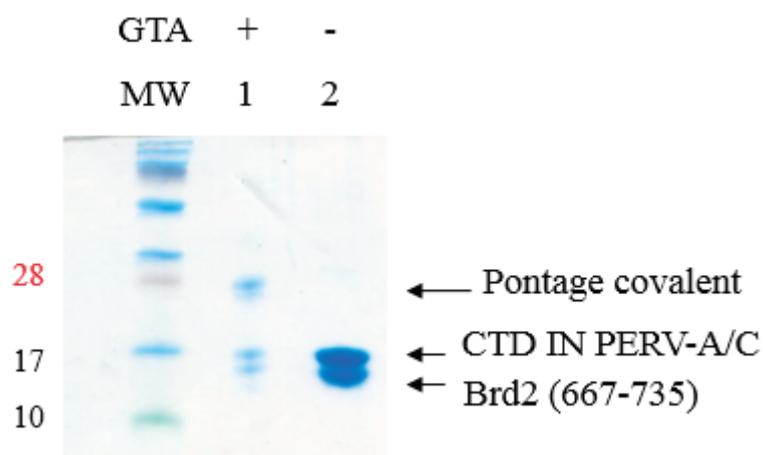


Figure 59 | Analyse par SDS-PAGE de l'état oligomérique du complexe MultiTEV Minimal après pontage covalent au GTA. Piste 1 : échantillon incubé avec l'agent de pontage. Piste 2 : échantillon sans agent de pontage. Avec MW, les marqueurs de poids moléculaires.

L'échantillon en sortie de colonne de chromatographie d'exclusion de taille, traité sans agent pontant, a été déposé en piste 2. On peut y observer deux bandes, la supérieure correspondant au CTD de l'IN de PERV-A/C et la bande inférieure correspondant à Brd2 (667-735). L'échantillon, après ces étapes de purification, est donc bien pur et le rapport des deux protéines semble stœchiométrique.

L'échantillon traité avec la GTA est déposé en piste 1 et une bande migrant au niveau du marqueur de poids moléculaire de 28 kDa peut être observée. Celle-ci correspond à la taille attendue pour un complexe hétérodimérique formé de Brd2 (667-735) et du CTD de l'IN de PERV-A/C. Les deux fragments protéiques ont également été identifiés dans cette bande de plus haut poids moléculaire par spectrométrie de masse.

I. 2. 2. 3. Diffusion dynamique de la lumière (DLS)

Afin d'obtenir une estimation de sa pureté, l'échantillon provenant du pic A en sortie de chromatographie d'exclusion de taille a également été analysé en DLS à une concentration de

9 mg/mL. La courbe de distribution de l'intensité en fonction de la taille obtenue est représentée ci-dessous (**Figure 60**).

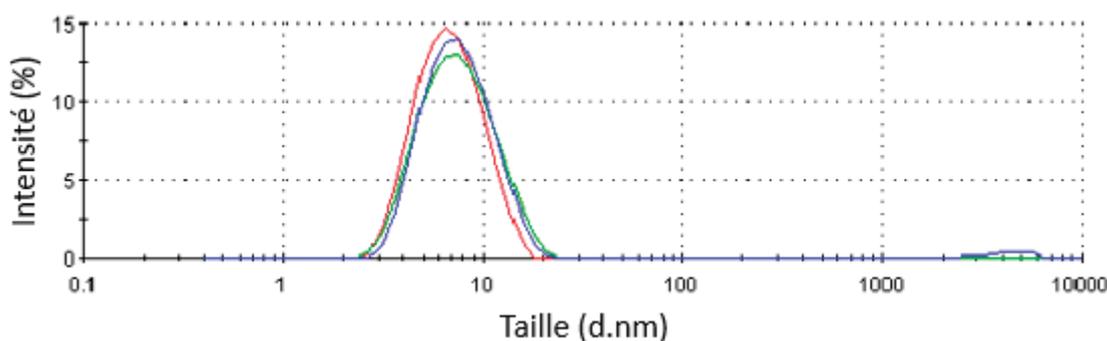


Figure 60 | Analyse de la monodispersité du complexe MultiTEV Minimal par diffusion dynamique de la lumière. Les courbes bleue, rouge et verte représentent les trois mesures successives.

Cette analyse par DLS montre la présence d'une seule espèce dans l'échantillon. De plus, l'indice de polydispersité de 0,16 confirme que l'échantillon protéique est homogène et monodisperse, qualités nécessaires pour les expériences subséquentes de cristallogénèse.

I. 2. 3. Études structurales du complexe MultiTEV Minimal

I. 2. 3. 1. Étude structurale par cristallographie aux rayons X

Le complexe protéique formé du CTD de l'IN de PERV-A/C et de Brd2 (667-735), purifié après chromatographie d'exclusion de taille, a été concentré par centrifugation et des expériences de cristallogénèse ont été initiées.

Les kits utilisés pour ces essais de cristallisation sont les suivants : PEG/Ion HT™ (*Hampton Research*), PEGs Suite (*Qiagen*), PEGs II Suite (*Qiagen*), Protein Complex Suite (*Qiagen*), AmSO₄ Suite (*Qiagen*), JCSG+ Suite (*Qiagen*), Morpheus (*Molecular Dimensions*) et MPD Suite (*Qiagen*)

L'ensemble des conditions testées sont résumées en **Tableau 8**.

		Kits							
		PEG/Ion HT™	PEGs Suite	PEGs II Suite	Protein Complex Suite	AmSO ₄ Suite	JCSG+ Suite	Morpheus	MPD Suite
Concentrations (en mg/mL)	8,2			Jaune	Rouge				
	9,8	Jaune	Rouge	Jaune			Jaune		
	11	Jaune				Rouge		Rouge	Rouge
	26	Jaune			Rouge		Jaune		
	27,2	Jaune	Rouge	Jaune			Jaune		

Tableau 8 | Tableau récapitulatif des concentrations protéiques et kits de criblage des conditions de cristallisation testés pour le complexe entre Brd2 (667-735) et le CTD de l'IN de PERV-A/C. En rouge, les kits de criblage testés n'ayant pas mené à l'apparition de cristaux ; en jaune, ceux ayant conduit à l'apparition de cristaux.

Dans plusieurs conditions, de petits cristaux sont apparus après quelques jours. Il s'agit de la condition 79 du kit PEGs II Suite, des conditions 10, 11, 34 et 41 du kit JCSG+ Suite et des conditions 4, 33, 40 et 72 du kit PEG/Ion HT™. Les cristaux obtenus ont été pêchés et cryo-protégés afin d'être soumis aux rayons X à l'ESRF de Grenoble.

Une grande partie de ceux-ci ont diffracté avec des profils caractéristiques des cristaux de sel. En revanche, les cristaux provenant des conditions 4 (0,2 M chlorure de lithium, 20% PEG 3350, pH 6,8) et 40 (0,2 M phosphate de sodium dibasique, PEG 3350, pH 9,1) du kit PEG/Ion HT™ n'ont pas diffracté, ce qui pouvait être dû à leur très petite taille. Ces deux dernières conditions ont donc été reproduites dans des plaques 24 puits, afin d'augmenter la taille des cristaux. Plusieurs rapports protéine:solution de cristallisation (1:1, 2:1 et 1:2) ainsi que plusieurs concentrations protéiques ont été testés dans ce but. Cependant, aucun nouveau cristal n'a été obtenu à la suite de ces essais.

1. 2. 3. 2. Étude structurale par SAXS

Parallèlement aux essais de cristallisation, des données SEC-SAXS ont été enregistrées sur le complexe MultiTEV minimal, concentré à 15 mg/mL, sur la ligne de lumière BioSAXS BM29 de l'ESRF de Grenoble.

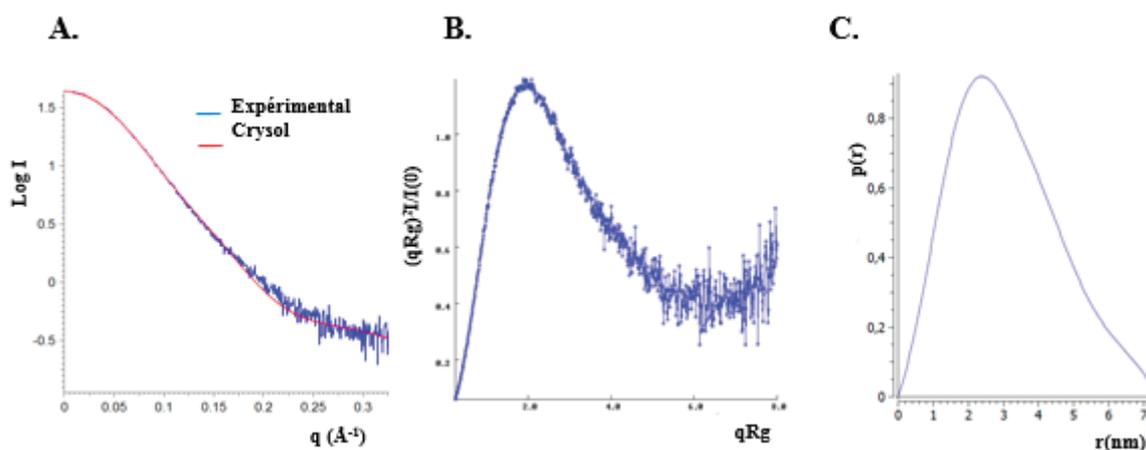


Figure 61 | Analyse SAXS du complexe MultiTEV Minimal. **A.** Courbe de diffusion expérimentale des rayons X aux petits angles (en bleu) et courbe CRYSol calculée d'après le modèle atomique (en rouge). **B.** Représentation de Kratky. **C.** Fonction de distribution des distances $P(r)$. Résultats et figures obtenus avec ATSAS.

La courbe de diffusion obtenue est de bonne qualité jusqu'à une valeur de q de $3,2 \text{ nm}^{-1}$ (**Figure 61A**). Elle montre, aux petites valeurs de q , que le tampon est neutre pour l'échantillon (pas d'effets d'agrégation ou de répulsion observés) et le R_g calculé selon l'équation de Guinier est de $2,4 \text{ nm}$ (**Tableau 9**). Une valeur identique est obtenue en calculant le R_g d'après la fonction de distribution des distances $P(r)$ (**Figure 61C** et voir **Matériels et Méthodes §V.1.**), ce qui constitue également un critère de qualité des mesures. Le D_{\max} calculé d'après la $P(r)$ est de $7,4 \text{ nm}$.

Enregistrement des données	
Fragment analysé	IN CTD – Brd2 (667 - 735)
Ligne de lumière	BM29 (ESRF, Grenoble)
q-range (nm ⁻¹)	0 – 5
Temps d'exposition (s)/nb frames	1/10
Concentration (mg/mL)	15
Température (°C)	4
Analyse des données	
Réduction des données	EDNA
Traitement des données	CHROMIXS
q-range (nm ⁻¹)	0,1 – 3,2
Rg (nm) [d'après Guinier]	2,4 ± 0,07
Rg (nm) [d'après P(r)]	2,4
Dmax (nm)	7,4
Analyse <i>ab initio</i>	DAMMIF
Nombre de modèles	50
Validation et moyenne	DAMAVER
NSD	0,87 ± 0,06
Modélisation 'rigid-body'	DADIMODO
Test χ^2	1,0

Tableau 9 | Statistiques d'enregistrement et de traitement des données SAXS ainsi que de génération du modèle *ab initio*.

Cinquante modèles *ab initio* de l'échantillon ont ensuite été générés par affinement contre les données SAXS. Ces modèles en billes ont été superposés et moyennés avec les programmes DAMSUP et DAMAVER de la suite ATSAS (Franke *et al.*, 2017). L'enveloppe SAXS résultante présente une forme de cône tronqué à base elliptique avec une résolution estimée de 3 nm (**Figure 62**).

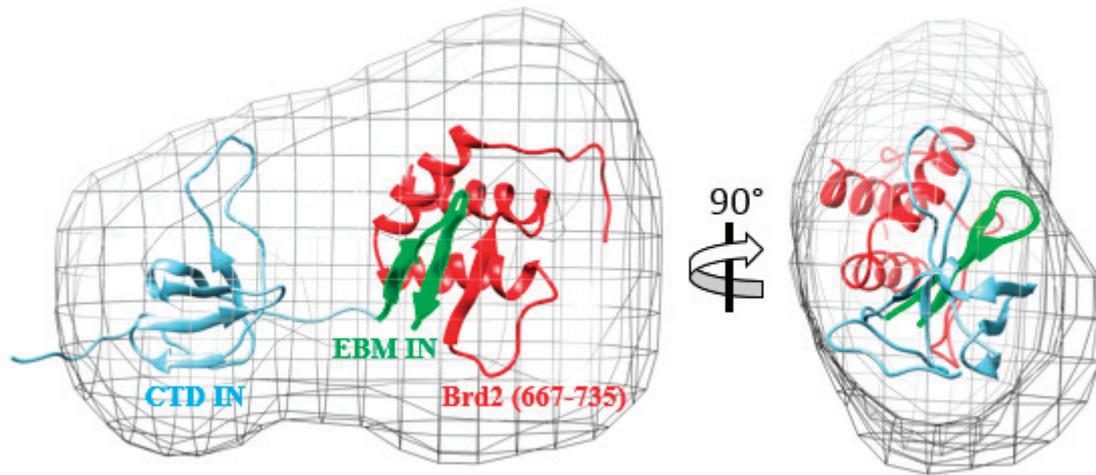


Figure 62 | Enveloppe SAXS *ab initio* (en gris) et représentation en ruban du modèle SAXS du MultiTEV Minimal avec le CTD de l'IN de PERV-A/C en bleu, son domaine EBM en vert et le fragment Brd2 (667-735) en rouge. Figure réalisée avec le logiciel Chimera.

Un modèle atomique du MultiTEV Minimal a alors été généré. Cette modélisation était possible car le CTD de l'IN de PERV-A/C et l'ET de Brd2 possèdent chacun une identité de séquence significative avec le CTD de l'IN de MLV et l'ET de Brd4 (respectivement 60% et 86%) dont les structures sont connues (Aiyer *et al.*, 2014; Crowe *et al.*, 2016). Nous avons également supposé que le segment EBM (378-394) du CTD de PERV-A/C présentait, dans le complexe MultiTEV Minimal, un repliement en épingle à cheveux similaire à celui observé dans la structure RMN d'EBM de MLV lié à ET de Brd4 (Crowe *et al.*, 2016).

Ce modèle a ensuite été affiné contre nos données SAXS en utilisant une stratégie à deux corps rigides. Le premier corps était constitué du segment EBM lié à ET tandis que le deuxième corps était formé du CTD de l'IN sans sa partie EBM. La boucle 373-377 reliant le CTD au segment EBM a été laissée flexible durant l'affinement.

Cette stratégie a permis de positionner expérimentalement le CTD de l'IN (acides aminés 309 à 372) de façon relative à son domaine EBM (378-394) lié à l'ET de Brd2 (**Figure 62**). Le modèle SAXS qui en résulte présente un χ^2 de 1, indicateur de sa bonne concordance avec les données expérimentales.

Ce modèle montre deux blocs structuraux distincts, constitués du cœur de type SH3 du CTD et d'ET:EBM, qui sont reliés par la boucle flexible 373-377 du CTD. Cette structure partiellement flexible à deux domaines est en accord avec le tracé de la représentation de Kratky, qui présente un plateau aux hautes valeurs de q (**Figure 61B**).

Le modèle SAXS suggère également qu'avant l'étape d'intégration, la queue carboxy-terminale flexible de l'IN gammarétrovirale va cibler, tel un hameçon, le domaine ET de son cofacteur cellulaire. Lors de l'interaction, le segment EBM de l'IN se replie pour former un feuillet β à trois brins avec le domaine ET, bloquant les deux partenaires ensemble.

Notre modèle confirme également le rôle essentiel joué par le segment EBM pour lier les protéines BET. Il peut permettre de proposer une explication aux résultats que nous avons obtenus avec les mutant disruptifs W379A et R391A/L392A lors des tests d'interaction *in vitro* (voir **Résultats §I.1.4.2.**) tout en gardant à l'esprit que la résolution de la structure SAXS n'est pas suffisante pour observer des détails atomiques et que ceux-ci sont déduits par modélisation.

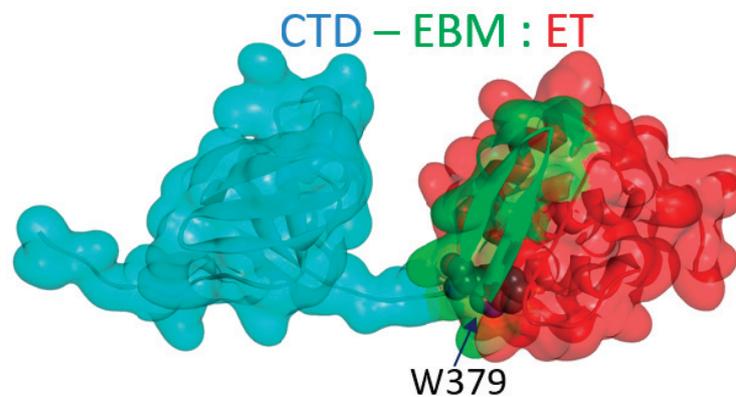


Figure 63 | Position du résidu W379 dans la structure SAXS du complexe MultiTEV Minimal. Le complexe est représenté sous forme de ruban à l'intérieur de sa surface de van der Waals. La partie 309-377 du CTD de l'IN de PERV-A/C est en bleu, le segment EBM (378-394) du CTD est en vert et le domaine lié ET de Brd2 en rouge. Le résidu W379 du CTD, coloré en gris, est représenté en mode boule.

Il apparaît ainsi que le résidu W379 occupe une place centrale dans la zone d'interaction EBM:ET du complexe MultiTEV minimal (**Figure 63**). En effet, ce tryptophane se trouve enfoui au cœur de l'interface intermoléculaire, son noyau indole étant en contact avec les chaînes latérales des résidus R394 et L392 d'EBM mais aussi L697 et V701 d'ET, le tout avec des distances inférieures à 4 Å. Ces interactions, majoritairement hydrophobes, stabilisent le feuillet β intermoléculaire, tandis que le mutant déficient W379A ne peut pas induire ce point de fixation.

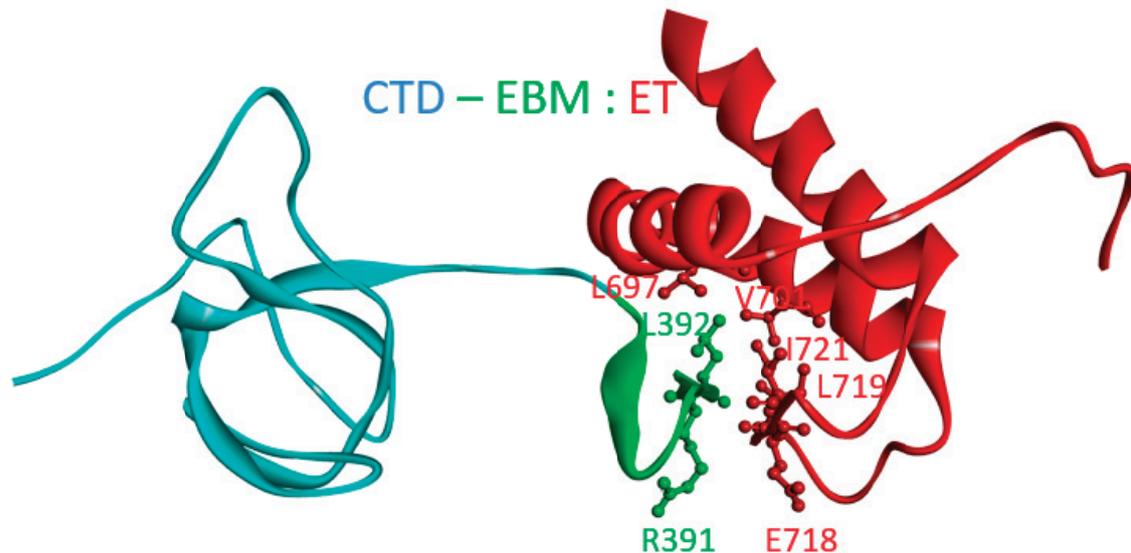


Figure 64 | Position des résidus R391 et L392 dans la structure SAXS du complexe MultiTEV Minimal. Le complexe est représenté sous forme de ruban avec la partie 309-377 du CTD de l'IN de PERV-A/C en bleu, le segment EBM du CTD en vert et le domaine lié ET de Brd2 en rouge. Les résidus R391 et L392 sont en représentation boule et bâton verte, ceux avec lesquels ils interagissent, en rouge (distance inférieure à 4 Å).

Les résidus R391 et L392 du CTD sont, quant à eux, localisés dans le brin central du feuillet β (**Figure 64**). Le résidu arginine est orienté vers la surface du complexe et côtoie un glutamate du domaine ET, tandis que le résidu leucine est orienté vers une poche hydrophobe de ET qui comprend les résidus L697, V701, L710 et I721. Outre le fait que le double mutant alanine montre que ces interactions sont essentielles à la formation du complexe, il est intéressant de noter que cette double mutation affecte le motif -BHBH- (389-392) du CTD, où -B- et -H- sont des résidus basiques et hydrophobes. Ce motif -BHBH- est conservé chez les *Gammarrétrovirus* (**Figure 48**) et est observé chez d'autres partenaires viraux et cellulaires des protéines BET, où il est essentiel pour la liaison avec le domaine ET (Zhang *et al.*, 2016).

Enfin, nous avons voulu savoir si les résultats obtenus sur ce modèle à deux fragments protéiques pouvaient être plus largement étendus à l'assemblage de l'intasome gammarrétroviral avec les protéines BET. Les données structurales sur ces deux partenaires sont toutefois limitées, car aucune structure expérimentale d'intasome de *Gammarrétrovirus* n'a encore été publiée et seules les structures des domaines BD et ET des protéines BET sont connues.

L'état oligomérique des intasomes est différent selon les genres rétroviraux, cette propriété semblant provenir de la taille de la boucle reliant le CCD et le CTD de l'IN (**Figure 22**). Comme

les dimensions de cette boucle sont similaires entre les *Gammarétrovirus* et les *Spumavirus*, on peut supposer que l'intasome gammarétroviral est tétramérique comme celui de PFV (Engelman & Cherepanov, 2017). Le modèle 3D de l'intasome de PERV-A/C, publié par l'équipe en 2015 (Demange *et al.*, 2015), est ainsi basé sur la structure cristalline de l'intasome de PFV. Il compte deux monomères d'IN internes, qui vont réaliser la réaction enzymatique d'intégration, et deux monomères externes. Nous avons superposé le CTD d'un des monomères internes avec le CTD de notre modèle SAXS afin de générer un modèle de l'intasome PERV-A/C lié au domaine ET de Brd2 (**Figure 65**).

Cette superposition n'engendre pas de *clashes* stériques et montre que les CTD des monomères internes pourraient interagir avec le domaine ET des protéines BET sans réarrangement structural significatif. Cette accessibilité est conservée si le CTD du MultiTEV Minimal est superposé avec un des CTD internes de la structure cristalline de l'intasome lié à son ADN cellulaire cible (Maertens *et al.*, 2010). Enfin, on peut remarquer que la flexibilité de la petite boucle reliant le cœur SH3 du CTD à l'extrémité EBM peut faciliter l'assemblage des deux partenaires et aider au ciblage efficace des sites d'intégration de PERV-A/C.

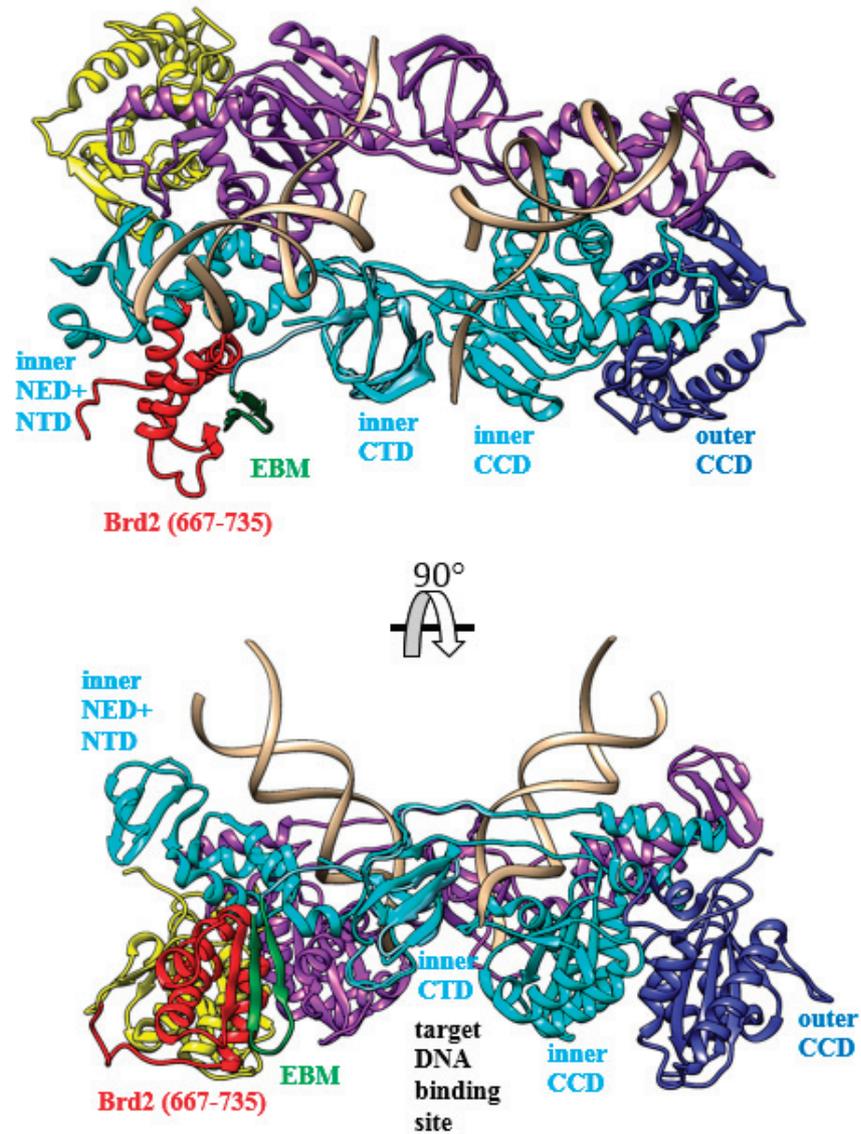
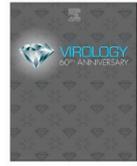


Figure 65 | Représentation en ruban du modèle de l'intasome de PERV-A/C superposé avec le modèle SAXS de MultiTEV Minimal (CTD de l'IN de PERV-A/C et Brd2 (667-735)). Avec en cyan et violet, les deux monomères d'IN internes ; en bleu et jaune, les deux monomères externes ; en gris, les brins d'ADN viraux ; en rouge, Brd2 (667-735) et en vert, le fragment EBM de l'IN de PERV-A/C. La superposition a été réalisée avec le logiciel COOT et la figure produite avec le logiciel Chimera.



In vitro, *in cellulo* and structural characterizations of the interaction between the integrase of Porcine Endogenous Retrovirus A/C and proteins of the BET family



Kathy Gallay^{a,b,c,1}, Guillaume Blot^{d,1}, Margaux Chahpazoff^{e,1}, Halima Yajjou-Hamalian^e, Marie-Pierre Confort^{a,b,c}, Claire De Boissésou^d, Aurélie Leroux^d, Catherine Luengo^{a,b,c}, Francesca Fiorini^e, Marc Lavigne^f, Yahia Chebloune^g, Patrice Gouet^e, Karen Moreau^{a,b,c,3}, Yannick Blanchard^{d,*,2}, Corinne Ronfort^{a,b,c,*,2}

^a INRA, Université Lyon 1, UMR754, Viral Infections Compared Pathology, 69007, Lyon, France

^b Université de Lyon, 69000, Lyon, France

^c UMS13444 Biosciences Gerland Lyon Sud, 69007, Lyon, France

^d ANSES, Ploufragan/Plouzané Laboratory, Viral Genetics and Bio-Security Unit, Université Européenne de Bretagne, Ploufragan, France

^e Molecular Microbiology and Structural Biochemistry, MMSB-IBCP, UMR 5086 CNRS University of Lyon, 7, passage du Vercors, 69367, Lyon, Cedex 07, France

^f Pasteur Institute, Virology dept., Institut Cochin-Inserm U1016-CNRS UMR8104-Université Paris Descartes, 27, rue du faubourg Saint Jacques, 75014, Paris, France

^g Pathogenesis and Lentiviral Vaccination Laboratory, INRA, Univ Grenoble Alpes, 570 rue de la Chimie, 38400, St Martin d'Hères, France

ARTICLE INFO

Keywords:

Gamma-retrovirus
Integration
IN co-factors
BET proteins
SAXS

ABSTRACT

Retroviral integrase (IN) proteins catalyze the permanent integration of the viral genome into host DNA. They can productively recruit cellular proteins, and the human Bromodomain and Extra-Terminal domain (hBET) proteins have been shown to be co-factors for integration of gamma-retroviruses such as Murine Leukemia Virus (MLV) into human cells. By using two-hybrid, co-immunoprecipitation and *in vitro* interaction assays, we showed that IN of the gamma- Porcine Endogenous Retrovirus-A/C (PERV IN) interacts through its C-terminal domain (CTD) with hBET proteins. We observed that PERV IN interacts with the BRD2, BRD3 and BRD4 proteins *in vitro* and that the BRD2 protein specifically binds and co-localizes with PERV IN protein in the nucleus of cells. We further mapped the interaction sites to the conserved Extra-Terminal (ET) domain of the hBET proteins and to several amino acids of the C-terminal tail of the PERV IN CTD. Finally, we determined the first experimental structure of an IN CTD – BET ET complex from small-angle X-ray scattering data (SAXS). We showed that the two factors assemble as two distinct modules linked by a short loop which confers partial flexibility. The SAXS-restrained model is structurally compatible with the binding of the PERV intasome to BRD2. Altogether, these data confirm the important role of host BET proteins in the gamma-retroviruses' targeting site and efficiency of integration.

1. Background

During the retroviral life cycle, integrases (INs) that catalyze the key step of integration for the provirus of the retrotranscribed DNA into the host cell genome can interact with cellular cofactors to target specific integration sites (Kvaratskhelia et al., 2014). Hence, the Human

Immunodeficiency Virus type 1 (HIV-1) binds the host Lens Epithelium-Derived Growth Factor (LEDGF/p75) protein (Busschots et al., 2005; Cherepanov et al., 2003; Emiliani et al., 2005; Llano et al., 2006; Maertens et al., 2003) to preferentially target the body of actively transcribed genes (Mitchell et al., 2004; Schroder et al., 2002). More recently, the cleavage and polyadenylation specificity factor 6 protein

* Corresponding author. INRA, Université Lyon 1, UMR754, Viral Infections Compared Pathology, 69007, Lyon, France.

** Corresponding author.

E-mail addresses: yannick.blanchard@anses.fr (Y. Blanchard), corinne.ronfort@inra.fr (C. Ronfort).

¹ KG, GB and MC contributed equally.

² YB and CR contributed equally.

³ Present address: CIRI, International Center for Infectiology Research, INSERM U1111, Université Lyon 1, Ecole Normale Supérieure de Lyon, CNRS UMR5308, 7 rue Guillaume Paradin, 69008, Lyon, France.

<https://doi.org/10.1016/j.virol.2019.04.002>

Received 12 January 2019; Received in revised form 30 March 2019; Accepted 9 April 2019

Available online 16 April 2019

0042-6822/ © 2019 Published by Elsevier Inc.

(CPSF6) complexed with the HIV-1 capsid (CA) protein has been shown to play a dominant role in directing integration within euchromatin (Sowd et al., 2016). Integration of Avian Sarcoma and Leukemia Viruses (ASLV) is more random (Mitchell et al., 2004) and no integration cell cofactor directing ASLV integration has been reported to date even though the ASLV IN-binding factor FACT (FACilitates Chromatin Transcription) might be a good candidate (Winans et al., 2017). Also, it remains to be seen whether the delta-retroviral IN-binding serine/threonine protein phosphatase 2A (PP2A) protein plays a role in directing integration of delta-retroviruses (Maertens, 2016). Foamy virus (FV) integration seems to also be random, but FVs tend to target promoter-closed regions when integration into a transcribed region has occurred (Nowrouzi et al., 2006).

Concerning gamma-retroviruses, the Murine Leukemia Virus (MLV) favors integration near the promoters, enhancers, CpG islands and DNase I hypersensitive sites (De Ravin et al., 2014; LaFave et al., 2014; Mitchell et al., 2004; Wu et al., 2003). Numerous groups (De Rijck et al., 2013; Gupta et al., 2013; Sharma et al., 2013) have confirmed earlier preliminary data (Studamire and Goff, 2008) indicating a potential binding of MLV IN with the Bromodomain-containing protein Brd2, and they have shown that Brd2, 3 and 4 target the integration of MLV in transcription start sites. These three ubiquitously expressed proteins of the Bromodomain and Extra-Terminal (BET) family specifically bind MLV IN and co-localize in the cell nucleus. Furthermore, the integration site distribution of MLV corresponds to the chromatin-binding profile of BET proteins (De Rijck et al., 2013). Brd4 was reported to interact with other gamma-retroviral INs, such as those of Feline Leukemia Virus (FeLV) (Gupta et al., 2013) and avian Reticulo-Endotheliosis Virus (REV) (Serrao et al., 2015). In addition, a recent paper has pinpointed an important role of the MLV CA protein in facilitating the targeting of the pre-integration complex to nucleosomes via a mechanism that could be similar to the one of the HIV-1 CA–CPSF6 complex (Wanaguru et al., 2018).

Members of the BET protein family contain tandem bromodomains (BD-I and BD-II) and a conserved Extra-Terminal domain (ET) (Thorpe et al., 1996). The BET family contains chromatin-binding factors that have been implicated in the control of cell cycle progression, transcription and DNA replication (Belkina and Denis, 2012; Wu and Chiang, 2007). These proteins bind to the acetylated tails of histones H3 and H4 via their bromodomains. The extra-terminal motif ET in the C-terminal region of BET proteins was shown to interact with numerous cellular — e.g. transcription factors and histone modification enzymes — and viral factors. Interestingly, BRD2 was shown to interact with Kaposi's Sarcoma-Associated Herpesvirus (KSHV) and the latency-associated nuclear antigen 1 (LANA 1), a protein required for the replication of the episomal genome in dividing cells (Viejo-Borbolla et al., 2005). The C-terminal part of the long isoform of BRD4 can bind the bovine papilloma virus type 1 E2 protein (Jang et al., 2009).

Retroviral INs are composed of three conserved domains — i.e. N-terminal domain (NTD), catalytic core domain (CCD) and C-terminal domain (CTD) — joined by flexible linkers (see Fig. 1A) which affect crystallization trials. For this reason structural information was obtained only for one- or two-domain IN fragments for a decade (reviewed by (Andrake and Skalka, 2015)). The NTD was characterized with an all- α fold, the CCD with an RNase-H fold and the CTD with an SH3 fold. Since 2010, crystal and cryo-electron microscopy structures of intasomes (i.e. the entire IN complex with bound viral DNA ends) have been published for spumavirus, lentivirus, alpha-retrovirus and beta-retrovirus genera (Ballandras-Colas et al., 2016, 2017; Hare et al., 2010; Passos et al., 2017; Yin et al., 2016). Concerning gamma-retroviruses, the isolated NTD and CTD structures of MLV IN are known (Aiyer et al., 2014; Guan et al., 2017), and we have recently published a model of the Porcine Endogenous Virus A/C (PERV-A/C) intasome (Demange et al., 2015). BET proteins also display a conserved modular architecture and the isolated structures of BD-I, BD-II and ET have been determined, each with an all- α fold ((Lin et al., 2008); for a review, see (Sanchez and

Zhou, 2009)). The ET binding motif (EBM) of MLV IN was located within the C-terminal tail of the CTD (Aiyer et al., 2014), as observed in the nuclear magnetic resonance (NMR) structure of MLV IN EBM complexed with BRD4 ET (Crowe et al., 2016). The latter study revealed that the 17-long flexible EBM segment (MLV IN aa 389–405) folds into a β -hairpin when it binds the ET domain (Crowe et al., 2016).

We have previously shown that the integration of the gamma-retrovirus PERV-A/C (hereafter referred to as PERV) in human cells is not random and is strongly favored at transcriptional start sites and CpG islands (Moalic et al., 2006, 2009). These results revealed similarities between PERV and MLV integration site selection, suggesting that gamma-retroviruses have a common pattern of integration and that their cellular integration target cofactors might be similar. However, this assumption remains to be demonstrated. In this study, we show *in vitro* and *in cellulo* that PERV IN interacts strongly with BET proteins like MLV IN, in contrast with the behavior exhibited by ASLV and HIV-1 INs. We further perform small-angle X-ray Scattering (SAXS) experiments on the PERV IN CTD – BRD2 ET complex and generate a 3D model that corroborates our interaction results.

2. Materials and methods

2.1. Plasmids

Plasmids for the two-hybrids assay. PERV, MLV, HIV-1 and RAV-1 INs — as well as BRD2 encoding sequences — were amplified by PCR from the recombinant PERV-A/C clone AY570980 (supplied by Takeuchi, Y.), the pHit60 (supplied by S. P. Goff), pET15b-IN-HIV-1 (Cellier et al., 2013), pET30a-IN-RAV-1 (Moreau et al., 2004), GST-Brd2/RingG3B, GST-Brd4/HUNK, GST-Brd43/ORF XRingG3B (supplied by T. Schulz), respectively, using appropriate primers (Tables S1 and S2). The resulting fragments were cloned into either the pGBKT7 plasmid (for INs) or the pGAD-T7 plasmid (for BRD2, 3, 4) using the Matchmaker Gold Yeast Two Hybrid System (Clontech). Deletion constructs were generated for PERV IN (pGBKT7-PERV IN (1–98), pGBKT7-PERV IN (99–272), pGBKT7-PERV IN (273–405), pGBKT7-PERV IN (1–272), pGBKT7-PERV IN (99–405), pGBKT7-PERV IN (99–394), pGBKT7-PERV IN (99–383), pGBKT7-PERV IN (372–405) and for MLV IN (pGBKT7-MLV IN (1–276), pGBKT7-MLV IN (277–408) by PCR fusion of DNA fragments encoding the corresponding amino acid (aa) residues from either the PERV or the MLV INs, respectively, with the Gal4-DNA binding domain. Point mutations (pGBKT7-PERV INmut (P387A/L388A), pGBKT7-PERV INmut (K389A/L390A), pGBKT7-PERV INmut (R391A/L392A), pGBKT7-PERV INmut (H393A/R394A) of the full-length PERV IN were generated by PCR-directed mutagenesis with the appropriate primers. The oligonucleotides used are listed in Table S1. All constructs were verified by DNA sequencing.

Cloning of the BRD2-encoding cDNA from HEK293T cells into the pCMV-Myc plasmid. The full-length BRD2 coding sequences were isolated from HEK293T cells. RNA from HEK293T cells was isolated with the RNeasy kit (Qiagen) and converted to complementary DNA (cDNA) by reverse transcription using the M-MLV RT (Promega) with an oligo (dT) primer. The BRD2 encoding cDNA was then amplified by two independent polymerase chain reactions (PCRs): (i) the first with the hF1 (position 1–22) and hR1 (position 1217–1196) primers, and (ii) the second with hF2 (position 1144–1162) and hR2 (position 2406–2379) primers (Table S3). PCR amplifications were performed with 35 amplification cycles (30 s at 95 °C, 30 s at 60 °C and 105 s at 72 °C) using the Pfu DNA polymerase (Promega). The two overlapping PCR fragments were then denatured (10 min at 95 °C) and hybridized at room temperature for 30 min followed by an elongation step of 15 min at 72 °C using the Pfu polymerase to generate the full-length cDNA of the BRD2 gene. The BRD2 cDNA was then inserted into the mammalian expression vector pCMV-Myc (Invitrogen) according to the Gateway technology (Invitrogen). This plasmid (named pCMV-Myc-BRD2) contains the hCMV enhancer-promoter and an N-terminal Myc tag allowing

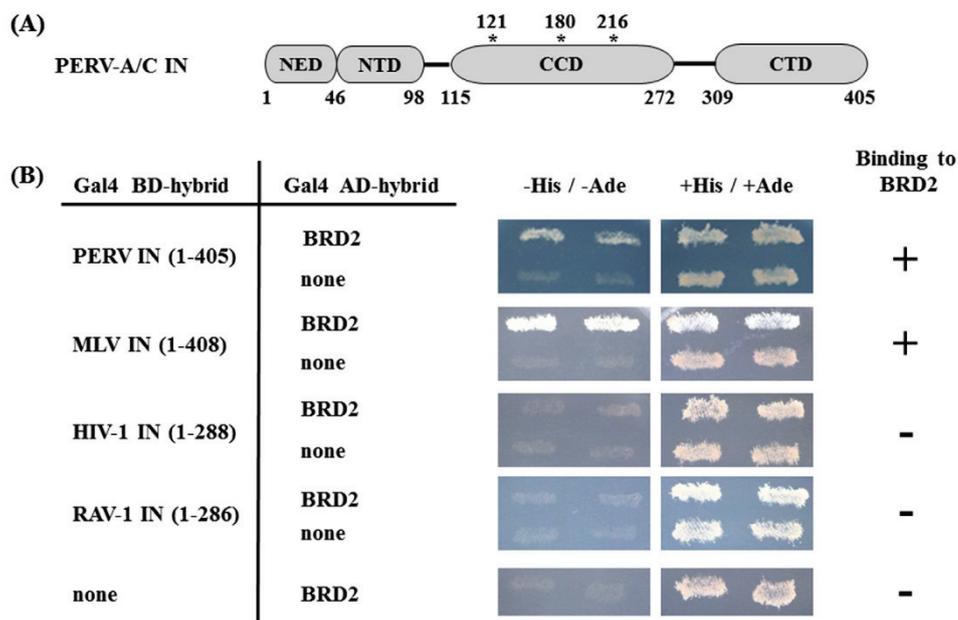


Fig. 1. Interaction of PERV IN with the BRD2 protein. **(A)** Domain organization of the full-length gamma-retroviral PERV IN protein. The IN proteins are made up of three domains: (i) the N-terminal domain (NTD) which is involved in the binding of the viral DNA and zinc, (ii) the catalytic core domain (CCD) which contains the invariant catalytic triad D,D-35-E (indicated with asterisks), (iii) the C-terminal domain (CTD) involved in non-specific DNA binding. The PERV also contains an N-terminal extension domain (NED). **(B)** In the two-hybrid assay, INs from PERV and MLV gamma-retroviruses interact with BRD2, while INs from HIV-1, a lentivirus, and from RAV-1, an alpha-retrovirus, do not. The full-length ORFs of PERV, MLV, HIV-1 and RAV-1 INs were fused to the Gal4-DNA binding domain (Gal4-BD). The BRD2 ORF was fused to the Gal4-activation domain (GAL4-AD). The yeast reporter strain Y2H-Gold producing the hybrid proteins indicated was analyzed for histidine and adenine auxotrophy (-His/-Ade) acquired by the interaction of the fused proteins.

the expression of the Myc-BRD2 protein. Sequencing was performed to control for the integrity of the cDNA.

Construction of the pPERV IN CTD – BRD2 (667–808) complex expression plasmid. An original plasmid construction was built to obtain the protein complex formed by PERV IN CTD (aa 309–405) and BRD2 (aa 667–808), cloned from the pCI-Flag-PERV IN and the pCMV-Myc-BRD2 plasmids respectively. This construction contains three genes encoding: i) TEV protease, ii) PERV IN CTD with an N-terminal Strep-tag II and iii) BRD2 (667–808) with an N-terminal hexahistidine (His) tag (see Fig. 6A). At first, each gene was amplified by PCR using the pRK793b, pGEM-T Easy and pPROEX-Hta plasmids, respectively. The whole insert containing these three genes was generated by PCR. In addition, the genes were separated from each other by TEV protease cleavage sites introduced by the primers. The insert was cloned using an In-Fusion[®] HD-Cloning kit (Clontech) and amplified by PCR in the pET28a vector.

The constructed expression plasmid was also used to generate one single and one double alanine mutant affecting the complexed IN CTD (i.e. W379A and R391A/L392A), and to generate the shorter pPERV IN CTD – BRD2 (667–735) complex expression plasmid also termed pPERV IN CTD – BRD2 ET. All plasmids were confirmed by DNA sequencing.

Construction of the pCMV-Myc-BRD2-, BRD3-, BRD4-Cter plasmids. The pCMV-Myc-BRD2 plasmid was used for the amplification of the BRD2 C-terminal fragments (BRD2-Cter) by PCR using hF3 (position 1789–1807) and hR3 (position 2406–2389) primers (Table S3). The BRD3 C-terminal fragment (BRD3-Cter) was amplified from pGVB3403 using the hF4 (position 1705–1722) and hR4 (position 2181–2161) primers, and the BRD4 C-terminal fragment (BRD4-Cter) was amplified from pGVB3115 using hF5 (position 1819–1836) and hR5 (position hF5-hR5) primers (Table S3). PCRs were performed with 35 amplification cycles (30 s at 95 °C, 45 s 65 °C and 160 s at 72 °C) using the Pfu

DNA polymerase (Promega). The resulting fragment was inserted into the pCMV-Myc vector using the Gateway technology as above and controlled by sequencing.

Construction of the pCI-Flag-PERV IN plasmids. The PERV IN coding sequence was amplified from plasmid A14/220 containing the whole genome of PERV (Genbank ref. AY570980) using attB11SP/PERV and attB21ASP/PERV IN primers (Table S3). PCRs were performed with 35 amplification cycles (30 s at 95 °C, 30 s 60 °C and 120 s at 72 °C) using the Pfu DNA polymerase (Promega). The resulting fragment was inserted into the pDONR223 and pCI-Flag plasmids using the Gateway technology as above.

The pCI-Flag-PERV IN CTD was obtained by PCR amplification of the CTD of PERV IN (aa 273–405) from pCI-Flag PERV IN by PCR amplification using F and R primers. The resulting fragment was inserted into the pCI-Flag plasmid using the Gateway technology. The pCI-Flag PERV IN CTD (P387A/L388A) plasmid containing a CTD mutated was obtained by PCR-directed mutagenesis with the appropriate primers.

2.2. The two-hybrid assay

The yeast reporter strain Y2H-Gold containing the HIS3 and ADE2 Gal4-inducible genes were co-transformed and plated on a selective medium lacking tryptophan and leucine. Y2H-Gold double transformants were patched on the same media and then analyzed for histidine and adenine auxotrophy by replica plating on selective medium lacking tryptophan, leucine, histidine and adenine (-His/-Ade).

2.3. Cells and immunoassays

Cells. Human embryonic kidney HEK293T cells were cultured in

Dulbecco's modified eagle medium (PAA), 10% fetal bovine serum (PAA) and 1% penicillin-streptomycin (Gibco) at 37 °C in 5% CO₂.

Confocal immunofluorescence microscopy. For immunofluorescence microscopy, HEK293T cells were grown in LabTek II chamber slides (BD Falcon) for 24 h and transfected with 800 ng of plasmid using Lipofectamine 2000 (Invitrogen). Forty-eight hours later, the transfected HEK293T cells were fixed with 1% paraformaldehyde in phosphate-buffered saline (PBS) overnight at 4 °C, washed with PBS and then incubated in a blocking/permeabilization buffer (BP) (10% FBS and 0.1% saponin in PBS) for 10 min at 4 °C. After incubation, the cells were washed twice with the BP buffer and incubated with the primary antibodies. The Myc-BRD2 fusion protein and the Flag-PERV IN fusion protein were detected using a mix of polyclonal rabbit antibody directed against the Myc tag (Sigma-Aldrich, catalog no. C3956, dilution 1/500) and a monoclonal mouse antibody directed against the Flag tag (Sigma-Aldrich, catalog no. F3165; dilution 1/500) in the BP buffer for 1 h at 4 °C. After incubation, the cells were washed twice with the BP buffer and incubated with the alexa-488 conjugated goat anti-rabbit antibody (Invitrogen, catalog no. A11008, dilution 1/500) or the alexa-633 goat anti-mouse antibody (Invitrogen, catalog no. A21053; dilution 1/500), respectively. The cells were then mounted in the Vectashield Mounting Medium (vector Laboratories) containing 1.5 µg/ml of 4',6-diamidino-2-phenylindole (DAPI) to stain the nuclear DNA. Images were acquired with a Confocal Spectral SP5 (Leica) controlled by LAS AF software (Leica Microsystems CMS GmbH) with a 63x oil immersion lens.

Co-immuno Precipitation (Co-IP) assays. HEK293T cells were plated at a confluence of 60–70% in 6-well dishes. The following day, cells were co-transfected with 2 µg of each of two plasmids: (i) the pCI-Flag, the pCI-Flag-PERV IN CTD or the pCI-Flag-PERV IN CTD (P387A/L388A) plasmids and (ii) the pCMV-Myc-BRD2-Cter, pCMV-Myc-BRD3-Cter or pCMV-Myc-BRD4-Cter encoding plasmids. Transfection was carried out using the Lipofectamine 2000 transfection reagent (Invitrogen). Forty-eight hours later, transfected cells were trypsinized, pelleted and resuspended in the IP buffer (50 mM Tris, pH 8, 150 mM NaCl, 1% Nonidet-P40, 0.5% sodium deoxycholate) supplemented with a complete EDTA-free protease inhibitor cocktail (Roche). The resuspended cells were then disrupted by sonication and incubated on ice for 30 min before being centrifuged for 10 min at 4 °C at 12,000 g. The protein concentration of the supernatant (cell lysate) was determined by the Bradford assay. For the IP reaction, 20 µl of pure proteome Protein G Magnetic Beads (Millipore) were washed and incubated for 15 min in the IP buffer with 4 µg of a polyclonal rabbit anti-Myc (Sigma-Aldrich, catalog no. C3956) on a rotator at room temperature. The magnetic Myc-beads were then washed three times with 500 µl of the IP buffer before the addition of 500 µg of the cell lysate to 600 µl of the magnetic Myc-beads in the IP buffer. The mixture was then incubated overnight at 4 °C on a rotator. The next day, the magnetic beads were washed three times with 500 µl of IP buffer and then boiled in 60 µl of Laemmli buffer containing 5% β-mercaptoethanol. The protein mixture was analyzed by SDS-PAGE and immuno-detected by Western blotting.

Immunodetection. For immunoblotting, 20 µg of the whole cell extracts or 20 µl of the immuno-precipitated proteins were separated on a 10% SDS-PAGE. The proteins were transferred to a nitrocellulose membrane and detected either (i) with a mouse anti-Myc antibody (Sigma-Aldrich, catalog no. M4439, dilution 1/5000) and an anti-mouse IgG-Horseradish Peroxidase antibody (Sigma-Aldrich, catalog no. A5906, dilution 1/10,000) or (ii) with an anti-Flag M2 Horseradish Peroxidase (HRP) antibody (Sigma-Aldrich, catalog no. A8592, dilution 1/10,000). The Western blots were developed using the ECL chemiluminescent HRP substrate (Thermo Scientific).

2.4. Protein purification

Purification of the PERV IN CTD – BRD2 ET complex. After transformation, the protein complex was expressed in T7 Express

competent *E. coli* in LB broth by induction with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37 °C for 3 h. Bacteria were harvested by centrifugation at 6,000 g for 10 min and stored at –20 °C. The thawed bacterial pellet was resuspended in buffer A (20 mM Tris HCl, pH 8.0, 0.5 M NaCl, 25 mM imidazole) in the presence of lysozyme (0.2 mg/ml, Sigma) and complete ULTRA EDTA-free protease inhibitor cocktail 1X (Roche). Following sonication for 8 min at 40 mA on ice, the lysate was clarified by centrifugation at 14,000 g for 45 min and filtered through a 0.45 µm filter (VWR International). The supernatant containing soluble His-tagged BRD2 ET, Strep-tagged II PERV IN CTD and the resulting double-tagged complex was loaded on a Ni²⁺ charged 5 ml Protino[®] Ni-NTA (Macherey Nagel) equilibrated with buffer A using the ÄKTA purifier system (GE Healthcare). The column was extensively washed with buffer B (20 mM Tris HCl, pH 8.0, 1 M NaCl, 25 mM imidazole). The His-tagged BRD2 ET and the double-tagged complex were eluted by a linear gradient of imidazole (0.025–1 M). Eluted fractions were diluted 1:4 with buffer C (20 mM Tris HCl, pH 8, 0.5 M NaCl) and immediately loaded on a 5 ml Strep-Tactin affinity column (Strep-Tactin Superflow[®] cartridge, IBA) equilibrated with buffer W (100 mM Tris HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA). The double-tagged complex was eluted with a desthiobiotin-containing buffer (100 mM Tris HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin). After 15% SDS-PAGE analysis, purified protein complex aliquots were flash frozen in liquid N₂ and stored at –80 °C. Each partner of the complex was identified by western blot and MALDI-TOF (Voyager DE-PRO, AB Sciex).

Purification of the PERV IN CTD – BRD2 (667–806) complex. A similar protocol was used to purify the complex and the two related mutants (i.e. W379A, and R391A/L392A mutated on IN).

2.5. Structural studies

Determination of the oligomeric state of the complex. PERV IN CTD – BRD2 ET was concentrated at 14 mg/ml using a Vivaspin-PES-10K centrifugal concentrator (Sartorius) and loaded on Superdex 200 10/300 GL (GE Healthcare) equilibrated with buffer C for analysis by Size Exclusion Chromatography (SEC). The main peak fraction was concentrated to 10 mg/ml and analyzed by Dynamic Light Scattering (DLS) with a Zetasizer Nano S ZEN 1600 instrument (Malvern Instruments). Each measure is the mean of 10 runs repeated three times. A cross-linking experiment with glutaraldehyde was performed by mixing 10 µg of purified protein complex, 20 µl of 20 mM Tris HCl, pH 8 and 5 µl of 0.1% glutaraldehyde. After 20 min of incubation at room temperature, the cross-linking reaction was stopped by adding 5 µl of 200 mM glycine. The resulting protein sample was loaded onto a 15% SDS PAGE. A similar protocol was used for PERV IN CTD – BRD2 (667–808).

SAXS study of the complex. SAXS data were collected at the European Synchrotron Radiation Facility (ESRF) beamline BM29, Grenoble, France. Sample and buffer were centrifuged at 14,000 g for 10 min before acquisition to remove aggregates. The complex, at a concentration of 15 mg/ml, was injected in SEC coupled mode using an analytical Superdex S200 increase 5/150 GL column (GE Healthcare) equilibrated in buffer C. The temperature was maintained at 4 °C. Processing and analysis of SEC-SAXS data were performed with the programs CHROMIXS and SAS-DATA of the ATSAS suite (Franke et al., 2017). The radius of gyration (R_g) was calculated from both the Guinier plot and the pair-distribution function $P(r)$. The maximum inter-atomic distance of particle (D_{max}) was calculated from $P(r)$ analysis. *Ab initio* bead models were generated using DAMMIN. After superposition and averaging, the filtered SAXS envelope was displayed with CHIMERA (Pettersen et al., 2004).

A sequence search performed with SWISS-MODEL (Waterhouse et al., 2018) showed that the C-terminal fragment 736–808 of BRD2 has no homologous sequence of known structure in the Protein Data Bank (PDB). In consequence, the generation of a SAXS-restrained atomic model was possible for PERV IN CTD – BRD2 ET (667–735) only.

First, the NMR structure of MLV IN EBM in complex with BRD4 ET (PDB code 2n3k) was taken as a guide to model with the program SCULPTOR (Bunkoczi and Read, 2011) the structure of PERV IN EBM (aa 378–394) in complex with BRD2 ET. This model of the bound C-terminal tail was linked with an arbitrary relative position and orientation with the structure of the remaining CTD (aa 309–377) extracted from our model of the PERV intasome (Demange et al., 2015). The resulting homology model was fitted against experimental SAXS data with the program DADIMODO (Evrard et al., 2011) using a two domain rigid-body strategy with the loop 376–380 connecting the IN CTD core to the bound EBM segment rendered as flexible. The goodness of fit was estimated by χ^2 calculation with CRYSOLO (Franko et al., 2017). The resulting SAXS-restrained model was positioned with CHIMERA into the SAXS envelope for observation.

3. Results

3.1. The human BRD2 protein interacts specifically with PERV and MLV INs

We first analyzed the interactions between different retroviral INs and BRD2 in the yeast two-hybrid assay (Fig. 1B). We tested the INs of the gamma-retroviruses PERV and MLV, the lentivirus HIV-1 and the alpha-retrovirus Rous Associated Virus type 1 (RAV-1) (Fig. 1B, left table). The full-length PERV IN (aa 1–405), MLV IN (aa 1–408), HIV-1 IN (aa 1–288), and RAV-1 IN (aa 1–286) were fused to the Gal4-DNA binding domain (Gal4-BD). The full length ORF of BRD2 (836 residues long) was fused to the Gal4 activation domain (Gal4-AD). The yeast strain was analyzed for histidine and adenine auxotrophy (-His/-Ade) acquired by the interaction of the fused proteins (Fig. 1B, left table). As expected, no growth was observed if PERV and MLV INs or BRD2 were not fused to their respective Gal4-BD and -AD genes. In accordance with others (Gupta et al., 2013), we detected no interaction between RAV-1 or HIV-1 INs and BRD2. We confirmed that MLV IN interacts with BRD2 and showed that PERV IN gave a positive interaction (Fig. 1B).

3.2. PERV IN and BRD2 co-localize in the nucleus of cells

To further investigate the interaction between BRD2 and PERV IN, we used immunofluorescent microscopy to examine the localization of the full-length Flag-PERV IN and the Myc-BRD2 proteins in HEK 293T cells. In the first experiment, the pCMV-Flag-PERV IN and the pCMV-Myc-BRD2 plasmids were individually transfected into HEK 293T cells. The correct expression of each protein was checked by western blot analysis (data not shown). The cellular distribution of each fusion protein was analyzed using confocal fluorescent microscopy 48 h later (Fig. 2). More than 30 positive cells were observed in each sample. In HEK 293T cells, exogenously expressed Flag-PERV IN was mostly detected in the nucleus (either diffused (top images) or as punctuate nuclear foci (bottom images)) with a minority expression in cytoplasm (Fig. 2A). In HEK 293T cells, exogenously expressed Myc-BRD2 showed diffuse nuclear localization (main pattern) or both a nuclear and cytoplasmic localization (Fig. 2B).

In the second experiment, the pCMV-Flag-PERV IN and the pCMV-Myc-BRD2 plasmids were co-transfected into HEK 293T cells. Several patterns of co-localization (in yellow) were observed (Fig. 2C): (i) at the internal periphery of the nucleus (top images) and (ii) diffuse and punctuate in the nucleus (middle images). We also observed a chromosomal co-localization in a dividing cell (bottom images). Although a χ^2 statistical test did not reveal any change in the localization of the proteins — whether they were expressed alone or together (data not shown) — these co-localization studies provide further evidence supporting an interaction between PERV IN and BRD2 in cytoplasm and nucleus of cells.

3.3. The BRD2 binding site maps within the C-terminal tail of PERV IN CTD

The retroviral IN protein contains three evolutionarily conserved and functional domains interconnected with flexible linkers: the N-terminal domain (NTD), the catalytic core domain (CCD) and the C-terminal DNA binding domain (CTD) (Fig. 1A), all necessary for the biological activities of the IN protein. The predicted structure of PERV IN also comprises a singular 46-residue long N-terminal extension domain (NED) (Demange et al., 2015). This NED structure is also observed in MLV and prototype foamy virus (PFV) IN.

In order to map the region of PERV IN involved in the interaction with BRD2, we constructed a total of five IN truncation mutants fused with Gal4-DB: i.e. the one-domain fragments NTD (aa 1–98), CCD (aa 99–272) and CTD (aa 273–405) and the two-domain fragments NTD + CCD (aa 1–272) and CCD + CTD (aa 99–405) (see Fig. 1A). These five fragments were tested in the yeast two-hybrid system against the Gal4-AD-BRD2 protein (Fig. 3A). BRD2 interacts with PERV IN (99–405) and IN (273–405) but not with IN (1–98), IN (99–272) and IN (1–272). Therefore, only constructs containing PERV IN CTD can bind efficiently to BRD2 and the NTD and CCD are dispensable. Thus, our data showed that the CTD alone (aa 273–405) is necessary and sufficient for the interaction with BRD2. Similar deletion mutants were constructed with MLV IN and tested using the yeast two-hybrid system against Gal4-AD-BRD2 (Supplementary Fig. S1). In accordance with data from other studies (De Rijck et al., 2013; Gupta et al., 2013; Sharma et al., 2013), we observe that the CTD of MLV IN is sufficient for the interaction with BRD2. Altogether, these results demonstrate that INs from gamma-retroviruses interact specifically with BRD2 through their CTD domains.

To characterize more precisely the peptide sequence of IN required for the interaction, we used two CCD + CTD constructs deleted from the C-terminal sequences of the CTD, PERV IN (99–394) and IN (99–372) (Fig. 3B). The truncation of the 11 C-terminal residues of the CTD (PERV IN (99–394)) did not disturb the interaction. By contrast, removal of the 33 C-terminal residues of the CTD in PERV IN (99–372) resulted in a loss of interaction. This deletion study indicated that several residues from the segment 372–394 are necessary to mediate the interaction with BRD2 (Fig. 3B). We also tested positively the short IN fragment 372–405 carrying this interaction sequence (Fig. 3B) showing that the C-terminal tail of PERV IN (residues 372–405), alone, is sufficient for interaction with BRD2.

We then focused on the most C-terminal motif of CTD encompassing residues 378–394 in PERV IN. This segment corresponds to the MLV IN EBM (aa 389–405) which is critical for the interaction with BRD4 (Crowe et al., 2016). The primary sequence of this region is highly conserved among the gamma-retroviral INs (Fig. 4A) and not between retroviral genera (Supplementary Fig. S2). To assess the importance of these conserved residues, we performed a two-hybrid assay using five different mutants: the W379A single mutant (mutant containing an Alanine in place of the W379) and the P387A/L388A, K389A/L390A, R391A/L392A and H393A/R394A double mutants (Fig. 4B left panel). All of these mutants affect the interaction between PERV IN and BRD2 (Fig. 4B right panel). Notably the W379A single mutant disrupts the interaction in accordance with what was previously shown for the equivalent MLV W390A IN mutant and BRD4 (El Ashkar et al., 2014). This means that not only W379 but at least one residue in each double mutant (387–388, 389–390, 391–392 and 393–394) is important for the formation of the PERV IN – BRD2 complex.

PERV IN interacts with the ET domain of different BRD proteins in the two-hybrid assay and in HEK293T cells.

BET proteins such as BRD2 (Fig. 5A), BRD3 and BRD4 contain two conserved tandem bromodomains and a conserved ET domain. The ET from BRD4 is known to be involved in the interaction with MLV IN (Crowe et al., 2016; Gupta et al., 2013). We first tested the interaction of PERV IN with ET from different BET proteins using the two-hybrid system (Fig. 5B). A positive signal was detected with all constructs. The

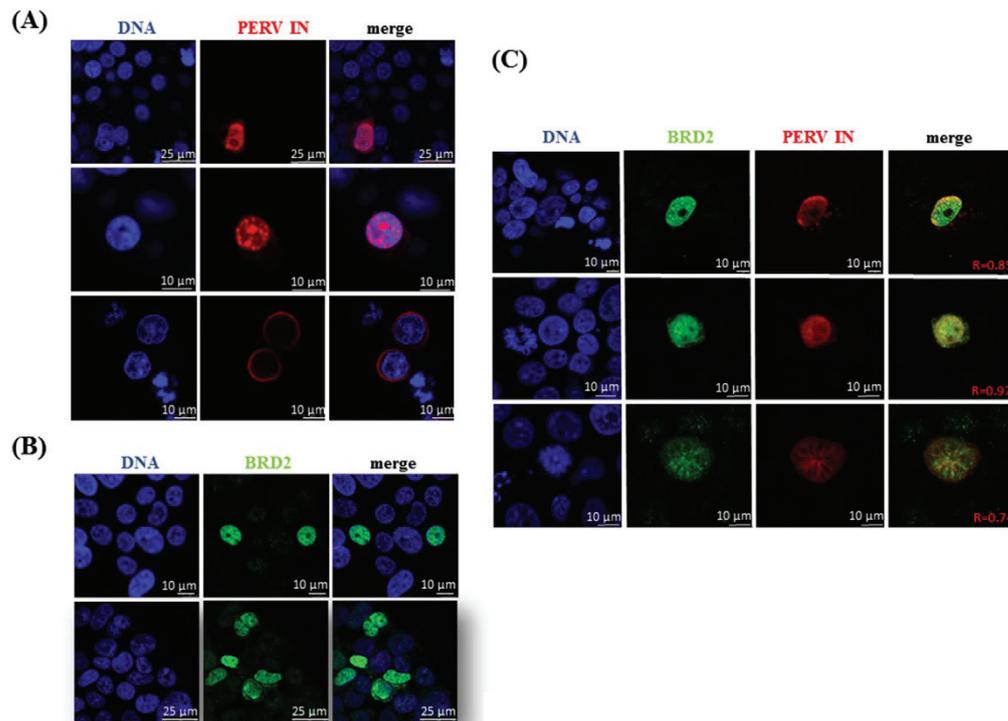


Fig. 2. Intracellular co-localization of the full-length Myc-BRD2 and Flag-IN PERV proteins in HEK 293T cells. HEK 293T cells were transfected with the pCMV-Flag-PERV IN plasmid or/and with the pCMV-Myc-BRD2 plasmids. 48 h after transfection, cells were fixed and confocal laser scanning microscopy was used to study expression of the fluorescent fusion proteins. Cells were also colored with DAPI to show DNA. In (A), cells were transfected with the pCMV-Flag-PERV IN plasmid. The DNA (in blue) and the Flag (in red) signals and their overlays are shown in the left, middle, and right images of each lane, respectively. Flag-PERV IN was localized to both the nucleus and—to a lesser extent—the cytoplasm. In (B), cells were transfected with the pCMV-Myc-BRD2 plasmid. Myc (in green) and DNA (in blue) signals and their overlays are shown in the top left, bottom left and right-hand images of the panel, respectively. Expression of Myc-BRD2 is mainly nuclear (top and bottom images) or both nuclear and cytoplasmic (bottom image). In (C), 293T cells were co-transfected with both the pCMV-Flag-PERV IN and the pCMV-Myc-BRD2 plasmids. The Flag-PERV IN (in red), Myc-BRD2 (in green) and DNA (in blue) signals are shown on the left of the images. A Pearson coefficient (in brackets) greater than 0.75 indicates a good co-localization of the two proteins.

BRD4 ET fragment alone (aa 607–671), was sufficient for this interaction whereas the use of smaller fragments of ET of BRD2 or BRD4 did not show any interaction (data not shown).

We also investigated the interaction of the PERV IN CTD and the hBET C-terminal domain in HEK 293T cells. PERV IN CTD and hBET ET-containing C-terminal region (BRD2-Cter, BRD3-Cter and BRD4-Cter) were tagged with the Flag and Myc tags, respectively. The Flag-PERV IN CTD was tested for its ability to co-immunoprecipitate the Myc-hBET-Cter after transfection in HEK 293T cells (Fig. 5C). As a control, we used a PERV IN CTD (P387A/L388A) mutant containing two alanines in place of P387 and L388 in the interacting domain of IN. The Flag-PERV IN-CTD and the Flag-PERV IN-CTD (P387A/L388A) co-immunoprecipitation with Myc-hBET-Cter were assessed in the Myc-immunoprecipitated fractions using a Flag antibody (Fig. 5C). Data showed that the Flag-PERV IN-CTD was co-immunoprecipitated with Myc-BRD2-Cter, Myc-BRD3-Cter and Myc-BRD4-Cter (panel D, lanes 2, 5 and 8). The mutation of two residues in the Flag-PERV IN-CTD (P387A/L388A) diminished but did not fully abolish the interaction with the Myc-BRD2-Cter and Myc-BRD4-Cter (panel D, lanes 3 and 9, respectively). By contrast, the interaction of Myc-BRD3-Cter with the Flag-PERV IN-CTD (P387A/L388A) was completely lost (panel D, lane 6).

3.4. Co-purification of the two proteins as a complex

With the aim of strengthening our data and obtaining novel structural information about the IN/Brd2 complex, we overexpressed and purified recombinant PERV IN CTD and the BRD2 C-terminal fragment from a prokaryotic expression system and performed *in vitro* interaction assays.

To date, structural knowledge on the C-terminal domain of BET proteins is limited to the ET (Crowe et al., 2016; Sanchez and Zhou, 2009). A prediction of intrinsic disorder by the GLOBPLOT server (Linding et al., 2003) shows that the C-terminal region of BRD2 (aa 667–836) that contains the ET (aa 676–738) is mostly ordered with the exception of the C-terminal extremity which is rich in serine residues (Supplementary Fig. S3). Since this long flexible tail can prevent crystallization of the protein and—in doing so—the use of X-ray crystallography to determine the atomic structure of the complex, we worked on the BRD2 fragment without the C-terminal tail. We produced independently PERV IN CTD (aa 309–405) and the BRD2 fragment (aa 667–808) to form the complex, but an aggregation phenomenon was observed when the two partners were mixed together (data not shown). To overcome this problem, we used a new strategy of protein co-expression inspired by Levy and co-workers that have already expressed HIV-1 IN and LEDGF/p75 recombinant proteins from a

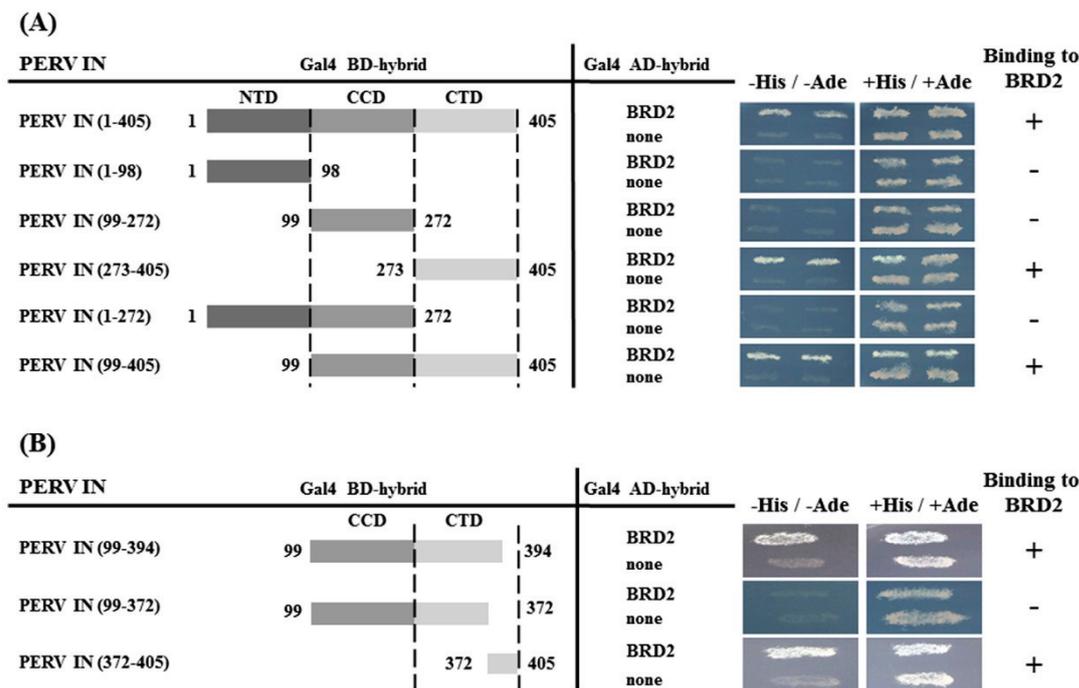


Fig. 3. The CTD domain of PERV IN (and more specifically, residues 372–405) interacts with BRD2 in the two-hybrid assay. **(A)** The PERV IN CTD is involved in the interaction with BRD2 protein. The NTD (aa 1–98), CCD (aa 99–272) or CTD (aa 273–405) isolated domains of PERV IN as well as combined domains of PERV IN (NTD + CTD (aa 1–272) and CCD + CTD (aa 99–405)) were fused to the Gal4-BD domain. Interaction with BRD2 was evaluated as described in Fig. 1. **(B)** Refinement of the PERV IN CTD interaction domain. PERV IN fragments 99–394, 99–372 and 372–405 were fused to the Gal4-BD domain. Interaction with BRD2 was evaluated as described in Fig. 1B.

sophisticated polycistronic construction in a mammalian cells expression system (Levy et al., 2016). We optimized the same kind of construction for a bacterial expression system (Fig. 6A). A Strep-tagged-PERV IN fragment (aa 309–405) and a His-tagged-BRD2 fragment (aa 667–808) were cloned in frame with Tobacco Etch Virus (TEV) protease and interspaced with TEV digestion sites to form a unique, cleavable, polyprotein. During overexpression, the TEV protease separates the Strep-PERV IN CTD and the His-BRD2 (667–808) of the polyprotein leading to formation of the complex. The complex was then purified by two successive affinity steps with nickel and Strep-Tactin column (Fig. 6B top left panel). The two partners were identified by western blot using anti-His tag and Anti-Strep tag antibodies (Fig. 6B bottom left panels) and mass spectrometry from gel bands (data not shown). When using a wild type Strep-PERV IN CTD and the His-BRD2 (667–808), we were able to co-purify both proteins in the elution of the nickel column (lane 2) and also in the elution sample of the strep column (lane 3). This indicated that both partners, His-BRD2 (667–808) and Strep-PERV IN CTD were indeed co-purified as a complex form and not as two isolated proteins.

In order to assess the interaction of the previously described IN mutants with BRD2, the PERV IN CTD was modified within this construction to obtain the PERV IN CTD- W379A and PERV IN CTD-R391A/L392A mutants. Using these constructions, both proteins were efficiently co-expressed (Fig. 6B, lanes 4 and 7). However, after passing on the nickel column, although a small proportion of the strep-PERV IN CTD mutants were retained (lanes 5 and 8), the His-BRD2 (667–808) fragment was not retained on the Strep column (lanes 6 and 9). This observation strongly supports the interaction of BRD2 with PERV IN CTD and assesses the critical role of each conserved residue of the IN

EBM region.

We also performed the co-purification of the shorter PERV IN CTD – BRD2 complex, hereafter referred as to PERV IN CTD – BRD2 ET, by using the TEV expression system. We confirmed the interaction of this new two-fragment complex *in vitro*, as described below.

3.5. The complex is heterodimeric in solution and assembles as distinct modules

We next investigated the oligomeric state of the complex by Size Exclusion Chromatography (SEC) and cross-linking approaches.

In more details for PERV IN CTD – BRD2 ET, the SEC experiment indicated one main peak at an elution volume of 16.5 ml and a minor peak at an elution volume of 17.5 ml, corresponding to ~23 kDa (heterodimer) and ~13 kDa (IN CTD), respectively (Supplementary Figs. S4A and S4B). This demonstrated that the two-fragment complex is mainly heterodimeric at 14 mg/ml in 20 mM Tris HCl pH 8.0, 500 mM NaCl. Fractions from the main peak were pooled, concentrated and sample homogeneity was assessed by diffuse light scattering, resulting in a satisfactory polydispersity index of 0.2.

The heterodimeric state of PERV IN CTD – BRD2 ET was further confirmed by a chemically cross-linking assay using glutaraldehyde (Supp. Fig. 4C). Indeed, in the presence of glutaraldehyde, the covalently bound PERV IN CTD and BRD2 ET migrate with a higher molecular weight of about 23 kDa (shown on lane 1). The heterodimeric state of the cross-linked sample was confirmed by mass spectrometry from gel bands (data not shown). The heterodimeric state of the cross-linked PERV IN CTD – BRD2 (667–808) was checked with a similar protocol (data not shown).

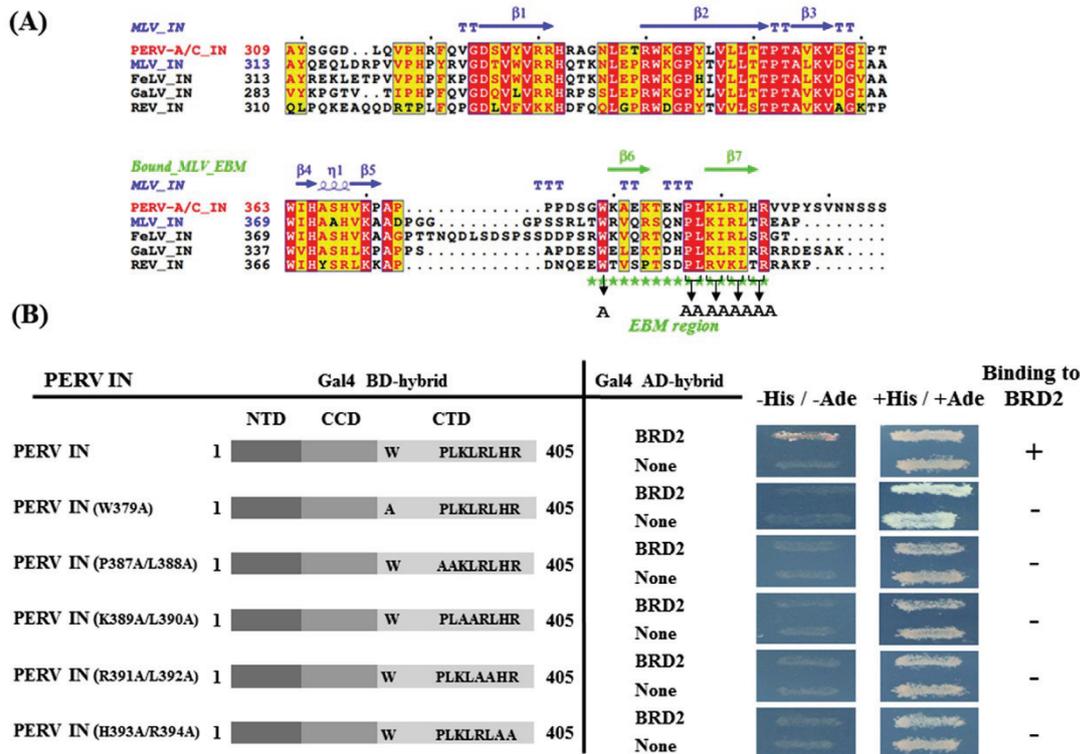


Fig. 4. Alanine replacements of W379 or amino acid pairs from residues 387–394 of PERV IN abolish interaction with BRD2. (A) Multiple sequence alignment of the CTD of gamma-retroviral INs from PERV, MLV, FeLV, Gibbon ape Leukemia Virus (GalV) and REV. Identical and well-conserved residues (> 70% similarity) have a red and a yellow background, respectively. The 17-residue long EBM segment is marked below sequence blocks with green stars, and the residues in PERV IN we have mutated into Alanine (A) are written. Secondary structure elements are shown above sequence blocks and colored in blue for the NMR structure of MLV IN CTD (PDB code 2m9u) and colored in green for the NMR structure of MLV IN EBM in complex with BRD4 ET (PDB code 2n3k). Secondary structure elements are shown above sequence blocks (helices with squiggles, strands with arrows, turns with T letters). The figure was rendered with ESPript. (B) Alanine scanning of W379 (single mutation) and of residues in the well-conserved stretch 387–394 (double mutations). The ability of the PERV IN mutants to interact with BRD2 was analyzed in the two-hybrid assay as described in Fig. 1B.

Crystallization trials were performed on the two purified complexes but without success. The two samples were analyzed by SAXS using a synchrotron radiation source and the collected HPLC-SAXS data were processed (Table 1 and Fig. 7 for PERV-A/C IN CTD - BRD2 ET). The calculated Kratky plots yielded bell-shaped curves at low q values but did not converge at high q (Fig. 7B), indicating partial flexibility which might have hindered our crystallization attempts. The linearity of the Guinier plots were indicative of monodisperse solutions and similar R_g values were obtained from Guinier and $P(r)$ analyses (Table 1 and Fig. 7C). *Ab initio* modeling led to elongated shape envelopes with an estimated resolution of 3 nm (Fig. 7D).

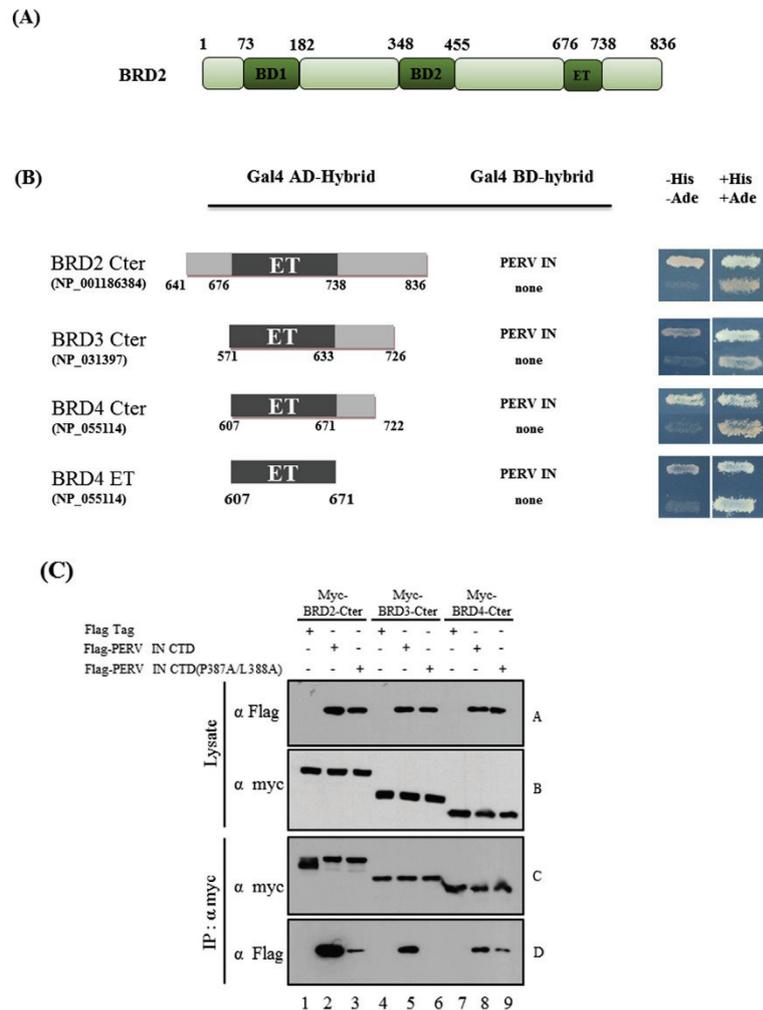
PERV IN CTD and BRD2 ET share significant sequence identities with MLV IN CTD and BRD4 ET (60% and 86%, respectively) whose experimental structures are known (Aiyer et al., 2014; Crowe et al., 2016). An atomic model of PERV IN CTD - BRD2 ET was generated by homology modeling, assuming that the bound peptide tail of the CTD (aa 378–394) adopts a β -hairpin fold like the bound MLV IN EBM (Crowe et al., 2016). This model was refined against related SAXS data using a two-domain rigid body strategy with a flexible CTD loop (aa 373–377) in order to experimentally position the SH3 core domain of the IN CTD (aa 309–372) relative to the bound EBM segment. The resulting SAXS-restrained model had an excellent dependence to the experimental data with a χ^2 of 1 (Fig. 7A), and it exhibited two distinct

structural blocks connected by the five-residue loop 373–377 (Fig. 7D). Such modular structure gives flexibility to the heterodimer and further explains why the EBM segment is sufficient and essential for IN - BET assembly.

4. Discussion

Herein, we show by using two-hybrid experiments (Figs. 1, 3–5), co-localization into cells (Fig. 2), co-purification and co-immunoprecipitation (Fig. 6), and SAXS (Fig. 7) that the gamma-retroviral PERV A/C IN interacts with the hBET proteins BRD2, BRD3 and BRD4 as does MLV IN. The interaction with hBET proteins seems specific to gamma-retroviral INs since BRD2 does not interact with either RAV-1 (alpha retrovirus) or HIV-1 (lentivirus) INs (Fig. 1). Altogether, these results show that interaction between BET proteins and INs is a common feature of gamma-retroviruses and is specific to this retroviral genus. This specificity may explain the similarity of integration site distribution between PERV and MLV (Moalic et al., 2006, 2009), and the difference with other retroviruses from other genera.

The CTD of PERV IN is responsible for the interaction with hBET proteins *in vitro* and *in cellulo* (Figs. 3–7). We also observed by two-hybrid and *in vitro* interaction assays that mutations in a 17-residue long peptide (aa 378–394), which pertains to the PERV IN CTD C-



terminal peptide tail, abolish the interaction with BRD2. This segment is equivalent in sequence position with the BET-interaction motif EBM characterized in MLV IN (Crowe et al., 2016; Larue et al., 2014) (Fig. 4A), thereby confirming that gamma-retroviral INs share a common mechanism of selection of BET proteins. We also observed that the W379 PERV IN residue in the CCD is essential for interaction with BRD2 in accordance with other authors that established that the corresponding residue W390 from the MLV IN CTD is a critical hotspot for interaction with the BRD4 protein (El Ashkar et al., 2014). We also showed that, besides W379, at least one residue in each double mutant of the PERV IN EBM 387–394 stretch is also essential for the interaction. Others reported that residues (E266, L268 and Y269 of MLV IN CCD) — located outside the MLV EBM motif — are also important for the interaction with BET proteins (Gupta et al., 2013), but substitution of these residues in the context of the full-length MLV IN did not abolish the interaction with Brd4 (El Ashkar et al., 2014). Finally, PERV IN CTD was shown to interact with the conserved BET ET domain (Fig. 5) as does MLV IN (Aiyer et al., 2014; Crowe et al., 2016).

The interaction with BET protein was shown to be important for the nuclear localization of MLV IN. Indeed, when expressed separately,

MLV IN and BET proteins show a diffuse cytoplasmic and nuclear localization, respectively. When co-expressed with BET proteins, IN MLV changes localization from cytoplasmic to nuclear (Gupta et al., 2013). By contrast, we found that the exogenously expressed full-length Flag-PERV IN and Myc-BRD2 proteins were detected mainly in the nucleus of 293T cells when expressed alone, while only a slight diffuse cytoplasmic localization was observed for both proteins. We did not observe any change in the localization of the Flag-PERV IN when co-expressed with the Myc-BRD2 protein. However, we clearly detected a co-localization of the two protein domains in the nucleus and cytoplasm of cells when expressed together, which confirms the interaction of both proteins *in cellulo* and strengthens the role of the BET protein to direct PERV IN integration (Fig. 2).

All of these results are in accordance with results obtained by others who showed that BET proteins interact with the MLV IN protein and that both proteins co-localize into the nucleus of the cells (De Rijck et al., 2013; Gupta et al., 2013). Interaction between the two partners has been shown to involve the CTD of IN and the ET domain of BET proteins (Gupta et al., 2013; Sharma et al., 2013). Moreover, disruption of the MLV IN – BET interaction by using specific BET inhibitors, blocks

Fig. 5. The C-terminal domains of BRD2, BRD3 and BRD4 proteins interact with PERV IN in the two-hybrid and *in cellulo* assays. (A) Domain organization of the BRD2 proteins. The hBET proteins contain several functional domains: (i) the bromodomains I and II (BD1 and BD2) bind acetylated marks on histones H3 and H4, (ii) the Extra-Terminal domain (ET) interacts with various viral and cellular proteins. Adapted from (Weidner-Glunde et al., 2010). (B) Two-hybrid assay using the full-length PERV IN and the C-terminal domains of hBET proteins. The C-terminal domains of BRD2, BRD3 and BRD4 proteins encompassing the ET domain, or the BRD4 ET domain alone were fused to the Gal4-activation domain (GAL4-AD). The ability of the full-length PERV IN to interact with these domains was analyzed in the two-hybrid assay as described in Fig. 1. (C) *In cellulo* assay: co-immunoprecipitation. HEK 293T cells were co-transfected: (i) with the pCMV-Myc-BRD2-Cter (lanes 1, 2 and 3), the pCMV-Myc-BRD3-Cter (lanes 4, 5 and 6) or the pCMV-Myc-BRD4-Cter (7, 8 and 9) and (ii) with the pCI-Flag-PERV IN-CTD (lanes 2, 5 and 8) or the pCI-Flag-PERV IN-CTD (P387A/L388A) (lanes 3, 6 and 9). Forty-eight hours later, cell lysates were collected. Expression of the Myc-BRD2-Cter, Myc-BRD3-Cter and Myc-BRD4-Cter proteins and of the Flag-PERV IN CTD or Flag-PERV IN-CTD (P387A/L388A) proteins was verified in the total cell lysates with anti-Myc and anti-Flag antibodies, respectively (panels A and B). Myc-BRD2-Cter, Myc-BRD3-Cter and Myc-BRD4-Cter proteins were then immuno-precipitated with the anti-Myc antibody and revealed with the Myc antibody (panel C). Co-immunoprecipitation of the PERV IN-CTD and PERV IN-CTD (P387A/L388A) proteins was assessed by a western blot analysis using an anti-Flag antibody (panel D).

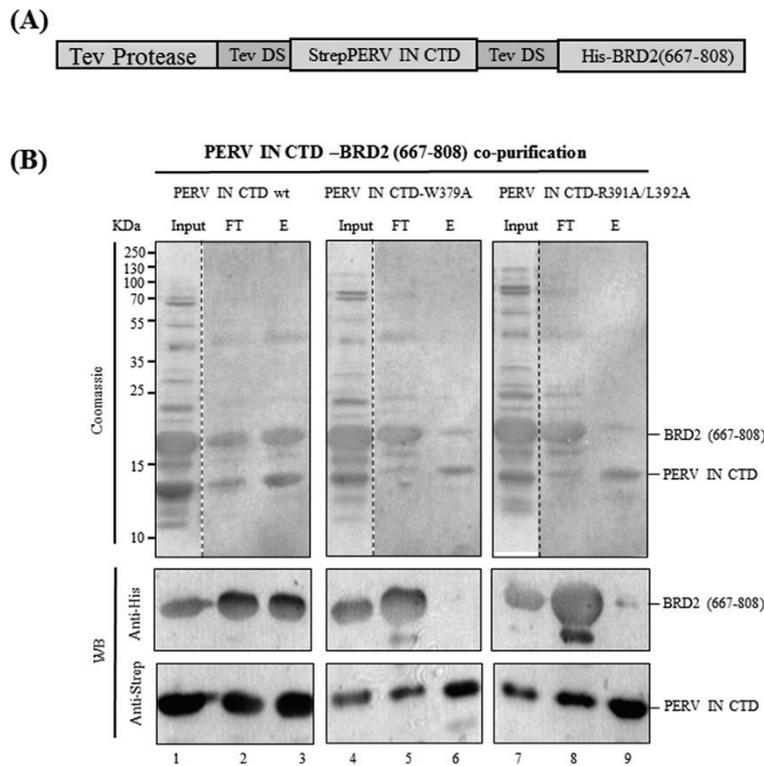


Fig. 6. Analysis of PERV IN CTD interaction with hBET C-terminal domains *in vitro*. (A) The pPERV IN CTD-BRD2 (667–808) plasmid is a polycistronic construction for the co-expression and co-purification of PERV IN CTD and BRD2 (667–808). This construction expresses a polyprotein containing the TEV protease and the Strep-PERV IN CTD (309–405) fragment with the His-BRD2 (667–808) fragment, both placed after a TEV digestion site (TEV DS). During overexpression, the TEV protease cleaves the TEV sites, if any, leaving the formation of a complex between PERV IN CTD and BRD2 (667–808) fragments. (B) SDS-PAGE gel illustrating the co-expression and co-purification of the wild-type (WT) Strep-PERV IN CTD, or of the Strep-PERV IN CTD- W379A and - R391A/L392A mutants with His-BRD2 (667–808). Two sequential affinity purifications were performed on nickel and Strep-Trap columns. Non-eluted samples (Lanes 1, 4 and 7; called “Input”) and samples eluted from the nickel column (lanes 2, 5 and 8; called flow-through “FT”) after eluted from the Strep-Trap column (lanes 3, 6 and 9; called “E”) were loaded on an SDS-PAGE gel; proteins were either revealed by Coomassie staining (upper panels) or were visualized by western blot using anti-His and anti-Strep-tag II antibodies targeting His-BRD2 (667–808) and Strep-PERV IN CTD, respectively (lower panels).

Table 1

Statistics of SAXS data collection and refinement from the two-fragment complexes.

Data collection	IN CTD - ET	IN CTD - BRD2 (667–808)
Analyzed fragment	IN CTD - ET	IN CTD - BRD2 (667–808)
Instrument	BM29 ESRF	BM29 ESRF
q-range (nm ⁻¹)	0–5	0–5
Exposure time (s)/nb frames	1/10	1/10
Concentration (mg/ml)	15	6
Temperature (°C)	4	4
Data processing		
Primary data reduction	EDNA	EDNA
Data processing	CHROMIXS	CHROMIXS
q-range (nm ⁻¹)	0.1–3.2	0.1–2.2
Rg (nm) [from Guinier]	2.4 ± .07	3.1 ± .1
Rg (nm) [from P(r)]	2.4	3.1
Dmax (nm)	7.4	11
Ab initio analysis	DAMMIF	DAMMIF
Number of models	50	50
Validation and averaging	DAMAVR	DAMAVR
NSD	0.87 ± .06	0.74 ± .07
Rigid-body modeling	DADIMODO	N.A.
Model χ^2	1.0	N.A.

MLV replication at the integration step and — when integration occurs — retargets MLV integration away from transcription start sites (De Rijck et al., 2013; Gupta et al., 2013; Sharma et al., 2013).

We combined SAXS and homology modeling to generate the first experimentally restrained atomic model of a PERV IN CTD – BET ET complex. This structural study reveals that PERV IN CTD and BRD2 ET assemble as heterodimers to form a modular structure (Fig. 7D). The first part of the C-terminal tail of IN bridges the SH3 heart of the CTD to an intermolecular domain that contains the PERV IN EBM and the BET

ET.

Keeping in mind that the estimated resolution of the SAXS-derived envelope is 3 nm and that atomic details are inferred by homology-modeling, the heterodimeric assembly is likely stabilized by an intermolecular β -sheet with two EBM β -strands and one ET β -strand as observed in the NMR structure of bound MLV IN EBM (Crowe et al., 2016), which shares a strong sequence identity with PERV IN EBM – BRD2 ET.

The SAXS-restrained model can also be used with caution (at the atomic level, structural features must be conserved in bound MLV IN EBM) to propose structural explanations to our interaction results: (i) the single mutant W379A is disruptive and W379 is modeled as a buried residue making numerous intramolecular interactions, suggesting that this tryptophan is essential for the stability of the folded tail. Moreover, this residue is strictly conserved in gamma-retroviral INs and known as a critical hotspot for interaction; (ii) the disruptive double mutant P387A/L388A may alter the turn connecting the two β -strands of the modeled bound EBM (Fig. 3) and hinder the induced folding process; (iii) the disruptive mutants K389A/L390A and R391A/L392A (Figs. 4 and 6) cancel the -BHBH- motif of the conserved gamma-retroviral stretch, where -B- and -H- are basic and hydrophobic residues. Thus, the modeled intermolecular interactions between the basic side-chains and conserved acidic residues of the neighboring ET β -strand could be essential to the stability of the heterodimer. Of interest, an amphipathic -BHBH- binding motif has been observed in other viral and cellular partners of BET proteins—such as the transcriptional regulator NSD3—where residues equivalent in position to the pair R391/L392 are essential for ET recognition (Zhang et al., 2016); (iv) the disruptive double mutant H393A/R394A affects the C-terminal arginine of EBM, which is strictly conserved among gamma-retroviral INs and contacts the critical tryptophan W379 in the model.

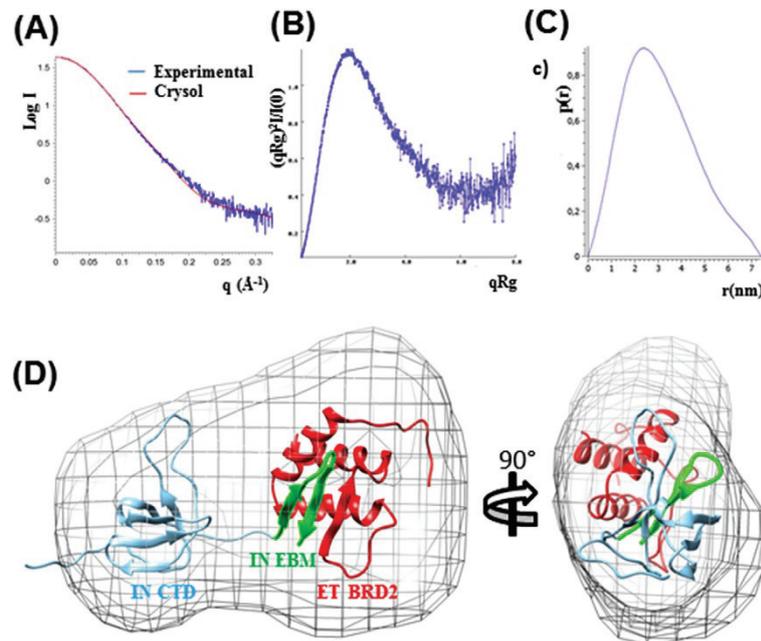


Fig. 7. SAXS analysis of PERV IN CTD – BRD2 ET. (A) Experimental diffraction curve in blue and CRYSOLOG curve in red calculated from the atomic model. (B) Dimensionless Kratky plot with a plateau in the higher q region reflecting flexibility. (C) Distance distribution function $P(r)$. (D) *Ab initio* SAXS envelope and ribbon representation of the SAXS-restrained model with the CTD of PERV IN in blue, the EBM of IN in green and the bound ET domain of BRD2 in red. The figures A, B and C were rendered with ATSAS, and the figure D was rendered with CHIMERA.

Our SAXS data also reveal that the PERV IN CTD – BRD2 ET complex possesses an inherent flexibility due to a short loop connecting the bound IN EBM segment and the IN CTD core. This flexibility is also observed in the bigger PERV IN CTD – BRD2 (667–808) complex and might be advantageous in the context of the full-length proteins for a dynamic selection of the integration sites. Moreover, the recent publications of numerous intasome structures (Ballandras-Colas et al., 2016, 2017; Hare et al., 2010; Passos et al., 2017; Yin et al., 2016) allowed us to gain information on the formation of the PERV IN – BRD2 assembly.

The oligomerization state of the known intasome structure differs between retroviral genera and seems to be related to the length of the IN CCD-CTD linker (for review, see (Engelman and Cherepanov, 2017)). Gamma-retroviral intasomes should be tetrameric, as observed in the PFV intasome (Engelman and Cherepanov, 2017). Our previously published model of the PERV intasome respects this architecture (Demange et al., 2015) with two symmetrical inner and outer monomers of INs. The two inner CTDs are positioned between their corresponding inner NTDs and CCDs (Fig. 8), whereas the outer CTDs are not modeled because they were too flexible to be observed in the PFV intasome. A superposition with our SAXS-restrained model of the PERV IN CTD – BRD2 ET complex shows that the inner CTDs are well accessible to the ET domain of the BET cofactor (Fig. 8). This accessibility remains unchanged if we consider the intasome to be bound with target DNA as observed for PFV (Maertens et al., 2010). Altogether, these data suggest that, before integration, the accessible, flexible C-terminal tail of the inner IN CTD can hook the ET domain of the BET cofactor. The induced β -folding of the IN EBM segment locks the two partners together, while the short loop preceding IN EBM gives partial flexibility to the complex. This structural combination participates in the efficient targeting of the bound vDNA to the promoters of active genes.

5. Conclusion

Retroviral vectors are important tools and have been widely used to deliver DNA for therapeutic purposes. However, retroviral integration

can be associated with insertional oncogene activation, especially for gamma-retroviral vectors due to their strong bias of integration into transcription start sites. Eliminating the interaction of gamma-retroviral IN with hBET family members has implications for the safety of human gene therapy and gene delivery when using gamma-retroviral vectors. Besides, it has been shown that the PERV recombinant virus can infect human cells and represents a potential major risk of zoonotic disease in the case of xenotransplantation of pig organs to humans. We have previously shown that Raltegravir (Demange et al., 2015), as well as AZT (Denner, 2017), are potent inhibitors of PERV replication and could be useful for the treatment of PERV infections. Eliminating the interaction of PERV IN with hBET proteins using inhibitors such as JQ1 would be another way to control this zoonotic risk.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding authors on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the French Institute INRA (Institut

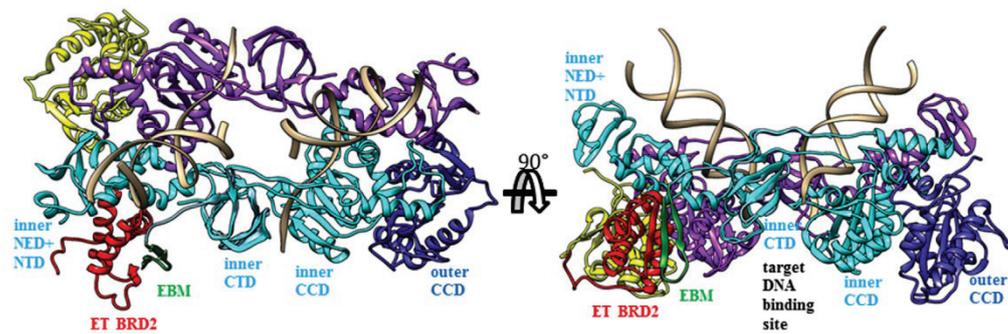


Fig. 8. Ribbon representation of the modeled PERV IN intasome superposed with the SAXS-restrained model of PERV IN CTD – BRD2 ET. The two symmetric inner IN monomers are painted cyan and violet, whereas the two outer INs are blue and yellow. The PERV IN EBM segments and the bound BRD2 ET of the superposed SAXS-restrained model are colored in green and red, respectively. The inner CTD of the intasome and the CTD of the heterodimeric complex were superimposed with COOT. The figure was rendered with CHIMERA.

National de la Recherche Agronomique), ANSES (Agence Nationale sécurité sanitaire de l'alimentation, de l'environnement et du travail) and CNRS (Centre National de la Recherche Scientifique).

MC is supported by a PhD grant from the Rhône-Alpes-Auvergne region (N°18003930 ARC Health 2016).

Authors' contributions

CR, YB, KM and PG designed the experiments, analyzed data and drafted the manuscript. KG, GB, MC, HY, and FF made substantial contributions to conception, design, acquisition and analysis of data and drafted the manuscript. MPC, CDB, AL, CL performed the experiments. ML and YC discussed the conception and analysis of some experiments. All authors read and approved the final manuscript.

Acknowledgments

We are grateful to the Production Sciences Facility (PSF) and to the Plateau Technique Imagerie/Microscopie (PLATIM) platforms of UMS3444 BioSciences Gerland Lyon Sud. We thank Lise-Marie Donnio¹, Romain Appourchaux¹, Anne-Sophie Guyomard², Séverine Calvo² and Margot François⁶ for technical assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.virol.2019.04.002>.

References

- Aiyer, S., Swapna, G.V., Malani, N., Aramini, J.M., Schneider, W.M., Plumb, M.R., Ghanem, M., Larue, R.C., Sharma, A., Studamire, B., Kvaratskhelia, M., Bushman, F.D., Montelione, G.T., Roth, M.J., 2014. Altering murine leukemia virus integration through disruption of the integrase and BET protein family interaction. *Nucleic Acids Res.* 42, 5917–5928.
- Andrake, M.D., Skalka, A.M., 2015. Retroviral integrase: then and now. *Annu. Rev. Virol.* 2, 241–264.
- Ballandras-Colas, A., Brown, M., Cook, N.J., Dewdney, T.G., Demeler, B., Cherepanov, P., Lyumkis, D., Engelman, A.N., 2016. Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* 530, 358–361.
- Ballandras-Colas, A., Maskell, D.P., Serrao, E., Locke, J., Swuec, P., Jonsson, S.R., Kotecha, A., Cook, N.J., Pye, V.E., Taylor, I.A., Andresdottir, V., Engelman, A.N., Costa, A., Cherepanov, P., 2017. A supramolecular assembly mediates lentiviral DNA integration. *Science* 355, 93–95.
- Belkina, A.C., Denis, G.V., 2012. BET domain co-regulators in obesity, inflammation and cancer. *Nat. Rev. Canc.* 12, 465–477.
- Bunkoczi, G., Read, R.J., 2011. Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr. D Biol. Crystallogr.* 67, 303–312.
- Busschots, K., Vercammen, J., Emiliani, S., Benarous, R., Engelborghs, Y., Christ, F., Debysier, Z., 2005. The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J. Biol. Chem.* 280, 17841–17847.

- Cellier, C., Moreau, K., Gally, 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.
- Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beumelen, J., Engelborghs, Y., De Clercq, E., Debysier, Z., 2003. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem.* 278, 372–381.
- Crowe, B.L., Larue, R.C., Yuan, C., Hess, S., Kvaratskhelia, M., Foster, M.P., 2016. 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.* 113, 2086–2091.
- De Ravin, S.S., Su, L., Theobald, N., Choi, U., Macpherson, J.L., Poidinger, M., Symonds, G., Pond, S.M., Ferris, A.L., Hughes, S.H., Malech, H.L., Wu, X., 2014. Enhancers are major targets for murine leukemia virus vector integration. *J. Virol.* 88, 4504–4513.
- De Rijck, J., de Kogel, C., Demeulemeester, J., Vets, S., El Ashkar, S., Malani, N., Bushman, F.D., Landuyt, B., Husson, S.J., Busschots, K., Gijssbers, R., Debysier, Z., 2013. The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Rep.* 5, 886–894.
- Demange, A., Yajjou-Hamalian, H., Gally, K., Luengo, C., Beven, V., Leroux, A., Confort, M.P., Al Andary, E., Gouet, P., Moreau, K., Ronfort, C., Blanchard, Y., 2015. Porcine endogenous retrovirus-A/C: biochemical properties of its integrase and susceptibility to raltegravir. *J. Gen. Virol.* 96, 3124–3130.
- Denner, J., 2017. Can antiretroviral drugs be used to treat porcine endogenous retrovirus (PERV) infection after xenotransplantation? *Viruses* 9.
- El Ashkar, S., De Rijck, J., Demeulemeester, J., Vets, S., Madlala, P., Cermakova, K., Debysier, Z., Gijssbers, R., 2014. BET-independent MLV-based vectors target away from promoters and regulatory elements. *Mol. Ther. Nucleic Acids* 3, e179.
- Emiliani, S., Mousnier, A., Busschots, K., Maroun, M., Van Maele, B., Tempe, D., Vandekerckhove, L., Moisan, F., Ben-Slama, L., Witvrouw, M., Christ, F., Rain, J.C., Dargemont, C., Debysier, Z., Benarous, R., 2005. Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *J. Biol. Chem.* 280, 25517–25523.
- Engelman, A.N., Cherepanov, P., 2017. Retroviral intasomes arising. *Curr. Opin. Struct. Biol.* 47, 23–29.
- Evrard, G., Mareuil, F., Bontems, F., Sizun, C., Perez, J., 2011. DADIMODO: a program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints. *Appl. Crystallogr.* 44, 1264–1271.
- Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., Svergun, D.I., 2017. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* 50, 1212–1225.
- Guan, R., Aiyer, S., Cote, M.L., Xiao, R., Jiang, M., Acton, T.B., Roth, M.J., Montelione, G.T., 2017. X-ray crystal structure of the N-terminal region of Moloney murine leukemia virus integrase and its implications for viral DNA recognition. *Proteins* 85, 647–656.
- Gupta, S.S., Maetzig, T., Maertens, G.N., Sharif, A., Rothe, M., Weidner-Glunde, M., Galla, M., Schambach, A., Cherepanov, P., Schulz, T.F., 2013. Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J. Virol.* 87, 12721–12736.
- Hare, S., Gupta, S.S., Valkov, E., Engelman, A., Cherepanov, P., 2010. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 464, 232–236.
- Jang, M.K., Kwon, D., McBride, A.A., 2009. Papillomavirus E2 proteins and the host BRD4 protein associate with transcriptionally active cellular chromatin. *J. Virol.* 83, 2592–2600.
- 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.
- LaFave, M.C., Varshney, G.K., Gildea, D.E., Wolfsberg, T.G., Baxevanis, A.D., Burgess, S.M., 2014. MLV integration site selection is driven by strong enhancers and active promoters. *Nucleic Acids Res.* 42, 4257–4269.
- Larue, R.C., Plumb, M.R., Crowe, B.L., Shkriabai, N., Sharma, A., DiFiore, J., Malani, N.,

- Aiyer, S.S., Roth, M.J., Bushman, F.D., Foster, M.P., Kvaratskhelia, M., 2014. Bimodal high-affinity association of Brd4 with murine leukemia virus integrase and mononucleosomes. *Nucleic Acids Res.* 42, 4868–4881.
- Levy, N., Eiler, S., Pradeau-Aubret, K., Maillot, B., Stricher, F., Ruff, M., 2016. Production of unstable proteins through the formation of stable core complexes. *Nat. Commun.* 7, 10932.
- Lin, Y.J., Umehara, T., Inoue, M., Saito, K., Kigawa, T., Jang, M.K., Ozato, K., Yokoyama, S., Padmanabhan, B., Guntert, P., 2008. Solution structure of the extraterminal domain of the bromodomain-containing protein BRD4. *Protein Sci.* 17, 2174–2179.
- Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., Russell, R.B., 2003. Protein disorder prediction: implications for structural proteomics. *Structure* 11, 1453–1459.
- Llano, M., Saenz, D.T., Meehan, A., Wongthida, P., Peretz, M., Walker, W.H., Teo, W., Poeschla, E.M., 2006. An essential role for LEDGF/p75 in HIV integration. *Science* 314, 461–464.
- Maertens, G., Cherepanov, P., Plumbers, W., Busschots, K., De Clercq, E., Debysers, Z., Engelborghs, Y., 2003. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J. Biol. Chem.* 278, 33528–33539.
- Maertens, G.N., 2016. B⁺-protein phosphatase 2A is a functional binding partner of delta-retroviral integrase. *Nucleic Acids Res.* 44, 364–376.
- 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.
- 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., Felix, 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., Felix, 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., 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.
- 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.
- Passos, D.O., Li, M., Yang, R., Rebsburg, S.V., Ghirlando, R., Jeon, Y., Shkriabai, N., Kvaratskhelia, M., Craigie, R., Lyumkis, D., 2017. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* 355, 89–92.
- Petersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Sanchez, R., Zhou, M.M., 2009. The role of human bromodomains in chromatin biology and gene transcription. *Curr. Opin. Drug Discov. Dev* 12, 659–665.
- Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., Bushman, F., 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110, 521–529.
- Serrao, E., Ballandras-Colas, A., Cherepanov, P., Maertens, G.N., Engelman, A.N., 2015. Key determinants of target DNA recognition by retroviral intasomes. *Retrovirology* 12, 39.
- Sharma, A., Larue, R.C., Plumb, M.R., Malani, N., Male, F., Slaughter, A., Kessl, J.J., Shkriabai, N., Coward, E., Aiyer, S.S., Green, P.L., Wu, L., Roth, M.J., Bushman, F.D., Kvaratskhelia, M., 2013. BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12036–12041.
- Sowd, G.A., Serrao, E., Wang, H., Wang, W., Fadel, H.J., Poeschla, E.M., Engelman, A.N., 2016. A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proc. Natl. Acad. Sci. U. S. A.* 113, E1054–E1063.
- Studamire, B., Goff, S.P., 2008. Host proteins interacting with the Moloney murine leukemia virus integrase: multiple transcriptional regulators and chromatin binding factors. *Retrovirology* 5, 48.
- Thorpe, K.L., Abdulla, S., Kaufman, J., Trowsdale, J., Beck, S., 1996. Phylogeny and structure of the RING3 gene. *Immunogenetics* 44, 391–396.
- Viejo-Borbolla, A., Ottinger, M., Bruning, E., Burger, A., Konig, R., Kati, E., Sheldon, J.A., Schulz, T.F., 2005. Brd2/RING3 interacts with a chromatin-binding domain in the Kaposi's Sarcoma-associated herpesvirus latency-associated nuclear antigen 1 (LANA-1) that is required for multiple functions of LANA-1. *J. Virol.* 79, 13618–13629.
- Wanaguru, M., Barry, D.J., Benton, D.J., O'Reilly, N.J., Bishop, K.N., 2018. Murine leukemia virus p12 tethers the capsid-containing pre-integration complex to chromatin by binding directly to host nucleosomes in mitosis. *PLoS Pathog.* 14, e1007117.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303.
- Weidner-Glunde, M., Ottinger, M., Schulz, T.F., 2010. WHAT do viruses BET on? *Front. Biosci.* 15, 537–549.
- Winans, S., Larue, R.C., Abraham, C.M., Shkriabai, N., Skopp, A., Winkler, D., Kvaratskhelia, M., Beemon, K.L., 2017. The FACT complex promotes avian leukosis virus DNA integration. *J. Virol.* 91 (7) pii: e00082-17.
- Wu, S.Y., Chiang, C.M., 2007. The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. *J. Biol. Chem.* 282, 13141–13145.
- Wu, X., Li, Y., Crise, B., Burgess, S.M., 2003. Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300, 1749–1751.
- Yin, Z., Shi, K., Banerjee, S., Pandey, K.K., Bera, S., Grandgenett, D.P., Aihara, H., 2016. Crystal structure of the Rous sarcoma virus intasome. *Nature* 530, 362–366.
- Zhang, Q., Zeng, L., Shen, C., Ju, Y., Konuma, T., Zhao, C., Vakoc, C.R., Zhou, M.M., 2016. Structural mechanism of transcriptional regulator NSD3 recognition by the ET domain of BRD4. *Structure* 24, 1201–1208.

II. Étude de l'intasome de PERV-A/C

Après avoir obtenu des résultats originaux sur l'interaction d'IN PERV-A/C avec son cofacteur cellulaire, nous avons souhaité surproduire l'intasome de PERV-A/C afin de tenter de résoudre la première structure atomique d'un intasome de *Gammarétrovirus*. La première étape de cette étude implique la mise au point d'un protocole de surproduction de l'IN entière.

II. 1. Expression et purification de l'IN de PERV-A/C

II. 1. 1. Tests de surproduction en système bactérien

J'ai testé différents protocoles d'expression afin de surproduire l'IN de PERV-A/C sous forme soluble, en changeant de bactéries d'expression et en faisant varier le temps d'incubation, la température d'incubation, la concentration d'IPTG et les additifs (voir **Matériels et Méthodes §II.3.**).

Dans chaque cas, les différentes fractions solubles et insolubles ont été analysées par SDS-PAGE, et la présence d'une bande correspondant à l'IN de PERV-A/C (d'un poids moléculaire attendu de 49,9 kDa) a été vérifiée.

Aucun de ces tests d'expression n'a permis d'obtenir la surproduction de la protéine d'intérêt sous forme soluble. En effet, dans une majorité de cas, l'IN de PERV-A/C a bien été surexprimée mais a été intégralement retrouvée dans la fraction insoluble avec le culot bactérien. Il est probable que son repliement ne se déroule pas correctement lors de l'expression et qu'elle se retrouve dans des corps d'inclusion.

L'ensemble des conditions d'expression testées sont résumées dans le **Tableau 10** et plusieurs résultats sont détaillés ci-après.

		IPTG									
		0,1 mM				0,5 mM		1 mM			
		1h	2h	3h	3h	3h	16h	2h	3h	4h	16h
T°C		20	20	20	37	37	18	37	37	37	18
Glucose											
pET-28a	T7 Express	0									
	Rosetta	1%									
		0									
	Bl21 (DE3)-pLysS	1%									
		0									
C41 (DE3)	1%										
pPROEX-Hta	T7 Express	1%									
	Rosetta	1%									

Tableau 10 | Tableau des tests d'expression des deux constructions de l'IN PERV-A/C en bactéries *E. coli* T7 Express, Rosetta, Bl21 (DE3)-pLysS et C41 (DE3). 1, test en présence et absence de rifampicine et 2, test en présence et absence de Zn²⁺. En rouge, pas d'expression protéique et en orange, protéine insoluble.

On peut déjà remarquer, dans le **Tableau 10**, que les tests d'expression réalisés avec un milieu de culture non supplémenté en glucose n'ont été positifs qu'à deux reprises, contre dix résultats positifs pour un milieu avec 1% de glucose. Mais, dans tous les cas, l'IN n'a été obtenue qu'en faible proportion et dans le culot bactérien.

On peut également noter que deux tests ont été réalisés en présence d'additifs : celui noté 1 avec l'ajout de rifampicine après une heure d'induction et celui noté 2, supplémenté en Zn²⁺. La rifampicine est un antibiotique permettant de bloquer l'étape de transcription des ARN. En stoppant cette étape après une heure d'induction, seule la traduction des ARNm déjà produits peut être réalisée par la machinerie bactérienne. L'ajout d'ions Zn²⁺, quant à lui, permet de stabiliser le NTD de l'IN qui comprend un motif HHCC en doigt de zinc. L'analyse SDS-PAGE

montre la présence, dans la fraction insoluble, d'une bande à hauteur du marqueur de poids moléculaire de 55 kDa, masse proche de celle d'IN PERV-A/C (**Figure 66**).

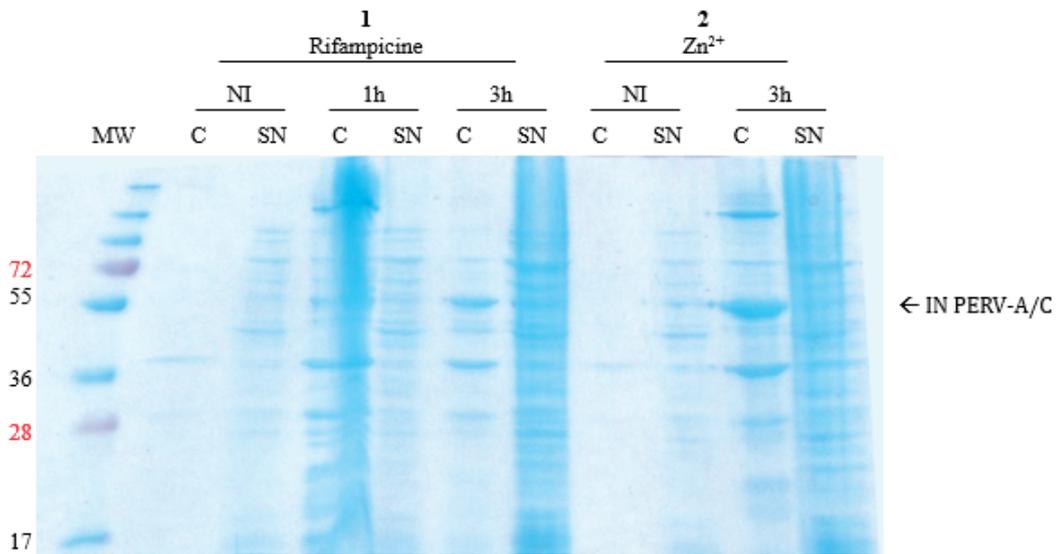


Figure 66 | Analyse par SDS-PAGE de l'expression de l'IN PERV-A/C portée par le plasmide pET-28a en bactéries *E. coli* B121 (DE3)-pLysS en présence de 0,5 mM IPTG et incubées à 37°C pendant 3h. Cas 1, ajout de 100 µg/mL de rifampicine après 1h d'induction et cas 2, supplémentation en Zn²⁺ lors de la mise en culture. Avec NI, fraction non induite ; C, fraction insoluble ; SN, fraction soluble et MW, marqueurs de poids moléculaires.

On peut enfin noter que le plasmide pPROEX-Hta portant l'IN et exprimé en bactéries *E. coli* Rosetta a généralement donné des résultats positifs, la protéine produite demeurant cependant dans la fraction insoluble. La **Figure 67** permet ainsi de comparer, par SDS-PAGE, un test d'expression positif obtenu avec ajout d'1mM d'IPTG pendant 2h à 37°C avec le seul test d'expression négatif obtenu avec ajout d'1 mM d'IPTG pendant 16h à 18°C.

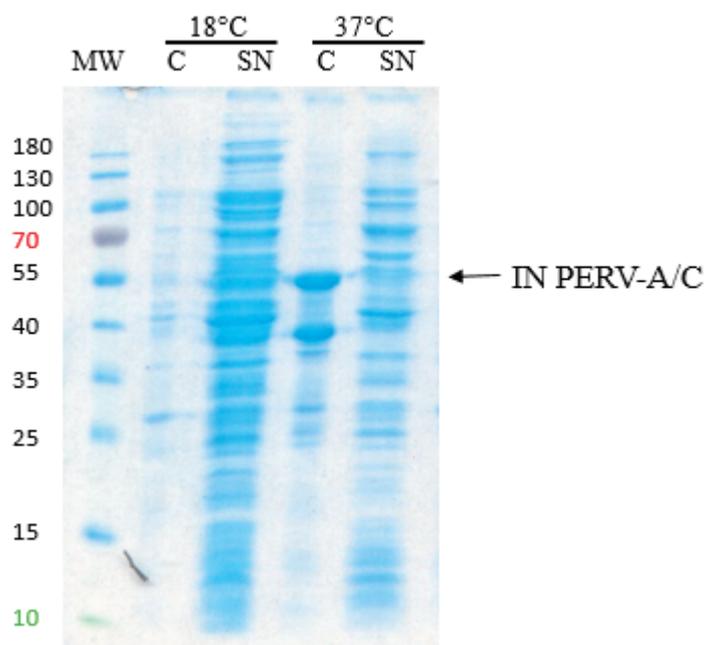


Figure 67 | Analyse par SDS-PAGE de l'expression de l'IN PERV-A/C portée par le plasmide pPROEX-Hta en bactéries *E. coli* Rosetta en présence de 1 mM IPTG et incubées à 18°C pendant 16h ou 37°C pendant 2h. Avec C, fraction insoluble ; SN, fraction soluble et MW, marqueurs de poids moléculaires.

Le protocole de surexpression d'IN PERV-A/C portée par le plasmide pET-28a en bactérie *E. coli* Rosetta, en présence d'1% glucose et après induction à 1 mM pendant 2h à 37°C, a été retenu. Bien que la protéine obtenue soit insoluble, il s'agit du processus permettant l'obtention d'une plus grande quantité protéique et des essais de solubilisation ont ensuite été réalisés.

II. 1. 2. Solubilisation de l'IN de PERV-A/C

Deux méthodes permettant de limiter l'agrégation des protéines et de favoriser leur repliement ont été testées (voir également **Matériels et Méthodes §II.3.4.**) :

- *Single freeze-thawing cycle*

Cette méthode consiste en la succession de trois centrifugations et re-suspension des culots dans des tampons contenant 1 ou 2 M d'urée. Les fractions correspondant aux fractions insolubles initiales et finales ainsi qu'aux fractions solubles après chaque centrifugation sont analysées par SDS-PAGE (**Figure 68**).

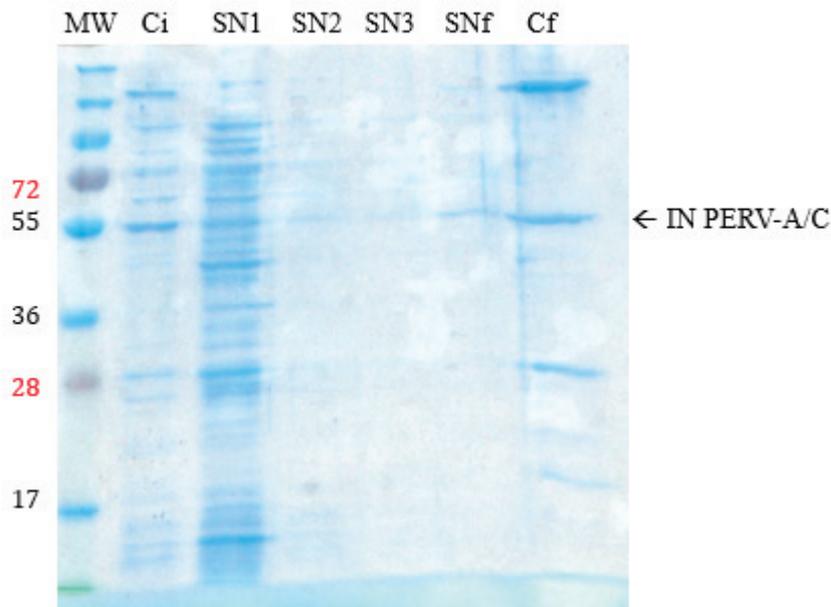


Figure 68 | Analyse par SDS-PAGE de la solubilisation de l'IN PERV-A/C par la méthode *Single freeze-thawing cycle*. Avec Ci, fraction insoluble initiale ; Cf, fraction insoluble finale ; SN, fractions solubles et MW, marqueurs de poids moléculaires.

Une bande migrant à la taille attendue pour l'IN de PERV-A/C est observée sur la piste correspondant au culot initial et représente l'échantillon à solubiliser. Dans les deux fractions finales, seule une quantité très minoritaire de l'IN est observable dans la fraction soluble, la majorité de la protéine se trouvant à nouveau dans la fraction insoluble. Ce protocole n'a donc pas été efficace pour solubiliser la protéine d'intérêt.

- Dénaturation par l'urée

Plusieurs protocoles de dénaturation par l'urée ont également été mis en place. Pour ce faire, différents tampons de lyse L8 (voir **Tableau 4**) contenant des concentrations d'urée allant de 2 à 8 M ont été testés afin de re-suspendre le culot bactérien. Les échantillons dénaturés sont ensuite injectés sur une colonne de purification Ni-NTA, où plusieurs méthodes sont testées de façon à éliminer l'urée et replier correctement la protéine. De manière générale, le tampon de lyse L8 (50 mM Tris HCl pH 7,5, 1 M NaCl, 20 mM imidazole, 2 mM β -mercaptoéthanol, 0,1 mM EDTA, 10% glycérol, 1 mM CHAPS, 1 mM $MgCl_2$, 1 μ M $ZnSO_4$ et 2 mM ATP) contenant 3 M urée a permis de récupérer la plus grande quantité de protéine dénaturée et soluble qui a été ensuite purifiée comme détaillé ci-dessous.

II. 1. 3. Tests de purification de l'IN de PERV-A/C

Plusieurs tentatives de purification de l'IN de PERV-A/C ont été réalisées à partir de l'échantillon lysé et dénaturé par l'urée 3 M. L'échantillon a été chargé sur une colonne Ni-NTA et deux approches différentes ont été envisagées.

Dans un premier temps, les tests se sont focalisés sur le repliement de la protéine d'intérêt directement sur la colonne grâce à un gradient d'urée, réalisé simultanément au gradient d'imidazole. Le passage de l'échantillon sur la colonne est suivi par mesure des absorbances à 280 nm et 260 nm, et les différentes fractions sont analysées par SDS-PAGE.

Le chromatogramme obtenu (**Figure 69A**) fait apparaître quatre pics d'élution. Le premier, et plus important, correspond à la sortie de colonne des protéines non retenues, tandis que le deuxième correspond aux molécules éluées suite au lavage avec le tampon contenant 2 mM d'ATP. Cette étape est réalisée dans le but de séparer les protéines chaperonnes interagissant avec les protéines non structurées. Après analyse de cette fraction par SDS-PAGE (**Figure 69B**), aucune bande ne correspondant à des protéines n'est observée sur le gel. Cela peut s'expliquer par le profil du pic du chromatogramme, qui présente une absorbance à 260 nm notablement plus importante que celle à 280 nm, profil caractéristique des acides nucléiques. On peut donc supposer que l'élution menée par le lavage à l'ATP entraîne uniquement des ADN ou ARN contaminants.

Au cours de l'étape d'élution, deux pics de faible intensité sont observés, le premier à 20% du gradient et le second à 50%. Ces deux fractions ont été analysées par SDS-PAGE, en pistes 1 et 2 respectivement (**Figure 69B**). Pour l'échantillon issu du premier pic d'élution, on peut observer une bande majoritaire au niveau du marqueur de poids moléculaire 55 kDa qui correspond à l'IN PERV-A/C. Des bandes de très faibles intensités sont également visibles sur la piste 1 et correspondent à des contaminants protéiques. L'échantillon issu du second pic d'élution, et déposé en piste 2, ne semble contenir qu'une quantité minime d'IN.

Ainsi, la quantité de protéine obtenue avec cette méthode de purification est très faible (1 mg de protéine).

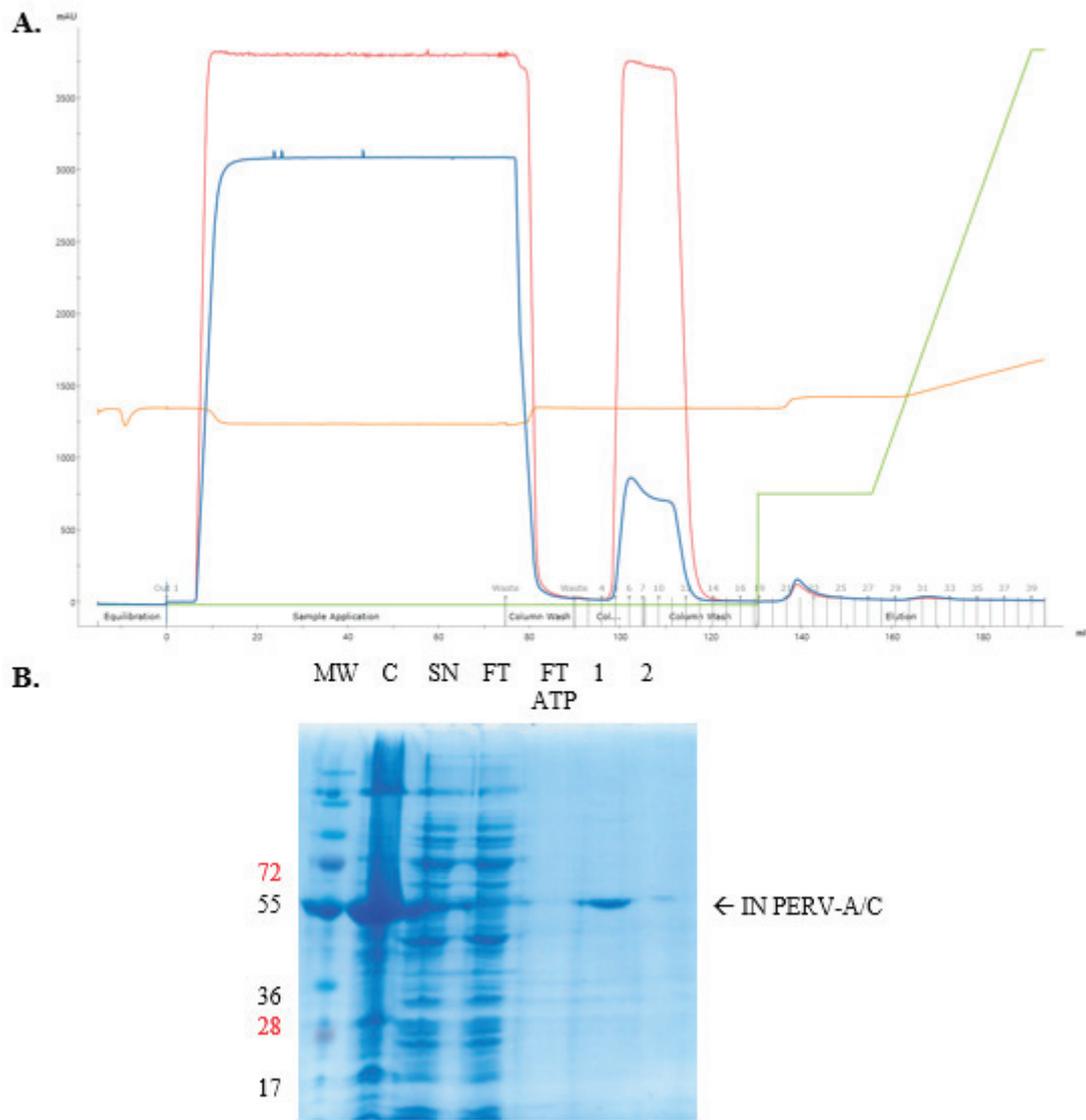


Figure 69 | Purification de l'IN de PERV-A/C avec renaturation protéique. **A.** Chromatogramme de purification par chromatographie d'affinité pour le nickel. En bleu, l'absorbance à 280 nm ; en rouge, l'absorbance à 260 nm ; en vert, le gradient d'imidazole et d'urée et en orange, la conductance. **B.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; piste C, fraction non soluble ; piste SN, fraction soluble ; piste FT, fraction non retenue ; piste FT ATP, fraction non retenue après lavage avec 2 mM ATP ; piste 1 et 2, fractions d'élution.

Dans un second temps, l'élution a eu lieu avec un gradient d'imidazole dans un tampon E8 (voir **Tableau 5**) contenant la même concentration en urée que le tampon de lyse utilisé. Dans ce cas de figure, l'élimination de l'urée et le repliement de l'IN ont été réalisés par dialyse et seront décrits ci-après (voir **Résultats et Discussion §II.3.1.**). Le passage de l'échantillon sur la

colonne est suivi par mesure des absorbances à 280 nm et 260 nm. Les différentes fractions sont ensuite analysées par SDS-PAGE.

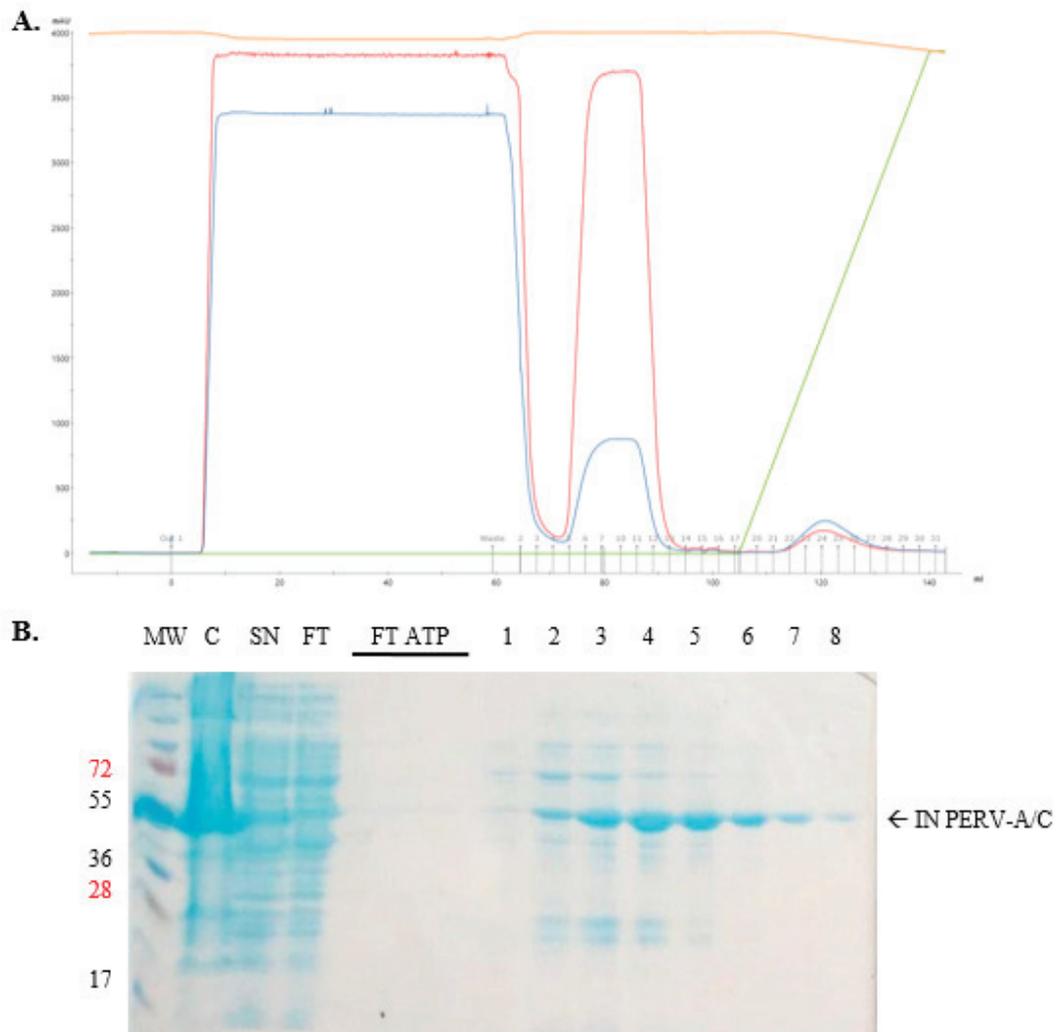


Figure 70 | Purification de l'IN de PERV-A/C en conditions dénaturantes. **A.** Chromatogramme de purification par chromatographie d'affinité pour le nickel. En bleu, l'absorbance à 280 nm ; en rouge, l'absorbance à 260 nm ; en vert, le gradient d'imidazole et en orange, la conductance. **B.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; piste C, fraction non soluble ; piste SN, fraction soluble ; piste FT, fraction non retenue ; pistes 'FT ATP', fractions non retenues après lavage avec 2 mM ATP et pistes 1 à 8, fractions d'éluion.

Le chromatogramme résultant (**Figure 70A**) montre la présence de trois pics d'éluion distincts. Comme précédemment, les deux premiers pics correspondent respectivement à la sortie de colonne des protéines non retenues et aux molécules éluées suite au lavage avec le tampon contenant 2 mM d'ATP, utilisé afin de décrocher les protéines chaperonnes. À nouveau, l'analyse par SDS-PAGE de la fraction éluee par ce lavage ne révèle aucune bande

correspondant à des protéines (**Figure 70B**). Comme précédemment, on constate une absorbance à 260 nm nettement plus importante que celle à 280 nm, ce qui laisse à penser que l'élué est principalement composé d'ADN ou d'ARN contaminants.

Un unique pic peut être observé lors de l'élué, à environ 50% d'imidazole. Les huit fractions éluées correspondantes ont été analysées par SDS-PAGE (**Figure 70B**). Dans chacune d'elles, on peut observer une bande à hauteur du marqueur de poids moléculaire 55 kDa et correspondant à l'IN. Bien que majoritaire dans l'échantillon, l'IN n'est pas entièrement pure et d'autres bandes minoritaires correspondant à des contaminants protéiques peuvent être observées.

Malgré ces quelques contaminants dont on pourra s'affranchir lors des étapes suivantes, l'échantillon élué est récupéré. Cette méthode, bien que conservant la protéine dans un tampon supplémenté en urée, a permis de purifier environ 13 mg de protéine. C'est donc cette stratégie de purification, réalisée avec les tampons L8 (50 mM Tris HCl pH 7,5, 1 M NaCl, 20 mM imidazole, 2 mM β -mercaptoéthanol, 0,1 mM EDTA, 10% glycérol, 1 mM CHAPS, 1 mM $MgCl_2$, 1 μ M $ZnSO_4$ et 2 mM ATP) et E8 (50 mM Tris HCl pH 7,5, 1 M NaCl, 0,5 M imidazole, 2 mM β -mercaptoéthanol, 0,1 mM EDTA, 10% glycérol, 1 mM CHAPS, 1 mM $MgCl_2$ et 1 μ M $ZnSO_4$) contenant 3 M d'urée, qui a été privilégiée pour l'obtention de l'IN en vue des tests de formation de l'intasome de PERV-A/C.

II. 2. Surproduction et purification de Brd2 (667-808)

Le partenaire cellulaire Brd2 (667-808) a également été purifié, afin de favoriser le repliement de l'IN de PERV-A/C, la stabilisation de l'intasome et de tenter de résoudre la première structure expérimentale d'un complexe ternaire IN-ADN-cofacteur cellulaire.

II. 2. 1. Purification de Brd2 (667-808)

La protéine d'intérêt Brd2 (667-808), portant une étiquette poly-histidine, a été surexprimée en système bactérien dans des bactéries *E. coli* T7 Express. Après lyse du culot bactérien, la fraction soluble a été chargée sur colonne Ni-NTA pour une purification par chromatographie d'affinité pour le nickel.

Le passage de l'échantillon sur la colonne est suivi par mesure des absorbances à 280 nm et 260 nm.

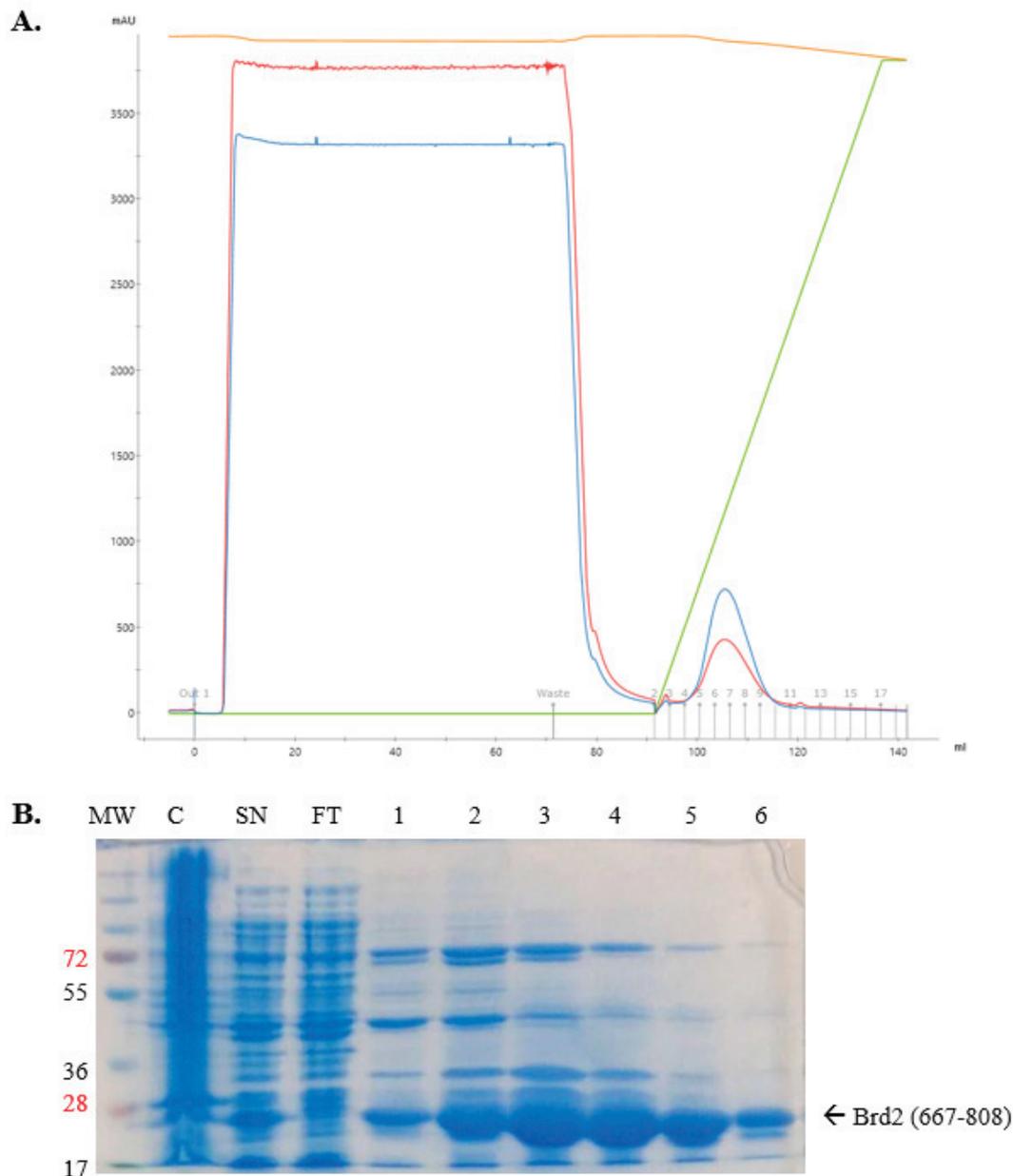


Figure 71 | Purification de Brd2 (667-808). **A.** Chromatogramme de purification par chromatographie d'affinité pour le nickel. En bleu, l'absorbance à 280 nm ; en rouge, l'absorbance à 260 nm ; en vert, le gradient d'imidazole et en orange, la conductance. **B.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; piste C, fraction non soluble ; piste SN, fraction soluble ; piste FT, fraction non retenue et pistes 1 à 6, fractions d'éluion.

L'éluion, réalisée à l'aide d'un gradient d'imidazole, a déterminé l'apparition d'un pic unique visible sur le chromatogramme (**Figure 71A**). L'analyse de ces fractions a été réalisée par SDS-PAGE (**Figure 71B**). Dans chacune de celles-ci, on peut observer une bande migrant sous le marqueur de poids moléculaire de 28 kDa et correspondant à Brd2 (667-808). Cependant, de

nombreuses bandes indiquant la présence de contaminants protéiques dans l'échantillon sont également visibles. Malgré cela, l'ensemble des fractions a été conservé, ce protocole de purification permettant d'obtenir environ 6 mg de protéine par litre de culture.

II. 2. 2. Clivage de l'étiquette d'affinité

Une étape supplémentaire a été menée afin de s'affranchir de l'étiquette poly-histidine portée par Brd2 (667-808). L'échantillon purifié a été incubé avec une quantité suffisante de protéase du TEV pour assurer le clivage complet de l'étiquette, puis dialysé dans un tampon ne contenant pas d'imidazole.

La solution protéique clivée est chargée à nouveau sur une colonne Ni-NTA et l'élution est suivie par mesure des absorbances à 280 nm et 260 nm (**Figure 72A**). Deux pics sont visibles sur le chromatogramme, le premier correspondant aux protéines non retenues sur la colonne et le deuxième aux protéines éluées à 55% du gradient d'imidazole. Il est attendu que Brd2 (667-808), dont l'étiquette a entièrement été clivée, ne soit pas retenue par la colonne, tandis que la protéase du TEV portant une étiquette poly-histidine et les contaminants interagissent avec la colonne.

Les différentes fractions ont été analysées par SDS-PAGE (**Figure 72B**). Dans chacune des fractions issues de l'échantillon non retenu sur la colonne, une bande migrant au niveau du marqueur de poids moléculaire de 17 kDa peut être observée. Celle-ci correspond à Brd2 (667-808) dont l'étiquette a été clivée, entraînant sa migration à sa taille attendue (18 kDa). On peut également observer la présence de deux bandes de plus hauts poids moléculaires indiquant la présence de contaminants résiduels. En revanche, les fractions représentant les protéines éluées par le gradient d'imidazole semblent uniquement être composées des contaminants protéiques interagissant avec la matrice de nickel.

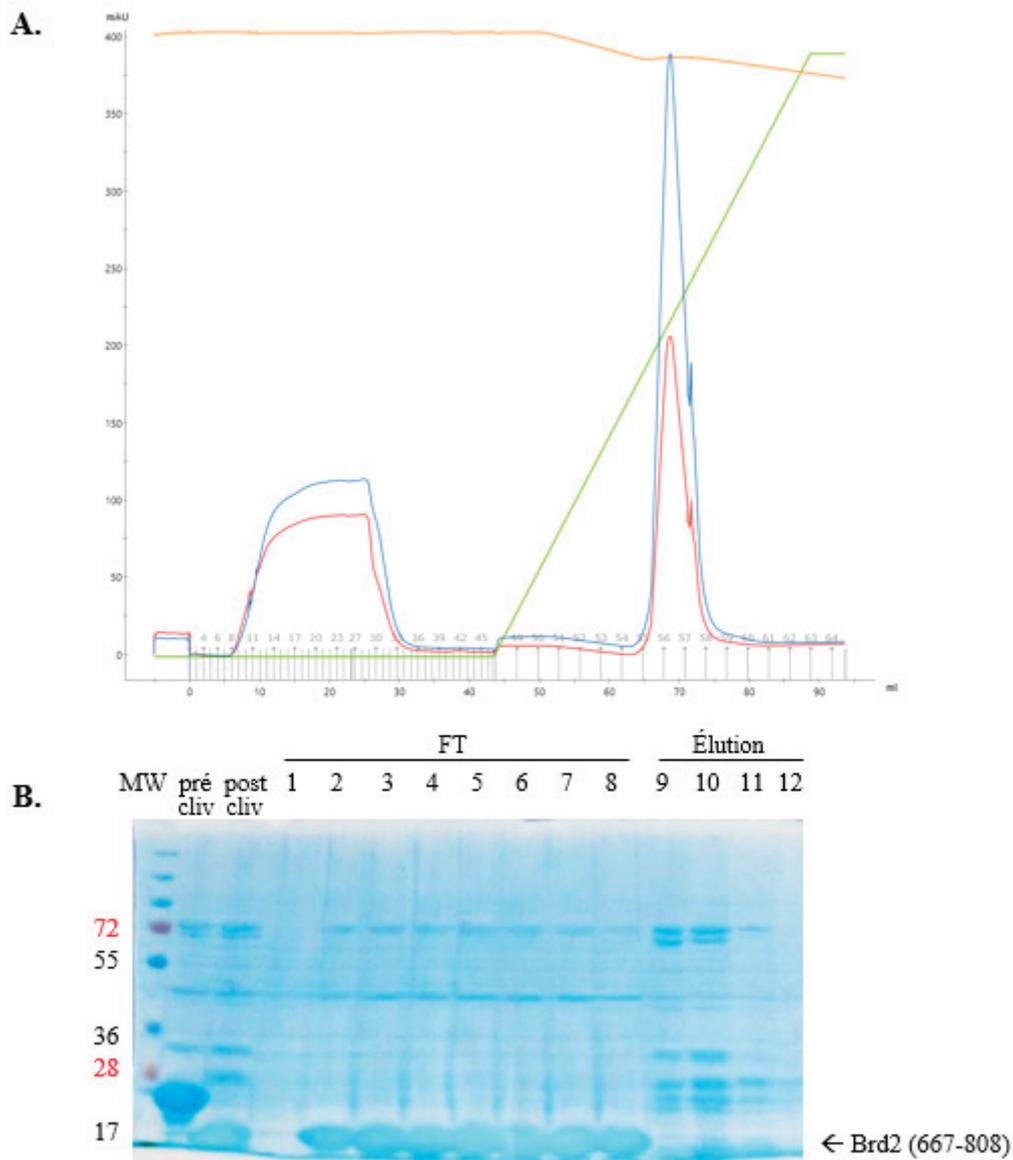


Figure 72 | Purification de Brd2 (667-808) après clivage de son étiquette poly-histidine. **A.** Chromatogramme de purification par chromatographie d'affinité pour le nickel. En bleu, l'absorbance à 280 nm ; en rouge, l'absorbance à 260 nm ; en vert, le gradient d'imidazole et en orange, la conductance. **B.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; piste 'pré cliv', échantillon avant clivage par la TEV ; piste 'post cliv', échantillon après clivage ; pistes 1 à 8, fractions non retenues (FT) et pistes 9 à 12, fractions d'élution.

Ce protocole de purification a donc permis d'obtenir une quantité importante de Brd2 (667-808) sans étiquette, qui a été utilisée dans la suite des expériences.

II. 3. Assemblage et purification de l'intasome de PERV-A/C

II. 3. 1. Tests d'assemblage de l'intasome de PERV-A/C

L'assemblage de l'intasome rétroviral nécessite de trouver les conditions optimales pour l'association *a minima* de l'IN et de l'ADN substrat, voire l'IN, l'ADN et Brd2 (667-808). Il est donc nécessaire de faire varier les conditions d'interaction des différents partenaires et de trouver le protocole idéal pour la purification du complexe formé. L'ensemble des tests réalisés est détaillé dans le **Matériels et Méthodes §II.3.6.**, ainsi que la composition de l'ADN substrat formé de trois brins hybridés. Les résultats ainsi obtenus sont présentés ci-après.

II. 3. 1. 1. Test 1

Dans ce premier essai, l'IN, Brd2 (667-808) et l'ADN ont été incubés ensemble en deux temps : tout d'abord l'IN avec Brd2 (667-808), puis avec ajout du substrat d'ADN. Après augmentation de la concentration saline et ajout de CHAPS afin de re-solubiliser l'échantillon suite à un début de précipitation, l'échantillon a été analysé par chromatographie d'exclusion de taille (**Figure 73A**). Six pics d'élution distincts sont observés sur le chromatogramme. Ceux-ci peuvent correspondre soit à des contaminants protéiques de hauts poids moléculaires, soit à différents niveaux d'oligomérisation de l'IN pour former l'intasome, soit aux trois partenaires non complexés.

Les différentes fractions d'élution ont été analysées par SDS-PAGE (**Figure 73B**). On peut tout d'abord noter qu'une majorité de l'IN se trouve dans la fraction précipitée et qu'une petite partie seulement est restée sous forme soluble et a pu être injectée sur la colonne (pistes P et SN). Une bande de très faible intensité correspondant à l'IN de PERV-A/C peut être observée dans les deux fractions englobant le pic 3. Une bande plus importante représentant Brd2 (667-808) est également visible dans ces fractions.

Ce protocole ne semble donc pas être optimal pour la formation de l'intasome, en raison de la quantité d'IN ayant précipité.

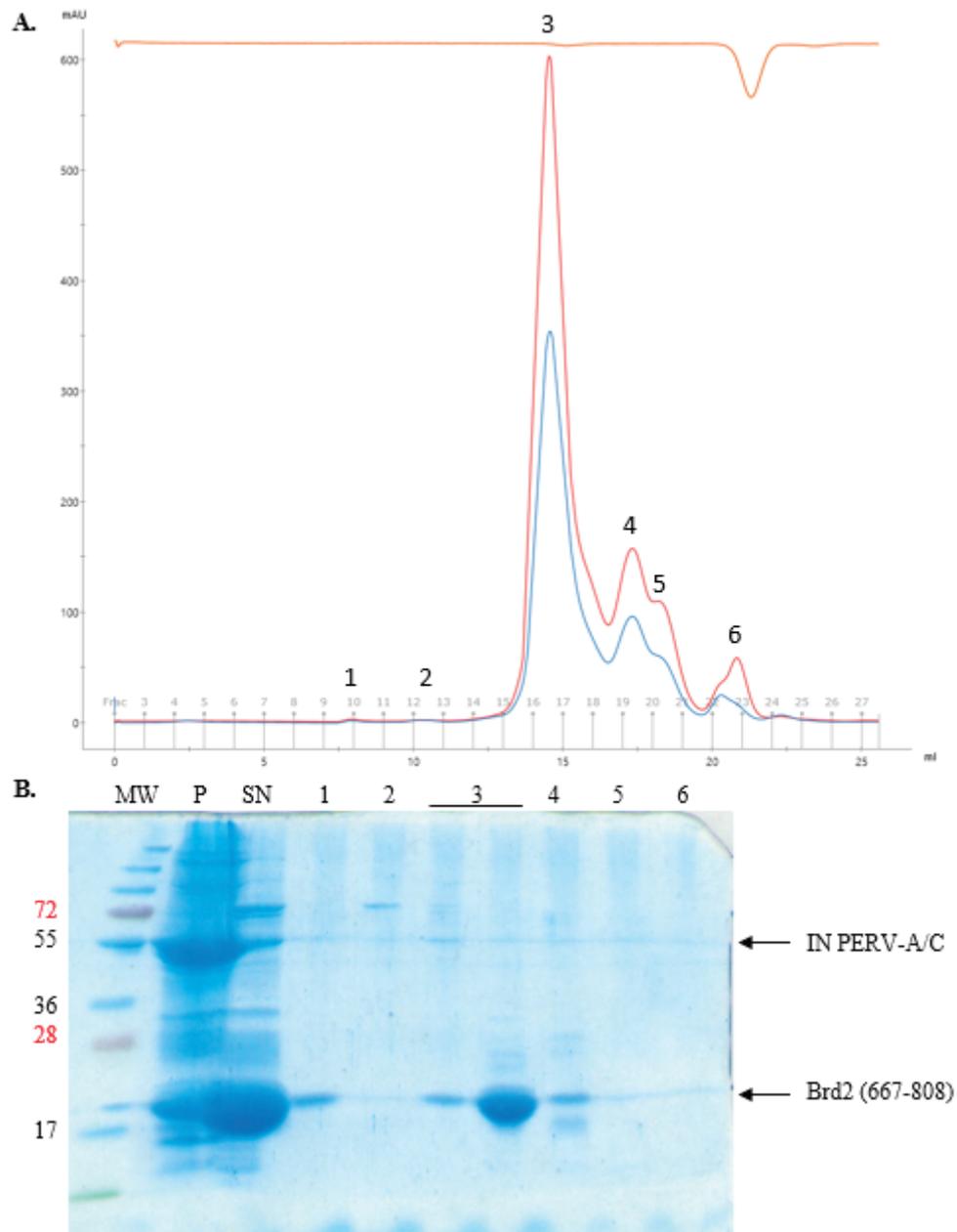


Figure 73 | Test d'assemblage de l'intasome : chromatographie d'exclusion stérique après incubation conjointe de l'IN de PERV-A/C, Brd2 (667-808) et l'ADN. **A.** Chromatogramme de purification, avec en bleu, l'absorbance à 280 nm ; en rouge, l'absorbance à 260 nm et en orange, la conductance. **B.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; piste P, fraction précipitée ; piste SN, fraction soluble ; pistes 1 à 6, fractions d'éluion.

II. 3. 1. 2. Test 2

Cette série a commencé par trois dialyses successives de l'IN et Brd2 (667-808) afin de diminuer graduellement la quantité d'urée, jusqu'à l'éliminer totalement. Afin de faciliter le repliement de l'IN lors de sa renaturation, l'ADN est ajouté à l'échantillon lors de la dernière

dialyse. La moitié de l'échantillon est conservée telle quelle et concentrée par centrifugation alors que l'autre moitié est supplémentée en zinc et CHAPS. Les deux aliquotes ensuite analysées par chromatographie d'exclusion de taille et les profils d'élutions obtenus sont identiques (**Figure 74A**).

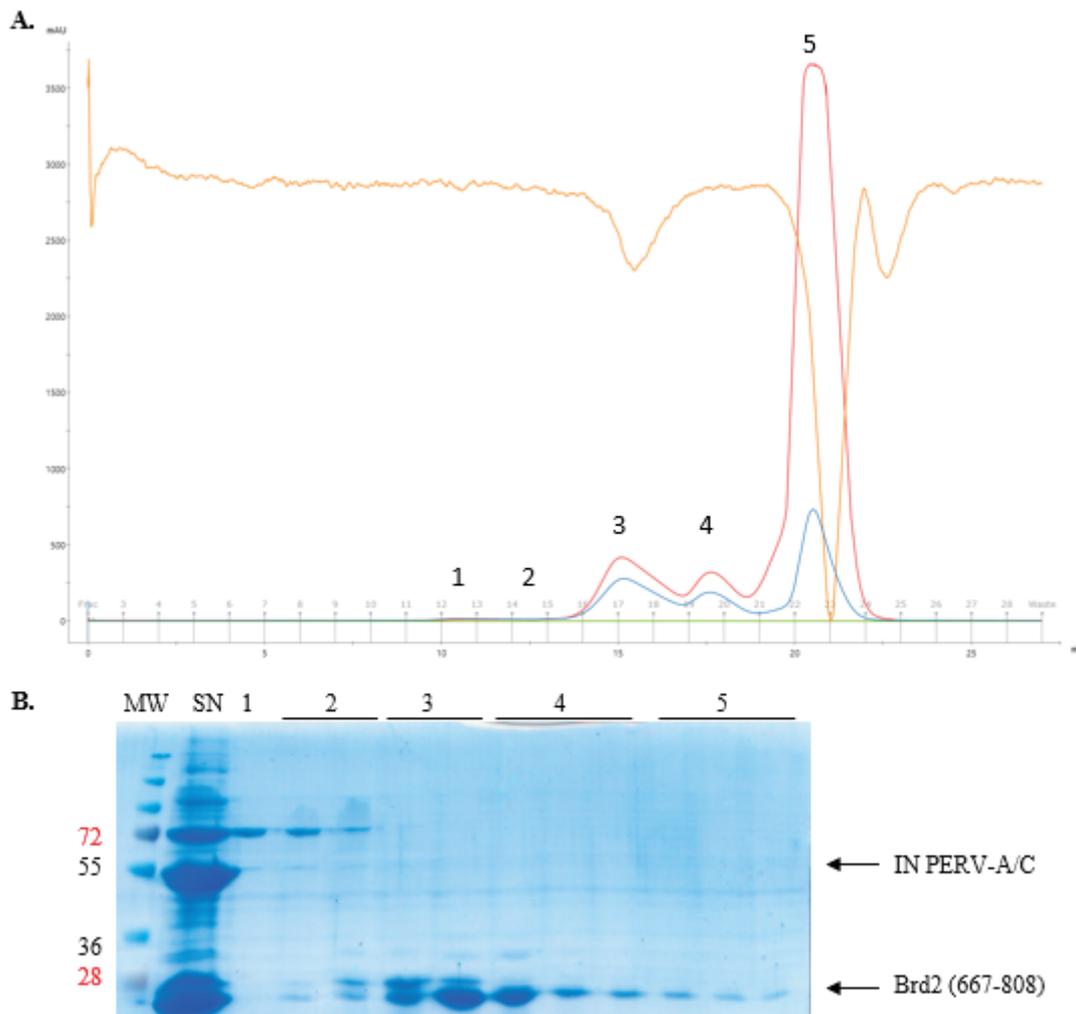


Figure 74 | Test d'assemblage de l'intasome : chromatographie d'exclusion stérique après incubation de l'IN de PERV-A/C et Brd2 (667-808) sur trois dialyses successives. **A.** Chromatogramme de purification, avec en bleu, l'absorbance à 280 nm ; en rouge, l'absorbance à 260 nm et en orange, la conductance. **B.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; piste SN, fraction soluble ; pistes 1 à 5, fractions d'élution.

Le profil laisse apparaître cinq pics d'élution distincts, dont les fractions ont été analysées par SDS-PAGE (**Figure 74B**). On peut tout d'abord observer une bande migrant au niveau du marqueur de poids moléculaire 72 kDa, qui correspond à un contaminant élué dans les fractions 1 et 2. La protéine Brd2 (667-808) est retrouvée dans chacune des fractions en quantités

variables. En revanche, bien qu'une bande correspondant à l'IN de PERV-A/C soit présente dans la fraction soluble injectée (piste SN), celle-ci n'est visible dans aucune des fractions d'élution. Ce protocole n'a donc pas été reproduit.

II. 3. 1. 3. Test 3

Le troisième test débute par la dilution de l'échantillon d'IN afin de pouvoir réaliser une chromatographie à échange d'ions (**Figure 75A**).

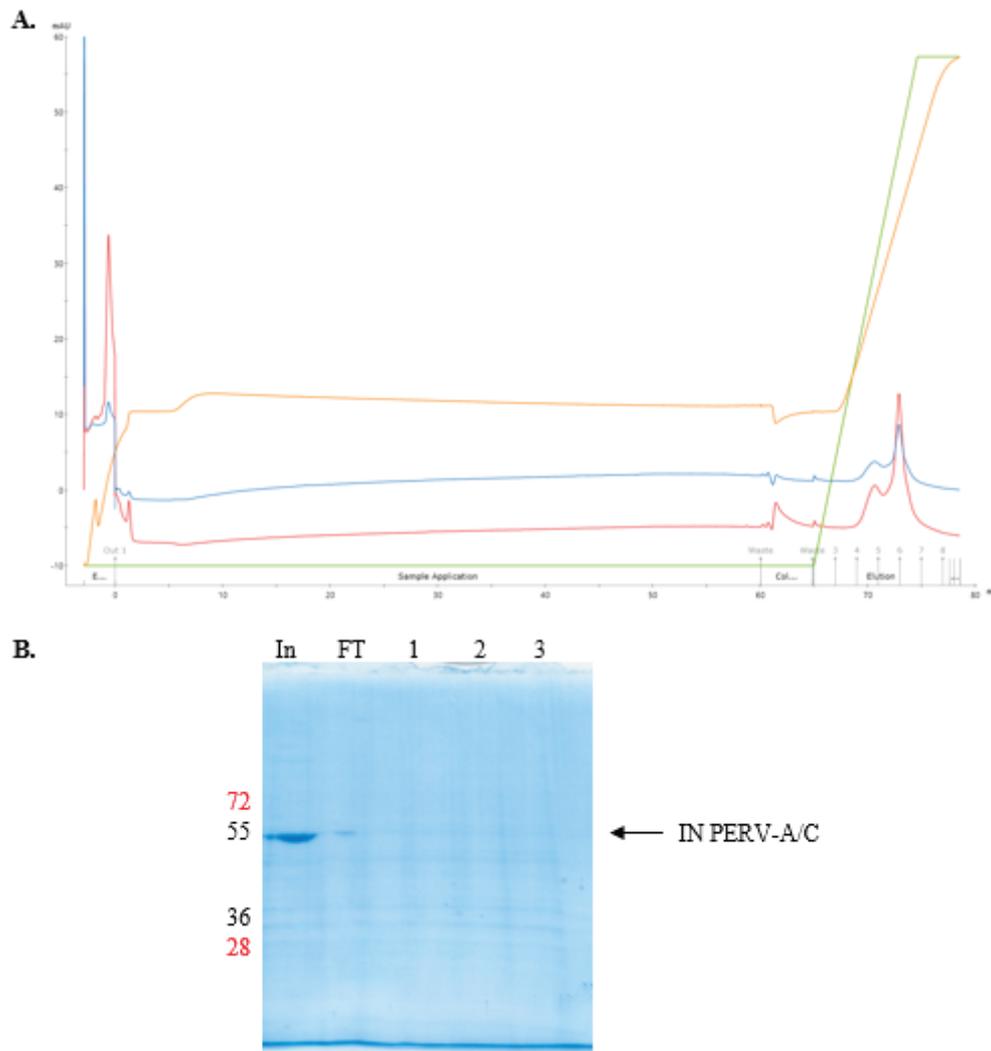


Figure 75 | Test d'assemblage de l'intasome : chromatographie à échange d'ions de l'IN de PERV-A/C. **A.** Chromatogramme de purification, avec en bleu, l'absorbance à 280 nm ; en rouge, l'absorbance à 260 nm ; en vert, le gradient d'imidazole et en orange, la conductance. **B.** Analyse par SDS-PAGE des fractions de purification. Piste In, fraction injectée sur la colonne ; piste FT, fraction non retenue et pistes 1 à 3, fractions d'élution.

Après analyse des fractions d'élution par SDS-PAGE (**Figure 75B**), on peut observer qu'aucune protéine n'a été éluée avec le gradient de sel. Effectivement, seule une bande de faible intensité peut être observée dans la fraction non retenue. L'intégralité des fractions non retenues a tout de même été concentrée et incubée avec le substrat d'ADN avant d'être analysée par chromatographie d'exclusion de taille (**Figure 76A**).

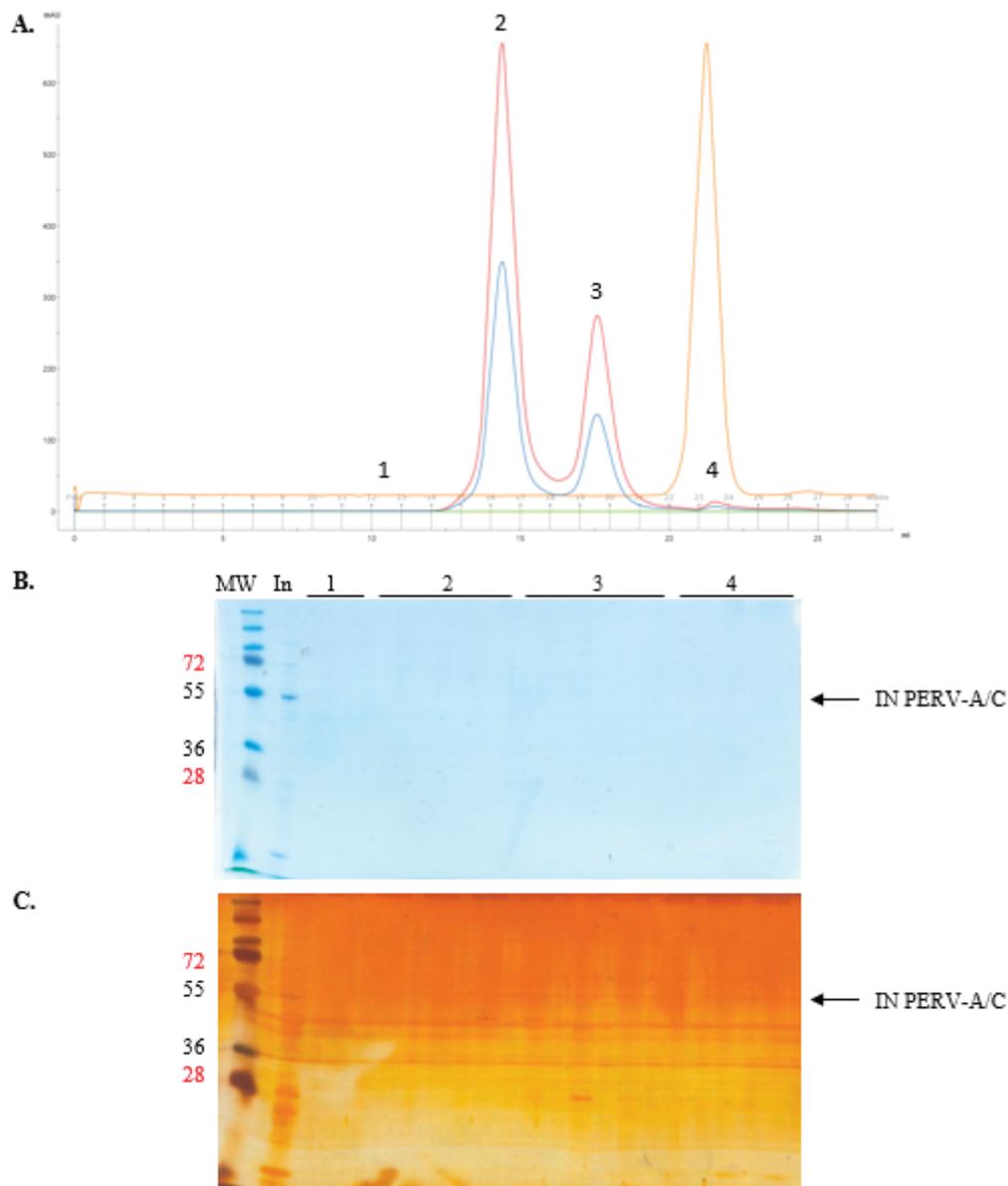


Figure 76 | Test d'assemblage de l'intasome : chromatographie d'exclusion stérique après incubation de l'IN de PERV-A/C avec l'ADN. **A.** Chromatogramme de purification, avec en bleu, l'absorbance à 280 nm ; en rouge, l'absorbance à 260 nm et en orange, la conductance. **B.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; piste In, fraction injectée sur la colonne ; pistes 1 à 4, fractions d'élution. Coloration au bleu de Coomassie et **C.** Coloration au nitrate d'argent.

Le chromatogramme obtenu montre la présence de quatre pics d'élution distincts qui ont été analysés par SDS-PAGE (**Figure 76B**). Bien qu'une bande correspondant à l'IN de PERV-A/C soit visible dans la fraction injectée sur la colonne, aucune bande n'est observable dans les différentes fractions d'élution. En raison de la faible concentration de l'échantillon, une coloration du gel au nitrate d'argent a également été réalisée, coloration ayant une plus forte sensibilité que celle au bleu de Coomassie (**Figure 76C**). Elle a permis la révélation de plusieurs bandes correspondant à des protéines de poids moléculaires inférieurs à celui attendu pour l'IN de PERV-A/C. La protéine d'intérêt n'a pas pu être purifiée à la suite de ce nouveau test.

II. 3. 1. 4. Test 4

Les tests suivants ont concerné la formation de l'intasome de PERV-A/C à deux pH distincts : 6,5 et 8,5. Dans les deux cas, l'IN de PERV-A/C a été purifiée suivant le protocole établi précédemment mais avec l'utilisation de tampons aux pH mentionnés. La protéine a ensuite été dialysée en présence du substrat d'ADN puis analysée par chromatographie d'affinité pour l'héparine.

Les chromatogrammes de purification obtenus pour ces deux essais sont similaires et seul celui représentant le test à pH 8,5 est exposé ci-après (**Figure 77A**). Ils montrent tous les deux la présence de deux pics d'élution de très faibles absorbances, traduisant la présence d'une quantité mineure de protéine dans l'échantillon.

Les fractions d'élution ont ensuite été analysées par SDS-PAGE (**Figure 76B**). Plusieurs bandes semblant correspondre à des contaminants protéiques sont visibles, autour de 72 kDa et 28 kDa. En revanche, une bande au poids moléculaire attendu pour l'IN de PERV-A/C est très faiblement visible dans les fractions non retenues et absente de chacune des fractions d'élution.

Ainsi, ces deux changements de pH n'ont pas semblé favoriser l'obtention d'une quantité d'IN suffisante pour la formation de l'intasome de PERV-A/C et la suite des essais sera à nouveau réalisée à pH 7,5 selon le protocole de purification de l'IN établi précédemment.

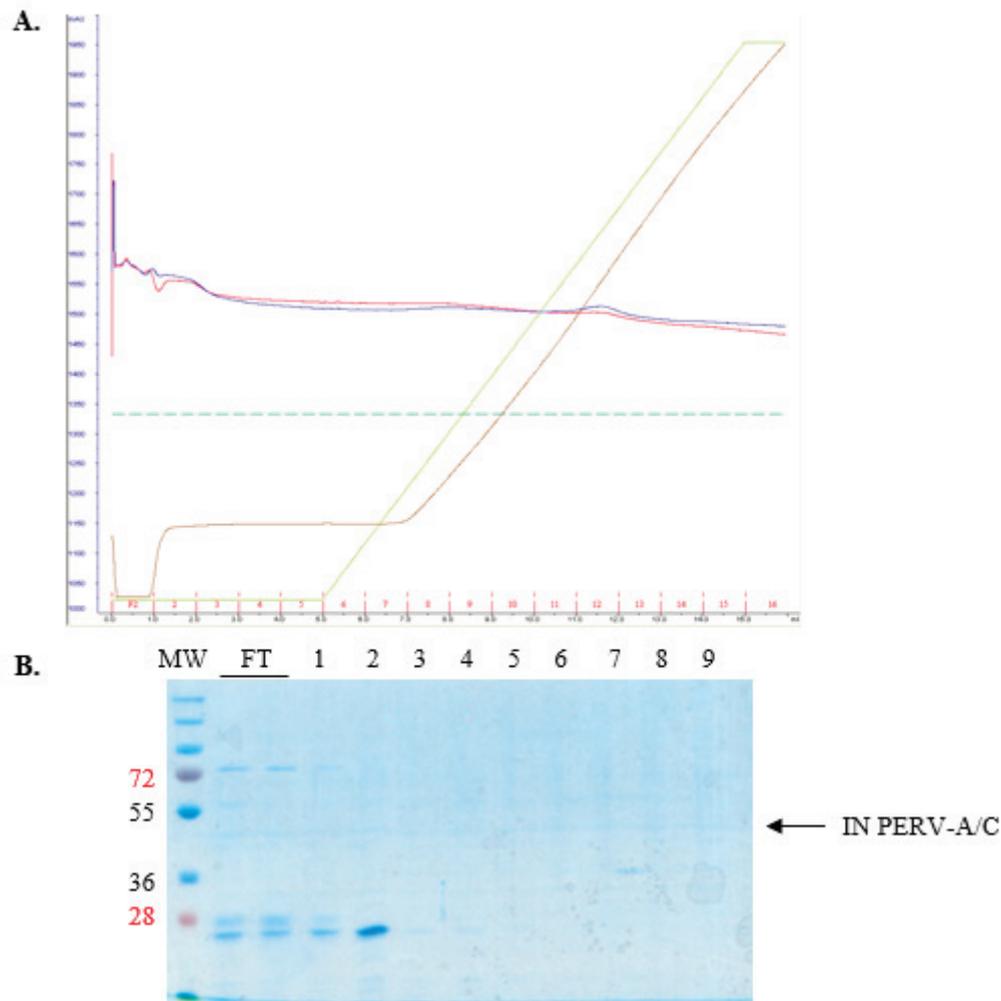


Figure 77 | Test d'assemblage de l'intasome à pH 8,5 : chromatographie d'affinité pour l'héparine après incubation de l'IN de PERV-A/C et de l'ADN. **A.** Chromatogramme de purification, avec en bleu, l'absorbance à 280 nm ; en rouge, l'absorbance à 260 nm ; en vert, le gradient de NaCl et en marron, la conductance. **B.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; pistes FT, fraction non retenue et pistes 1 à 9, fractions d'éluion.

II. 3. 1. 5. Test 5

La dernière série de tests réalisés pour la formation de l'intasome de PERV-A/C a débuté par la dialyse de l'IN avec la protéine Brd2 (667-808) dont l'étiquette de purification a préalablement été clivée. L'échantillon a ensuite été incubé avec le substrat d'ADN avant d'être analysé par trois techniques différentes, indépendantes les unes des autres.

Dans un premier temps, l'échantillon a été purifié par chromatographie d'affinité pour le nickel (**Figure 78**).

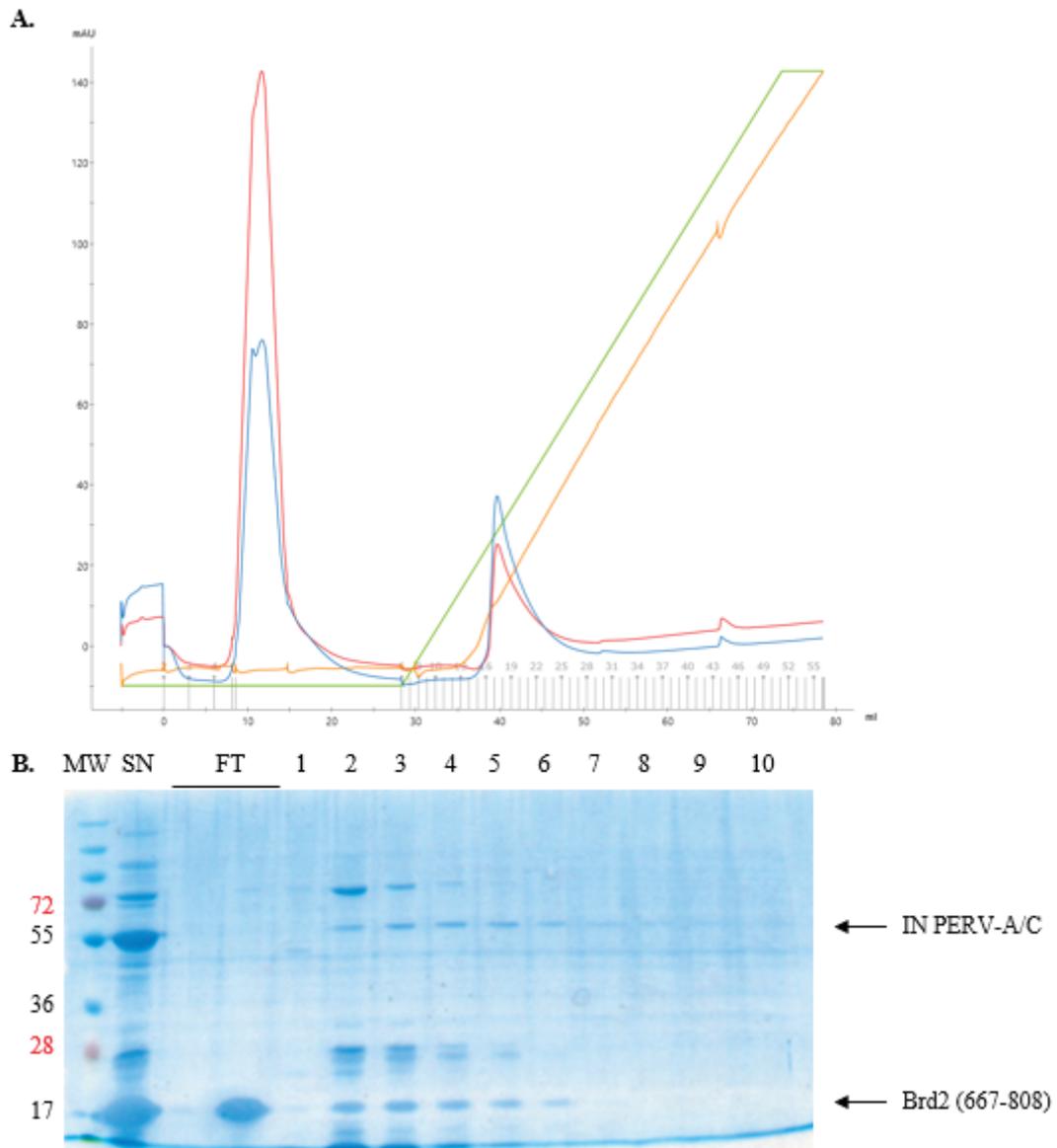


Figure 78 | Test d'assemblage de l'intasome : chromatographie d'affinité pour le nickel après incubation de l'IN de PERV-A/C avec Brd2 (667-808) et son substrat ADN. **A.** Chromatogramme de purification, avec en bleu, l'absorbance à 280 nm ; en rouge, l'absorbance à 260 nm ; en vert, le gradient d'imidazole et en orange, la conductance. **B.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; piste SN, fraction soluble ; piste FT, fraction non retenue et pistes 1 à 10, fractions d'élution.

Le chromatogramme de purification (**Figure 78A**) permet d'observer un pic d'élution principal, dont les différentes fractions ont été analysées par SDS-PAGE (**Figure 78B**). On peut dans un premier temps observer qu'une partie de Brd2 (667-808) est retrouvée dans la fraction non retenue, en raison de l'absence de son étiquette de purification. Cependant, des bandes correspondant à cette protéine sont également observables dans différentes fractions de purification. De plus, une bande correspondant à l'IN de PERV-A/C est visible dans ces mêmes

fractions, témoignant d'une possible interaction entre les deux protéines d'intérêt. En revanche, de nombreux contaminants protéiques sont également présents dans ces fractions.

La seconde approche utilisée a été la purification de l'échantillon par chromatographie d'exclusion de taille après concentration de celui-ci (**Figure 79**).

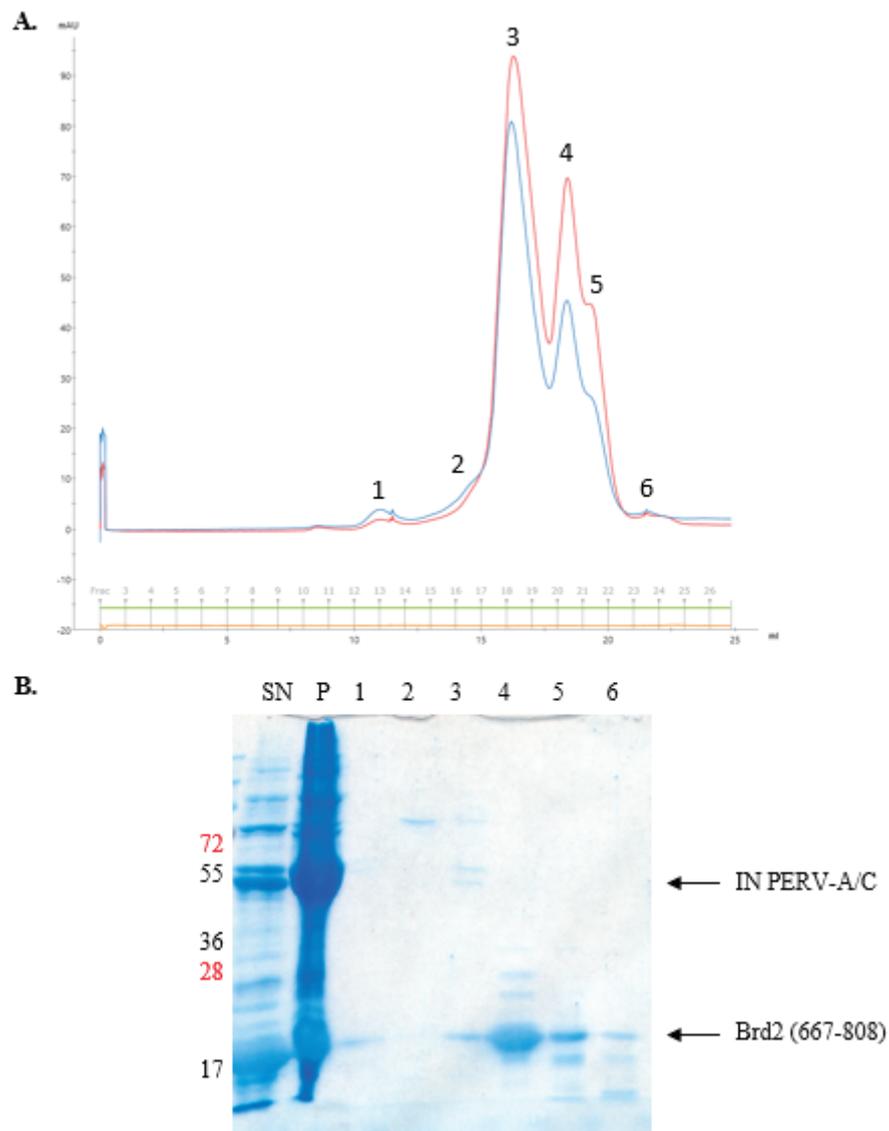


Figure 79 | Test d'assemblage de l'intasome : chromatographie d'exclusion stérique après incubation de l'IN de PERV-A/C avec Brd2 (667-808) et son substrat ADN. **A.** Chromatogramme de purification, avec en bleu, l'absorbance à 280 nm et en rouge, l'absorbance à 260 nm. **B.** Analyse par SDS-PAGE des fractions de purification. Piste SN, fraction soluble ; piste P, fraction précipitée et pistes 1 à 6, fractions d'éluion.

Six pics d'éluion sont observables sur le chromatogramme de purification (**Figure 79A**) et les différentes fractions d'éluion ont été analysées par SDS-PAGE (**Figure 79B**). Si une bande

correspondant à la protéine Brd2 (667-808) est bien visible pour certaines des fractions d'élution, ce n'est pas le cas de l'IN. En effet, celle-ci est observable dans l'une des fractions seulement (piste 3) et la majorité de la protéine semble avoir précipité suite à la concentration de l'échantillon (piste P).

La dernière approche a consisté à purifier l'échantillon par chromatographie d'affinité pour l'héparine.

Le chromatogramme obtenu laisse apparaître trois pics : le premier, de plus forte absorbance, correspondant à l'échantillon non retenu sur la colonne, et les deux suivant, d'absorbances plus faibles, correspondant aux protéines éluées par le gradient salin (**Figure 80A**).

Les différentes fractions d'élution ont été analysées par SDS-PAGE (**Figure 80B**). Ainsi, la fraction d'échantillon non retenu sur la colonne semble composée uniquement de contaminants protéiques. Le premier pic d'élution semble contenir uniquement la protéine Brd2 (667-808) tandis que le second pic contient les deux protéines d'intérêts Brd2 (667-808) et l'IN de PERV-A/C sans aucun contaminant.

C'est donc cet échantillon, obtenu en faible quantité (concentration de 0,2 mg/mL) qui servira par la suite à l'initiation de l'étude structurale de l'intasome de PERV-A/C, détaillée dans le chapitre suivant.

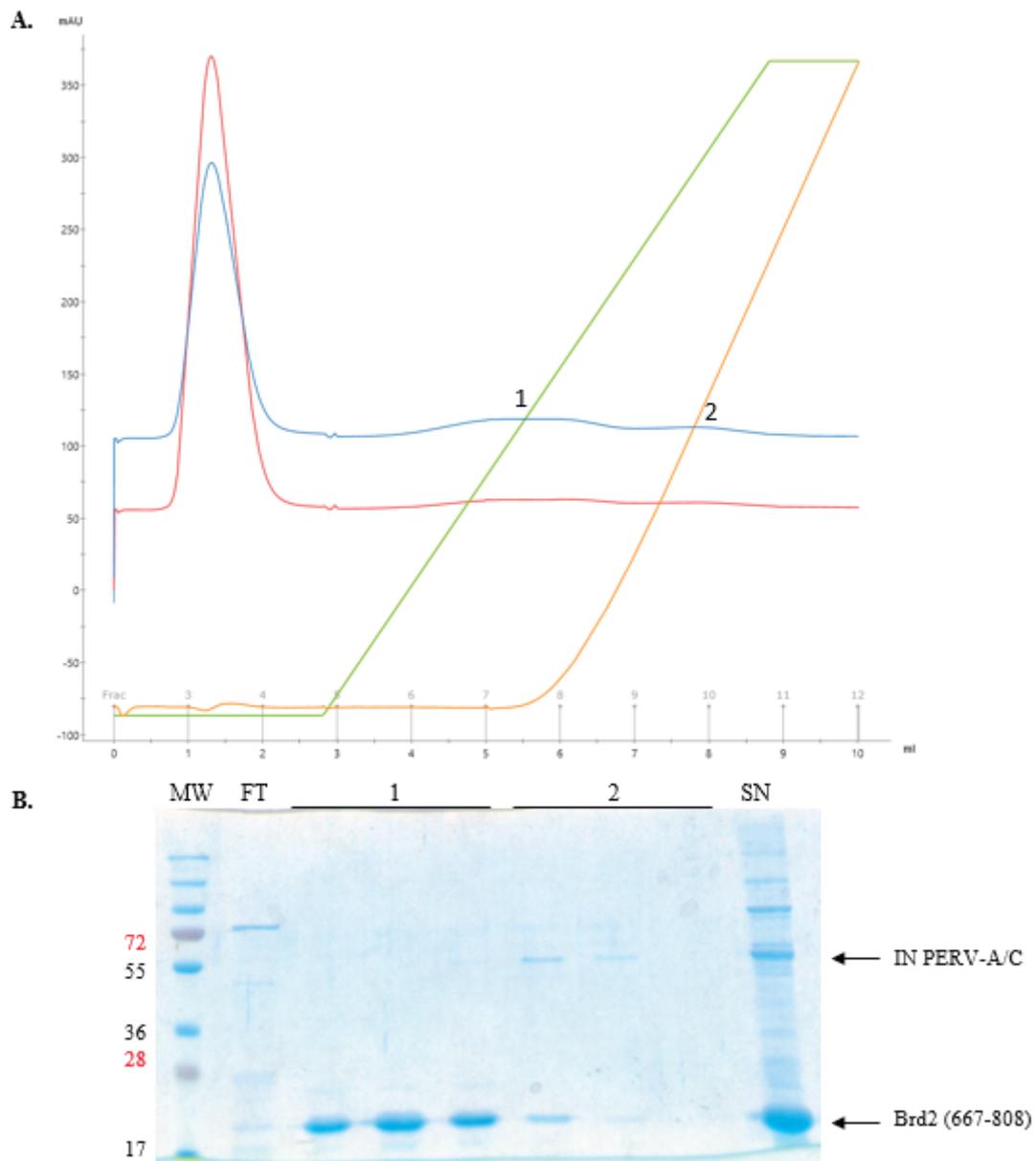


Figure 80 | Test d'assemblage de l'intasome : chromatographie d'affinité pour l'héparine après incubation de l'IN de PERV-A/C avec Brd2 (667-808) et son substrat ADN. **A.** Chromatogramme de purification, avec en bleu, l'absorbance à 280 nm ; en rouge, l'absorbance à 260 nm ; en vert, le gradient de sel (NaCl) et en orange, la conductance. **B.** Analyse par SDS-PAGE des fractions de purification. Piste MW, marqueurs de poids moléculaires ; piste FT, fraction non retenue ; piste SN, fraction soluble et pistes 1 et 2, fractions d'élution.

II. 3. 2. Étude structurale par microscopie électronique à transmission

En raison du poids moléculaire minimal attendu pour un intasome de PERV-A/C tétramérique (200 kDa), auquel s'ajoutent Brd2 (667-808) (18 kDa par monomère) et le substrat

d'ADN (26,4 kDa), la stratégie adoptée a été l'étude par microscopie électronique à transmission. En effet, la structure d'un échantillon de poids moléculaire supérieur à 200 kDa peut être obtenue par cette technique qui nécessite, en outre, un échantillon environ dix fois moins concentré que la cristallographie aux rayons X.

II. 3. 2. 1. Analyse par microscopie électronique à coloration négative

Les analyses par microscopie électronique à coloration négative ont été réalisées avec l'aide de la plateforme de Microscopie à l'Institut de Biologie Structurale de Grenoble. L'échantillon concentré à 0,2 mg/mL a été coloré avec du sodium silicotungstate avant observation et enregistrement des images (**Figure 81**).

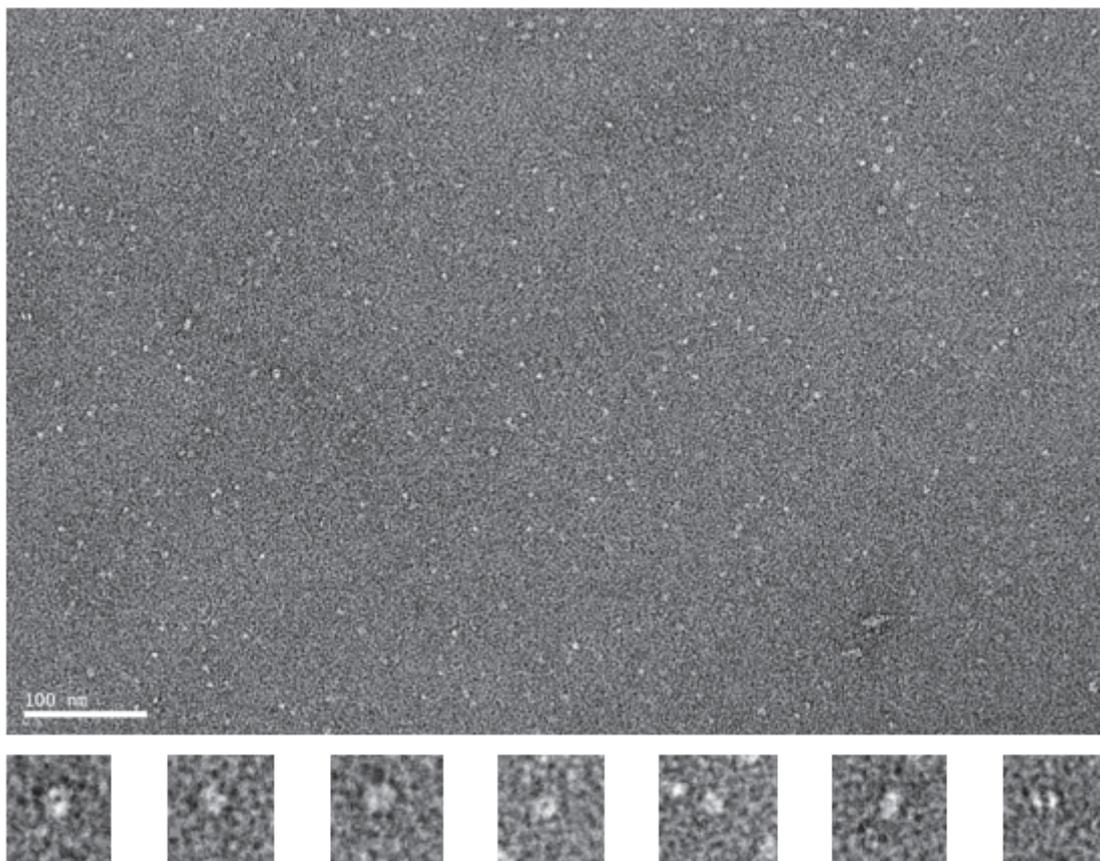


Figure 81 | Image obtenue par microscopie électronique à transmission en coloration négative pour l'intasome de PERV-A/C et grossissement sur différentes particules visibles. Images enregistrées par D. Fenel, IBS Grenoble.

L'analyse des images permet l'observation de nombreuses particules en forme d'anneaux d'une taille de 5 à 6 nm. Une tentative de calcul de classes 2D a été effectuée au laboratoire par le Dr. Robert, mais en raison du faible nombre de particules observées et de la résolution des

images, la reconstruction d'un modèle tridimensionnel, même à basse résolution, n'a pas été possible.

II. 3. 2. 2. Analyse par cryo-microscopie électronique

Nous avons bénéficié en juin 2019, pour les différents projets de l'équipe, de deux jours de temps d'enregistrement sur le cryo-microscope électronique 200 kV de l'Université de Milan. Ce cryo-TEM est couplé à une caméra à détection directe et nous en avons profité pour enregistrer des images de meilleure qualité sur notre échantillon d'intasome de PERV-A/C lié à Brd2 (667-808).

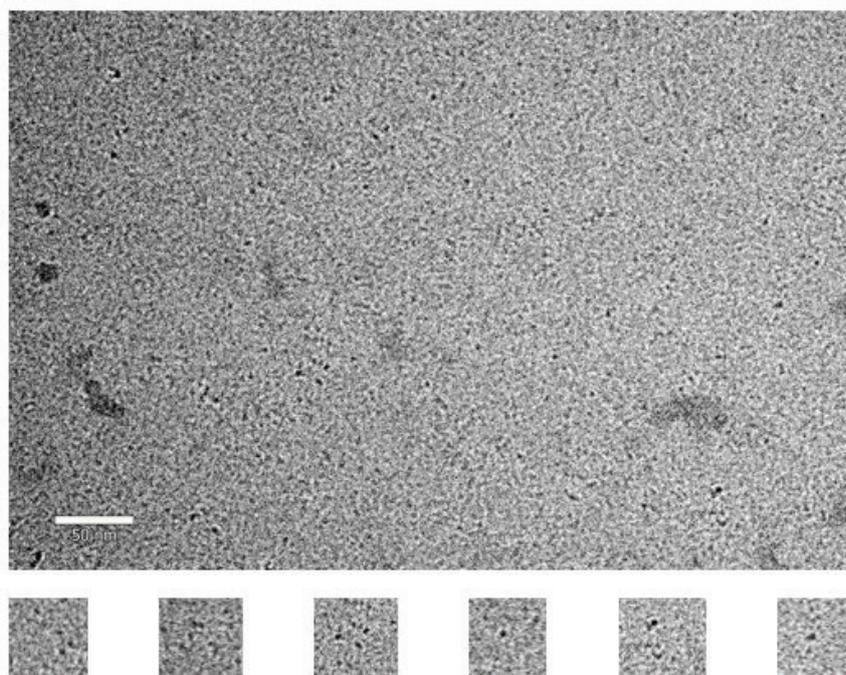


Figure 82 | Image obtenue par cryo-microscopie électronique pour l'intasome de PERV-A/C. Images enregistrées par le Dr. Paolo Swuec, Université de Milan.

De même que lors de l'analyse par microscopie électronique à coloration négative, quelques particules peuvent être observées sur ces images (**Figure 82**). Cependant, du fait de l'hétérogénéité de l'échantillon et du faible nombre de particules, il ne nous a pas été possible d'obtenir des classes 2D. Une optimisation du protocole permettant d'augmenter l'homogénéité de l'échantillon ainsi que sa concentration est donc nécessaire pour aller plus avant.

II. 4. Criblage de molécules compétitrices de l'intasome de PERV-A/C

Enfin, dans le but de développer des molécules inhibitrices de l'IN de PERV-A/C, la conception rationnelle d'inhibiteurs dirigés contre son interface hétérodimérique avec Brd2 (667-808) a été initiée.

II. 4. 1. Identification des molécules

Pour ce faire, nous avons réalisé une expérience de criblage virtuel *in silico* (ou *molecular docking*) de grande ampleur, bénéficiant d'un accès à 2 500 processeurs en parallèle au Centre de Calcul de l'IN2P3 de Villeurbanne. Cette puissance de calcul nous a autorisé le criblage complet de la branche '*Clean drug-like*' de la chimiothèque virtuelle ZINC (<https://zinc.docking.org>) (Sterling & Irwin, 2015) qui comprend plus de 15 millions de composés. Nous avons utilisé la structure du complexe entre le domaine ET de Brd4 humain en complexe avec le CTD de l'intégrase de MLV (code PDB : 2N3K) comme base structurale pour cibler la zone d'interaction EBM:ET. La forte homologie entre les CTD de l'IN-MLV et l'IN de PERV-A/C nous a permis de modéliser la zone d'intérêt avec grande confiance qui servira, ensuite, de cible dans l'expérience de *docking*.

Le criblage de la chimiothèque virtuelle a été conduit avec le logiciel Autodock VINA (Trott & Olson, 2010) contre la zone de Brd2 qui lie EBM. Un premier cycle de criblage a été lancé en considérant chaque ligand comme flexible, tandis que la structure protéique cible était également considérée comme rigide. Cette approximation tient au fait qu'introduire de la flexibilité dans la cible augmente drastiquement les temps de calcul et n'autorise pas un criblage aussi exhaustif dans des temps raisonnables.

À l'issue de cette première itération, le logiciel VINA a déterminé, pour chacun de 15 millions de composés, un score et une différence d'énergie libre de liaison ΔG . Ceci nous a permis d'interclasser les composés par ordre d'affinité prédite. Nous avons retenu les 1 340 molécules de meilleures affinités prédites que nous avons injectées dans un deuxième cycle de *docking*.

Ce dernier, plus exigeant en temps de calcul, a été conduit en considérant comme flexibles les chaînes latérales des résidus vraisemblablement impliqués dans l'interface hétérodimérique

de l'IN de PERV-A/C avec Brd2 (en se calquant sur le complexe IN-MLV/Brd4 humain, code PDB : 2N3K). Une fois encore, pour chacun des composés, un score et une différence d'énergie libre de liaison ΔG ont été calculés permettant de classer les molécules par affinités prédites.

En vue de réaliser une validation *in vitro* de cette expérience de criblage virtuel, six molécules parmi les 100 meilleures ont été choisies selon leur affinité calculée (K_D prédits de l'ordre du nanomolaire) et de leur coefficient de partition, témoignant de leur caractère hydrophile ou hydrophobe (**Figure 83**).

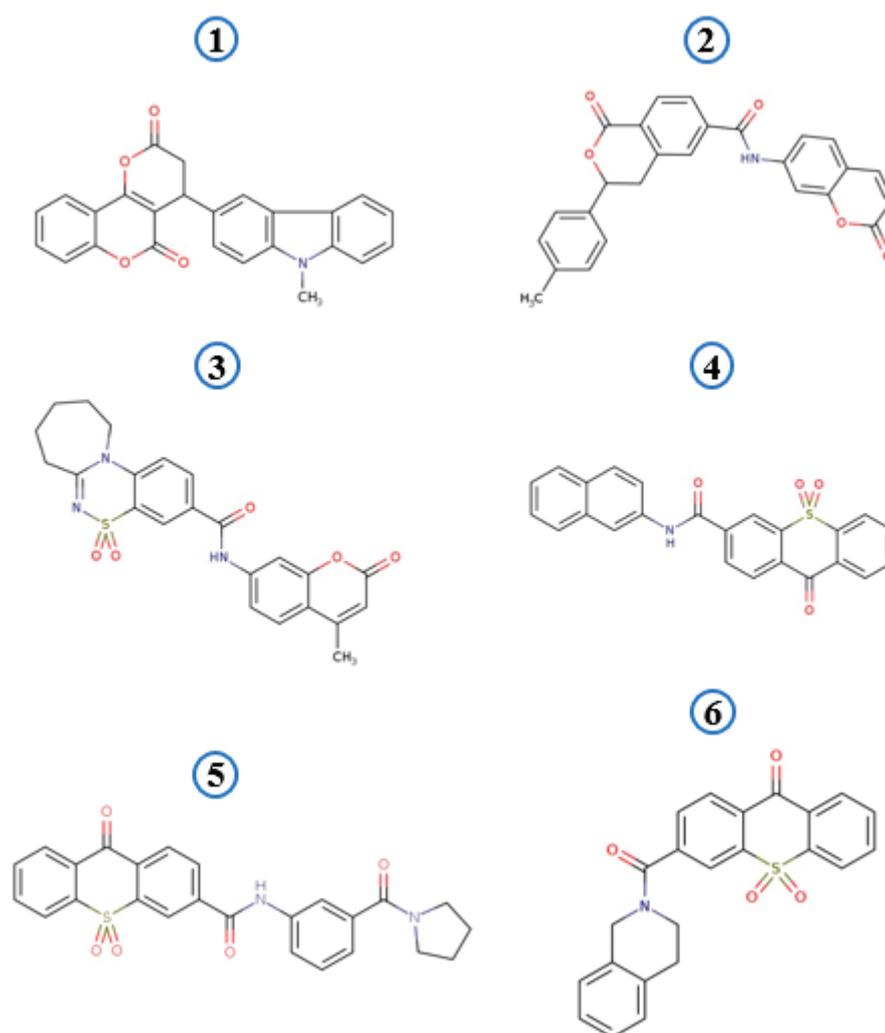


Figure 83 | Molécules potentiellement compétitrices de l'interaction de l'IN de PERV-A/C et Brd2.

Les différents composés sont disponibles sous forme de poudre. Du fait de leur faible solubilité dans l'eau, ils ont été repris dans du DMSO pur afin d'aboutir à une concentration de 50 mM. Leur concentration finale dans le milieu réactionnel étant de 1 μ M, la quantité de DMSO sera négligeable et ne perturbera pas le comportement des deux protéines.

Un peptide correspondant à la région EBM de l'IN de PERV-A/C et interagissant avec Brd2 (667-808) a également été synthétisé par la plateforme PSF de la SFR BioSciences de Gerland, afin de tester son effet compétiteur sur le complexe protéique.

II. 4. 2. Test des molécules par *pull down*

Dans un premier temps, l'interaction de chacune des molécules a été testée contre Brd2 (667-808) par criblage de thermostabilité des protéines par fluorescence (ou *Thermal Shift assay*). Cependant, cette méthode n'a pas semblé adaptée et aucun résultat n'a pu être obtenu.

Le potentiel effet compétiteur des six molécules identifiées et du peptide synthétique sur le complexe formé de l'IN de PERV-A/C et Brd2 (667-808) a ensuite été testé par méthode *pull down*. Le complexe a été formé à l'aide des deux protéines purifiées précédemment (**Figure 70** et **Figure 72**), puis incubé 16 heures avec l'une des molécules (voir **Matériels et méthodes §IV.2.**). Les fractions conservées avant l'ajout des molécules et les fractions d'élution ont été analysées par SDS-PAGE (**Figure 84**).

Dans un premier temps, les fractions conservées avant l'ajout des molécules ont été analysées (**Figure 84A**). Bien que des bandes représentant l'IN de PERV-A/C et Brd2 (667-808) soient visibles, des contaminants sont également présents et sont vraisemblablement apportés par l'échantillon d'IN car absents de la piste '-' correspondant à Brd2 seule. Cependant, ces contaminants n'ont pas d'influence sur la formation du complexe pour l'expérience.

Après incubation avec les différentes molécules, les protéines retenues par les billes magnétiques ont été analysées (**Figure 84B**). En théorie, les molécules interagissant avec Brd2 (667-808) devraient entraîner son élution lors des différents lavages, l'IN restant seule fixée sur les billes *via* son étiquette poly-histidine. Ici, dans chacun des cas testés, une partie de Brd2 (667-808) est retrouvée dans les fractions analysées. On pourrait en déduire l'inefficacité des molécules et du peptide pour la séparation du complexe, mais une bande correspondant à Brd2 est également observée dans la fraction du contrôle négatif, où seule cette protéine a été incubée. Il semblerait donc qu'il existe une affinité aspécifique de Brd2 (667-808) pour les billes

magnétiques utilisées. De nombreux autres tests ont été réalisés, faisant varier les concentrations ainsi que la résine utilisées mais, dans chacun des cas, tous les profils d'élution sont apparus identiques et Brd2 (667-808) a été retenue par la résine de façon non spécifique. Une optimisation supplémentaire ou l'utilisation de techniques différentes seront nécessaires pour évaluer ces molécules.

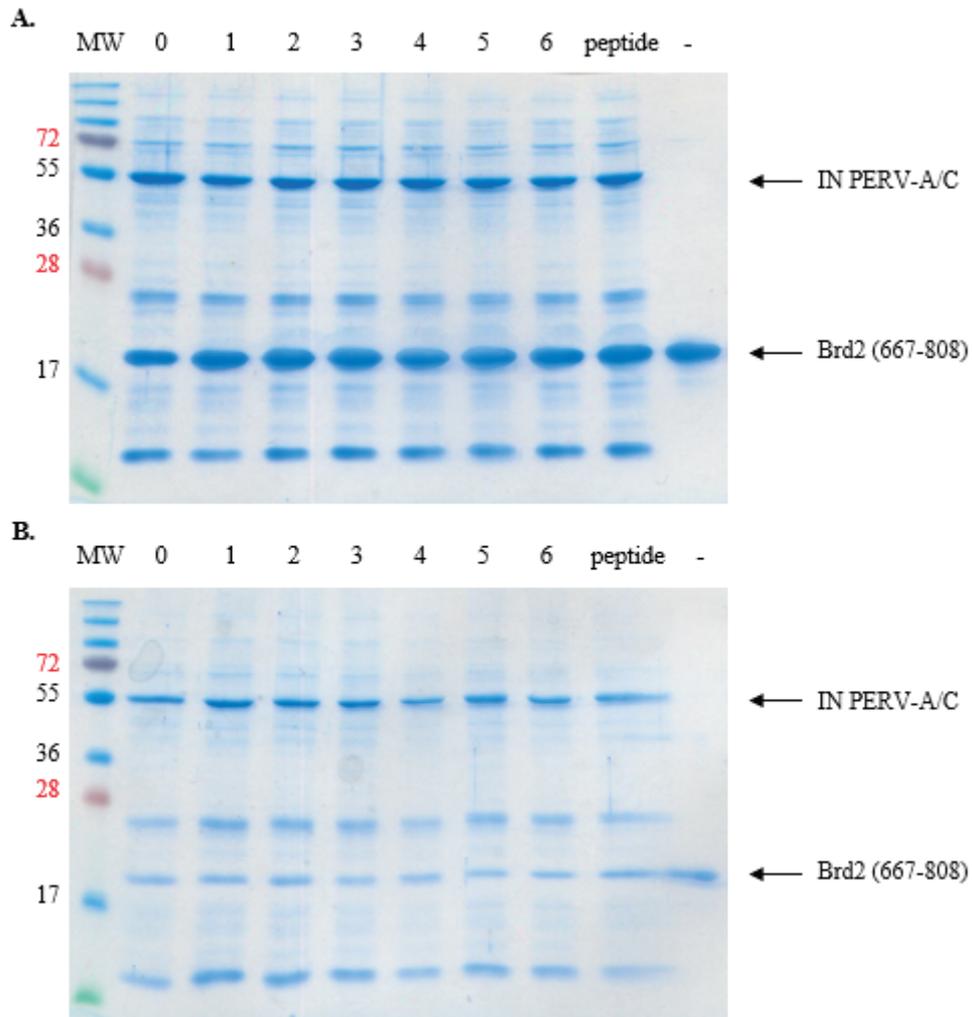


Figure 84 | Analyse SDS-PAGE de l'échantillon protéique (**A**) et des fractions éluées (**B**) suite à l'expérience de *pull down* en vue de tester les différentes molécules compétitrices. Piste MW, marqueurs de poids moléculaires ; piste 0, le contrôle sans molécule ; pistes 1 à 6, les tests des six molécules, piste 'peptide', test du peptide synthétisé et piste '-', contrôle négatif sans IN.

II. 5. Discussion

De manière générale, les bactéries *E. coli* sont utilisées en première intention pour la surexpression de protéines recombinantes en raison de la facilité de mise en place des protocoles et des hauts rendements de production constatés. Néanmoins, plusieurs souches d'*E. coli* ont été ici utilisées pour la surproduction de l'IN de PERV-A/C sans grand succès. Malgré les différents plasmides d'expression et tampon de lyse envisagés, la protéine n'a pas été exprimée ou a été exprimée sous forme insoluble. Afin de la solubiliser, plusieurs protocoles mettant en jeu des molécules telles que la rifampicine, des ions ou encore des détergents ont été essayés. Après de nombreux essais, l'IN a pu être solubilisée grâce à l'ajout d'urée 2 M.

Il a fallu ensuite élaborer un protocole permettant à la fois d'éliminer l'urée (pour que la protéine puisse se replier dans sa conformation native) et de former un intasome de PERV-A/C stable. L'ajout de Brd2 (667-808) a également été testé dans le but d'obtenir un complexe de plus grande taille en vue de son étude structurale par microscopie électronique à transmission. Les autres groupes de recherche ayant travaillé sur la surproduction d'intasome se sont heurtés à des difficultés similaires de précipitation, et il a été remarqué que seule une faible partie d'un échantillon d'IN a tendance à former un complexe stable avec son ADN. De plus, il a fallu trouver une concentration saline optimale pour la formation du complexe tout en assurant l'intégrité de l'IN, cette dernière nécessitant une concentration de NaCl élevée pour ne pas précipiter.

La majorité des nombreux tests réalisés n'a pas permis d'obtenir le complexe souhaité à l'exception d'un seul. La formation de l'intasome a pu être réalisée en deux temps, en incubant tout d'abord l'IN avec Brd2 (667-808) puis en ajoutant le substrat d'ADN, tout en éliminant graduellement l'urée par dialyse de façon à lentement restructurer l'IN. L'échantillon ensuite purifié par chromatographie d'affinité pour l'héparine a permis d'obtenir une petite quantité du complexe, en concentration suffisante pour son analyse par microscopie électronique à transmission.

Cette analyse structurale a été réalisée tout d'abord en coloration négative, ce qui a permis d'observer une petite quantité de particules en forme d'anneau. Toutefois, en raison de leur faible nombre et de problème de résolution, nous n'avons pu obtenir qu'un petit nombre de classes 2D. L'échantillon a ensuite été analysé par cryo-microscopie électronique afin d'améliorer la qualité des images enregistrées. Celles-ci ont cependant montré un échantillon

hétérogène, ne permettant pas de remonter à sa structure 3D. Il va donc être nécessaire d'optimiser le protocole d'obtention du complexe, afin d'améliorer l'homogénéité de l'échantillon et les rendements de purification.

Enfin, dans le but de concevoir des inhibiteurs de l'interaction de l'IN de PERV-A/C avec Brd2, plusieurs molécules ont été sélectionnées et testées sur le complexe protéique. Ces premiers tests n'ont cependant pas permis de mettre en évidence un effet des molécules sur la dissociation du complexe. Par ailleurs, les tests en *pull down* ont montré que Brd2 (667-808) avait une affinité non-désirée pour la résine, malgré l'absence de son étiquette poly-histidine. Une optimisation du protocole est nécessaire afin de limiter l'interaction aspécifique de Brd2 (667-808) et voir l'effet réel des différentes molécules sur le complexe. Il serait par exemple possible de réaliser le test à partir du complexe issu de la stratégie MultiTEV Minimal, où Brd2 (667-735) pourrait être retenue sur la colonne grâce à son étiquette poly-histidine, et le CTD de l'IN serait élué dans le cas d'un effet dissociatif causé par les molécules envisagées. Si cette optimisation n'aboutit pas à un résultat exploitable, il sera nécessaire d'envisager l'utilisation d'une technique de quantification d'interaction comme la thermophorèse.

CONCLUSION ET PERSPECTIVES

Mon travail doctoral a porté sur l'étude moléculaire et structurale de l'interaction entre l'IN du *Gammaretrovirus* PERV-A/C et son partenaire cellulaire, la protéine humaine Brd2. Cette étude a été menée dans le cadre de la prévention de zoonoses rétrovirales pouvant toucher les receveurs de greffons porcins porteurs du virus endogène PERV-A/C. L'objectif principal de ma thèse a donc été la caractérisation biochimique et l'obtention de données structurales du complexe protéique, afin de mieux comprendre le rôle de Brd2 dans le processus de réplication de PERV-A/C et de développer des inhibiteurs antirétroviraux spécifiques.

Dans un premier temps, mon étude a porté sur l'analyse biochimique et structurale du complexe formé par les domaines interagissant de l'IN de PERV-A/C et de Brd2. Pour ce faire, je me suis appuyée sur une nouvelle stratégie permettant la surproduction du complexe protéique d'intérêt à partir d'un unique plasmide d'expression. De cette manière, deux complexes ont été étudiés, tous deux formés du CTD de l'IN de PERV-A/C. Le premier, appelé MultiTEV, comprend en outre le fragment ET-SEED (667-808) de Brd2, alors que le deuxième, appelé MultiTEV Minimal, contient Brd2 réduit à son domaine ET (667-735).

Ces deux complexes ont pu être purifiés et caractérisés de façon à mettre en évidence l'interaction entre l'IN de PERV-A/C et Brd2 avec une affinité de l'ordre du micromolaire. Des expériences réalisées à partir de mutants disruptifs du complexe MultiTEV (W379A et R391A/L392A sur le CTD de l'IN) ont permis de confirmer l'importance de ces trois résidus pour l'interaction de l'IN avec Brd2. Le tryptophane est strictement conservé chez les IN des *Gammaretrovirus*, tandis que des résidus arginine et la leucine sont localisés au niveau d'un brin β intermoléculaire dans la structure RMN du complexe EBM MLV IN:ET Brd4 (Crowe *et al.*, 2016).

Une étude structurale par cristallographie aux rayons X a été initiée pour les deux complexes protéiques. Malgré les milliers de conditions de cristallisation testées, aucun de ces essais n'a mené à l'obtention de cristaux de protéines. Cependant, deux conditions ont permis d'obtenir de très petits cristaux du complexe Minimal, mais ceux-ci n'ont pas diffracté et des cristaux de taille supérieure n'ont pu être reproduits. Il sera nécessaire de surmonter ce blocage afin d'obtenir des cristaux de tailles suffisantes pour être analysés et, si ceux-ci sont des cristaux de protéines, résoudre la structure tridimensionnelle du complexe formé par le CTD de l'IN de PERV-A/C et Brd2 (667-735).

En revanche, l'étude structurale par SAXS m'a permis de calculer l'enveloppe du complexe formé par le CTD de l'IN de PERV-A/C et Brd2 (667-735). Un modèle atomique a pu être généré grâce à l'homologie entre nos protéines d'intérêt, le CTD de l'IN de MLV et le domaine ET de Brd4 (Crowe *et al.*, 2016). Après affinement en corps rigide contre les données SAXS, ce modèle a dévoilé la position du CTD de l'IN (309 à 372) de façon relative à son segment EBM (378-394) lié à l'ET de Brd2. Ce modèle SAXS présente une organisation en deux blocs structuraux formée d'un côté du cœur du CTD de l'IN, et de l'autre des domaines interagissant ET:EBM. La flexibilité observée entre ces deux blocs peut expliquer les difficultés rencontrées dans nos expériences de cristallogénèse.

Plusieurs informations ont également pu être déduites ou conjecturées d'après ce modèle SAXS. Ainsi, nous avons pu suggérer le rôle de la queue carboxy-terminale flexible de l'IN gammarétrovirale pour cibler le domaine ET de son cofacteur cellulaire, entraînant son repli sous forme d'un feuillet β intermoléculaire. Une explication des résultats obtenus avec les mutants disruptifs W379A et R391A/L392A lors des tests d'interaction *in vitro* a également pu être proposée à l'aide des détails atomiques déduits de la modélisation. Ainsi, il semble que le résidu W379 ait une place centrale dans l'interaction ET:EBM en stabilisant le feuillet β . Les résidus R391 et L392, quant à eux, appartiennent au motif -BHBH- du CTD, conservé chez les *Gammarrétrovirus* et essentiel pour la liaison avec le domaine ET chez d'autres partenaires des protéines BET.

Pour finir cette première partie de l'étude, un modèle de l'intasome de PERV-A/C lié au domaine ET de Brd2 a été généré à partir de la superposition de notre modèle SAXS avec la structure cristalline de l'intasome de PFV (Maertens *et al.*, 2010). Il semble ainsi possible que la flexibilité de la boucle reliant le cœur SH3 du CTD à l'extrémité EBM facilite l'interaction des deux partenaires et aide au ciblage efficace des sites d'intégration de PERV-A/C.

La deuxième partie de ce manuscrit a porté sur la caractérisation structurale et fonctionnelle de l'intasome de PERV-A/C, avec pour but d'obtenir des données structurales expérimentales suffisantes pour confirmer les théories proposées par notre modèle. Pour cela, il a été nécessaire de mettre au point des protocoles efficaces de surexpression et de purification de l'IN de PERV-A/C. De nombreuses conditions et de nombreux tampons ont été testés et une quantité d'IN suffisante pour la suite des essais a pu être obtenue en conditions dénaturantes seulement. Une deuxième série de tests a concerné le repliement natif de l'IN en présence d'un substrat d'ADN

(mimant l'ADN viral et l'ADN cellulaire) et de Brd2 (667-808) afin de stabiliser sa structure tout en permettant l'assemblage de l'intasome. Une petite quantité de protéines associées a pu être purifiée à la suite de ces essais et a servi à l'initiation d'une étude structurale de l'intasome de PERV-A/C. Cette étude structurale a été réalisée par microscopie électronique à coloration négative puis par cryo-microscopie électronique et, bien qu'un petit nombre de particules puisse être observé, il ne nous a pas été possible d'obtenir des classes 2D nécessaires au le calcul de l'enveloppe 3D du complexe nucléoprotéique. De plus, les images obtenues ont montré une hétérogénéité de l'échantillon. Il sera donc nécessaire d'optimiser le protocole de purification de l'intasome de PERV-A/C afin d'obtenir un échantillon plus homogène et plus concentré, optimal pour son analyse par microscopie électronique. Obtenir de telles données sur ce complexe permettrait d'élucider la première structure, jusqu'à présent inconnue, d'une intégrase de *Gammarrétrovirus*. Cela serait un premier pas majeur vers la caractérisation complète de l'assemblage de cette IN, la compréhension du rôle de son interaction avec Brd2 lors du ciblage des sites d'intégration, ainsi que pour expliciter ses différences avec les autres genres rétroviraux déjà connus.

Enfin, un premier criblage de molécules potentiellement compétitrices de l'interaction entre l'IN de PERV-A/C et Brd2 a été accompli et j'ai réalisé une première séries de tests avec quelques molécules candidates. Aucun résultat exploitable n'a pu être obtenu et une optimisation des conditions de tests va être nécessaire. D'autres techniques d'analyse *in vitro* de l'effet de ces molécules pourront également être envisagées. Il s'agit là des premières étapes vers la conception rationnelle d'inhibiteurs d'interaction protéine-protéine et si celles-ci sont prometteuses des essais *in cellulo* (tests de cytotoxicité et IC₅₀) pourront être réalisés. À long terme, le développement d'inhibiteurs visant une étape du cycle répliatif rétroviral jusqu'à présent peu exploitée permettra de réduire les risques de xénozooses lors de la transplantation de greffons porcins chez l'humain.

BIBLIOGRAPHIE

A

- Aiyer, S., Swapna, G.V., Malani, N., Aramini, J.M., Schneider, W.M., Plumb, M.R., Ghanem, M., Larue, R.C., Sharma, A., Studamire, B., Kvaratskhelia, M., Bushman, F.D., Montelione, G.T. & Roth, M.J. (2014) Altering murine leukemia virus integration through disruption of the integrase and BET protein family interaction. *Nucleic Acids Res*, **42**, 5917-5928.
- Ao, Z., Huang, G., Yao, H., Xu, Z., Labine, M., Cochrane, A.W. & Yao, X. (2007) Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication. *J Biol Chem*, **282**, 13456-13467.

B

- Ballandras-Colas, A., Brown, M., Cook, N.J., Dewdney, T.G., Demeler, B., Cherepanov, P., Lyumkis, D. & Engelman, A.N. (2016) Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature*, **530**, 358-361.
- Ballandras-Colas, A., Maskell, D.P., Serrao, E., Locke, J., Swuec, P., Jonsson, S.R., Kotecha, A., Cook, N.J., Pye, V.E., Taylor, I.A., Andresdottir, V., Engelman, A.N., Costa, A. & Cherepanov, P. (2017) A supramolecular assembly mediates lentiviral DNA integration. *Science*, **355**, 93-95.
- Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, **226**, 1209-1211.
- Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. & Montagnier, L. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, **220**, 868-871.
- 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.
- Belkina, A.C. & Denis, G.V. (2012) BET domain co-regulators in obesity, inflammation and cancer. *Nat Rev Cancer*, **12**, 465-477.
- Birnboim, H.C. & Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res*, **7**, 1513-1523.
- Bittner, J.J. (1936) Some possible effects of nursing on the mammary gland tumor incidence in mice. *Science*, **84**, 162.
- Brown, R. & Bennett, J.J. (1866) *The Miscellaneous Botanical Works of Robert Brown*. London, Published for the Ray society by R. Hardwicke, 1866-68.
- Bushman, F.D. & Wang, B. (1994) Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding. *J Virol*, **68**, 2215-2223.
- Busschots, K., Vercammen, J., Emiliani, S., Benarous, R., Engelborghs, Y., Christ, F. & Debyser, Z. (2005) The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J Biol Chem*, **280**, 17841-17847.

C

- Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G.M. & Gronenborn, A.M. (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat Struct Biol*, **4**, 567-577.
- Carrington, J.C. & Dougherty, W.G. (1988) A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proc Natl Acad Sci U S A*, **85**, 3391-3395.
- Chen, J.C.H., Krucinski, J., Miercke, L.J.W., Finer-Moore, J.S., Tang, A.H., Leavitt, A.D. & Stroud, R.M. (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 8233-8238.
- Cheng, Y. (2018) Single-particle cryo-EM-How did it get here and where will it go. *Science*, **361**, 876-880.
- Cherepanov, P., Devroe, E., Silver, P.A. & Engelman, A. (2004) Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J Biol Chem*, **279**, 48883-48892.
- Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R. & Bushman, F. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med*, **11**, 1287-1289.
- Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M.A., Santos-Ferreira, M.O., Laurent, A.G., Dauguet, C., Katlama, C., Rouzioux, C. & et al. (1986) Isolation of a new human retrovirus from West African patients with AIDS. *Science*, **233**, 343-346.
- Cooper, D.K.C., Gaston, R., Eckhoff, D., Ladowski, J., Yamamoto, T., Wang, L., Iwase, H., Hara, H., Tector, M. & Tector, A.J. (2018) Xenotransplantation-the current status and prospects. *Br Med Bull*, **125**, 5-14.
- Crowe, B.L., Larue, R.C., Yuan, C., Hess, S., Kvaratskhelia, M. & Foster, M.P. (2016) 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*, **113**, 2086-2091.
- Cullen, B.R. (1992) Mechanism of action of regulatory proteins encoded by complex retroviruses. *Microbiol Rev*, **56**, 375-394.

D

- De Rijck, J., de Kogel, C., Demeulemeester, J., Vets, S., El Ashkar, S., Malani, N., Bushman, F.D., Landuyt, B., Husson, S.J., Busschots, K., Gijsbers, R. & Debyser, Z. (2013) The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Rep*, **5**, 886-894.
- Debyser, Z., Christ, F., De Rijck, J. & Gijsbers, R. (2015) Host factors for retroviral integration site selection. *Trends Biochem Sci*, **40**, 108-116.
- Demange, A., Yajjou-Hamalian, H., Gallay, K., Luengo, C., Beven, V., Leroux, A., Confort, M.P., Al Andary, E., Gouet, P., Moreau, K., Ronfort, C. & Blanchard, Y. (2015) Porcine endogenous retrovirus-A/C: biochemical properties of its integrase and susceptibility to raltegravir. *J Gen Virol*, **96**, 3124-3130.

- Denis, G.V., McComb, M.E., Faller, D.V., Sinha, A., Romesser, P.B. & Costello, C.E. (2006) Identification of transcription complexes that contain the double bromodomain protein Brd2 and chromatin remodeling machines. *J Proteome Res*, **5**, 502-511.
- Denner, J. (2008) Recombinant porcine endogenous retroviruses (PERV-A/C): a new risk for xenotransplantation? *Arch Virol*, **153**, 1421-1426.
- Denner, J. (2016) How Active Are Porcine Endogenous Retroviruses (PERVs)? *Viruses*, **8**, 215.
- Dewannieux, M., Ribet, D. & Heidmann, T. (2010) Risks linked to endogenous retroviruses for vaccine production: a general overview. *Biologicals*, **38**, 366-370.
- Dey, A., Chitsaz, F., Abbasi, A., Misteli, T. & Ozato, K. (2003) The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. *Proc Natl Acad Sci U S A*, **100**, 8758-8763.
- Dobard, C.W., Briones, M.S. & Chow, S.A. (2007) Molecular mechanisms by which human immunodeficiency virus type 1 integrase stimulates the early steps of reverse transcription. *J Virol*, **81**, 10037-10046.
- Dumon-Seignovert, L., Cariot, G. & Vuillard, L. (2004) The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr Purif*, **37**, 203-206.

E

- Eidahl, J.O., Crowe, B.L., North, J.A., McKee, C.J., Shkriabai, N., Feng, L., Plumb, M., Graham, R.L., Gorelick, R.J., Hess, S., Poirier, M.G., Foster, M.P. & Kvaratskhelia, M. (2013) Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Res*, **41**, 3924-3936.
- Eijkelenboom, A.P., van den Ent, F.M., Vos, A., Doreleijers, J.F., Hard, K., Tullius, T.D., Plasterk, R.H., Kaptein, R. & Boelens, R. (1997) The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. *Curr Biol*, **7**, 739-746.
- El Ashkar, S., De Rijck, J., Demeulemeester, J., Vets, S., Madlala, P., Cermakova, K., Debyser, Z. & Gijssbers, R. (2014) BET-independent MLV-based Vectors Target Away From Promoters and Regulatory Elements. *Mol Ther Nucleic Acids*, **3**, e179.
- Ellermann, V. & Bang, O. (1908) Experimentelle Leukämie bei Hühnern. *Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt. Orig.*, **46**, 595-609.
- Engelman, A. & Cherepanov, P. (2008) The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. *PLoS Pathog*, **4**, e1000046.
- Engelman, A. (2011) Pleiotropic nature of HIV-1 integrase mutations. In Neamati, N. (ed.), *HIV-1 Integrase: Mechanism and Inhibitor Design*. John Wiley and Sons, Inc., New Jersey, pp. 67-81.
- Engelman, A., Kessl, J.J. & Kvaratskhelia, M. (2013) Allosteric inhibition of HIV-1 integrase activity. *Curr Opin Chem Biol*, **17**, 339-345.
- Engelman, A.N. & Cherepanov, P. (2017) Retroviral intasomes arising. *Curr Opin Struct Biol*, **47**, 23-29.

Engelman, A.N. & Singh, P.K. (2018) Cellular and molecular mechanisms of HIV-1 integration targeting. *Cell Mol Life Sci*, **75**, 2491-2507.

F

Faust, T.B., Binning, J.M., Gross, J.D. & Frankel, A.D. (2017) Making Sense of Multifunctional Proteins: Human Immunodeficiency Virus Type 1 Accessory and Regulatory Proteins and Connections to Transcription. *Annu Rev Virol*, **4**, 241-260.

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., Philpott, M., Munro, S., McKeown, M.R., Wang, Y., Christie, A.L., West, N., Cameron, M.J., Schwartz, B., Heightman, T.D., La Thangue, N., French, C.A., Wiest, O., Kung, A.L., Knapp, S. & Bradner, J.E. (2010) Selective inhibition of BET bromodomains. *Nature*, **468**, 1067-1073.

Fiorini, F., Stefanini, S., Valenti, P., Chiancone, E. & De Biase, D. (2008) Transcription of the *Listeria monocytogenes* *fri* gene is growth-phase dependent and is repressed directly by Fur, the ferric uptake regulator. *Gene*, **410**, 113-121.

Fontana, J., Jurado, K.A., Cheng, N., Ly, N.L., Fuchs, J.R., Gorelick, R.J., Engelman, A.N. & Steven, A.C. (2015) Distribution and Redistribution of HIV-1 Nucleocapsid Protein in Immature, Mature, and Integrase-Inhibited Virions: a Role for Integrase in Maturation. *J Virol*, **89**, 9765-9780.

Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M. & Svergun, D.I. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr*, **50**, 1212-1225.

Freed, E.O. (1998) HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology*, **251**, 1-15.

Freed, E.O. (2001) HIV-1 replication. *Somat Cell Mol Genet*, **26**, 13-33.

Freed, E.O. (2015) HIV-1 assembly, release and maturation. *Nat Rev Micro*, **13**, 484-496.

G

Gallay, K., Blot, G., Chahpazoff, M., Yajjou-Hamalian, H., Confort, M.P., De Boisseson, C., Leroux, A., Luengo, C., Fiorini, F., Lavigne, M., Chebloune, Y., Gouet, P., Moreau, K., Blanchard, Y. & Ronfort, C. (2019) In vitro, in cellulo and structural characterizations of the interaction between the integrase of Porcine Endogenous Retrovirus A/C and proteins of the BET family. *Virology*, **532**, 69-81.

Gallo, R.C., Salahuddin, S.Z., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Redfield, R., Oleske, J., Safai, B. & et al. (1984) Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science*, **224**, 500-503.

Gaynor, R.B. (1995) Regulation of HIV-1 gene expression by the transactivator protein Tat. *Curr Top Microbiol Immunol*, **193**, 51-77.

Ge, H., Si, Y. & Roeder, R.G. (1998) Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *Embo j*, **17**, 6723-6729.

- Gessain, A. & Mahieux, R. (2012) Tropical spastic paraparesis and HTLV-1 associated myelopathy: clinical, epidemiological, virological and therapeutic aspects. *Rev Neurol (Paris)*, **168**, 257-269.
- Grandgenett, D.P., Pandey, K.K., Bera, S. & Aihara, H. (2015) Multifunctional facets of retrovirus integrase. *World J Biol Chem*, **6**, 83-94.
- Gross, L. (1957) Development and serial cellfree passage of a highly potent strain of mouse leukemia virus. *Proc Soc Exp Biol Med*, **94**, 767-771.
- Gupta, K., Curtis, Joseph E., Krueger, S., Hwang, Y., Cherepanov, P., Bushman, Frederic D. & Van Duyne, Gregory D. (2012) Solution Conformations of Prototype Foamy Virus Integrase and Its Stable Synaptic Complex with U5 Viral DNA. *Structure*, **20**, 1918-1928.
- Gupta, K., Brady, T., Dyer, B.M., Malani, N., Hwang, Y., Male, F., Nolte, R.T., Wang, L., Velthuisen, E., Jeffrey, J., Van Duyne, G.D. & Bushman, F.D. (2014) Allosteric inhibition of human immunodeficiency virus integrase: late block during viral replication and abnormal multimerization involving specific protein domains. *J Biol Chem*, **289**, 20477-20488.
- Gupta, K., Turkki, V., Sherrill-Mix, S., Hwang, Y., Eilers, G., Taylor, L., McDanal, C., Wang, P., Temelkoff, D., Nolte, R.T., Velthuisen, E., Jeffrey, J., Van Duyne, G.D. & Bushman, F.D. (2016) Structural Basis for Inhibitor-Induced Aggregation of HIV Integrase. *PLoS Biol*, **14**, e1002584.
- Gupta, S.S., Maetzig, T., Maertens, G.N., Sharif, A., Rothe, M., Weidner-Glunde, M., Galla, M., Schambach, A., Cherepanov, P. & Schulz, T.F. (2013) Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J Virol*, **87**, 12721-12736.

H

- Hare, S., Shun, M.C., Gupta, S.S., Valkov, E., Engelman, A. & Cherepanov, P. (2009) A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75. *PLoS Pathog*, **5**, e1000259.
- Hare, S., Gupta, S.S., Valkov, E., Engelman, A. & Cherepanov, P. (2010) Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature*, **464**, 232-236.
- Hare, S., Smith, S.J., Metifiot, M., Jaxa-Chamiec, A., Pommier, Y., Hughes, S.H. & Cherepanov, P. (2011) Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572). *Mol Pharmacol*, **80**, 565-572.
- Hazuda, D.J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J.A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C. & Miller, M.D. (2000) Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science*, **287**, 646-650.
- Hightower, K.E., Wang, R., Deanda, F., Johns, B.A., Weaver, K., Shen, Y., Tomberlin, G.H., Carter, H.L., 3rd, Broderick, T., Sigethy, S., Seki, T., Kobayashi, M. & Underwood, M.R. (2011) Dolutegravir (S/GSK1349572) exhibits significantly slower dissociation than raltegravir and elvitegravir from wild-type and integrase inhibitor-resistant HIV-1 integrase-DNA complexes. *Antimicrob Agents Chemother*, **55**, 4552-4559.

I

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

J

Jacks, T. & Varmus, H.E. (1985) Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science*, **230**, 1237-1242.

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

Jacobson, J.M. & Flexner, C.W. (2017) Universal antiretroviral regimens: thinking beyond one-pill-once-a-day. *Curr Opin HIV AIDS*, **12**, 343-350.

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.

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

Johns, B.A., Kawasuji, T., Weatherhead, J.G., Taishi, T., Temelkoff, D.P., Yoshida, H., Akiyama, T., Taoda, Y., Murai, H., Kiyama, R., Fuji, M., Tanimoto, N., Jeffrey, J., Foster, S.A., Yoshinaga, T., Seki, T., Kobayashi, M., Sato, A., Johnson, M.N., Garvey, E.P. & Fujiwara, T. (2013) Carbamoyl pyridone HIV-1 integrase inhibitors 3. A diastereomeric approach to chiral nonracemic tricyclic ring systems and the discovery of dolutegravir (S/GSK1349572) and (S/GSK1265744). *J Med Chem*, **56**, 5901-5916.

K

Kalpana, G.V., Marmon, S., Wang, W., Crabtree, G.R. & Goff, S.P. (1994) Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science*, **266**, 2002-2006.

Kalyanaraman, V.S., Sarngadharan, M.G., Robert-Guroff, M., Miyoshi, I., Golde, D. & Gallo, R.C. (1982) A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. *Science*, **218**, 571-573.

Kanno, T., Kanno, Y., Siegel, R.M., Jang, M.K., Lenardo, M.J. & Ozato, K. (2004) Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. *Mol Cell*, **13**, 33-43.

Kessl, J.J., Kutluay, S.B., Townsend, D., Rebensburg, S., Slaughter, A., Larue, R.C., Shkriabai, N., Bakouche, N., Fuchs, J.R., Bieniasz, P.D. & Kvaratskhelia, M. (2016) HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. *Cell*, **166**, 1257-1268.e1212.

Kim, S., Rusmevichientong, A., Dong, B., Remenyi, R., Silverman, R.H. & Chow, S.A. (2010) Fidelity of target site duplication and sequence preference during integration of xenotropic murine leukemia virus-related virus. *PLoS One*, **5**, e10255.

- King, A., Efkowitz, E., Adams, J. & Carstens, E. (2012) *Virus Taxonomy: Classification and Nomenclature of Viruses : Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier.
- Kirsch, R.D. & Joly, E. (1998) An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res*, **26**, 1848-1850.
- Koneru, P.C., Francis, A.C., Deng, N., Rebensburg, S.V., Hoyte, A.C., Lindenberger, J., Adu-Ampratwum, D., Larue, R.C., Wempe, M.F., Engelman, A.N., Lyumkis, D., Fuchs, J.R., Levy, R.M., Melikyan, G.B. & Kvaratskhelia, M. (2019) HIV-1 integrase tetramers are the antiviral target of pyridine-based allosteric integrase inhibitors. *Elife*, **8**.
- Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P. & Skalka, A.M. (1992) 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*, **12**, 2331-2338.
- Kupitz, C., Grotjohann, I., Conrad, C.E., Roy-Chowdhury, S., Fromme, R. & Fromme, P. (2014) Microcrystallization techniques for serial femtosecond crystallography using photosystem II from *Thermosynechococcus elongatus* as a model system. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **369**.
- 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.

L

- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.
- Lesbats, P., Botbol, Y., Chevereau, G., Vaillant, C., Calmels, C., Arneodo, A., Andreola, M.L., Lavigne, M. & Parissi, V. (2011) Functional coupling between HIV-1 integrase and the SWI/SNF chromatin remodeling complex for efficient in vitro integration into stable nucleosomes. *PLoS Pathog*, **7**, e1001280.
- Lesbats, P., Engelman, A.N. & Cherepanov, P. (2016) Retroviral DNA Integration. *Chemical Reviews*, **116**, 12730-12757.
- Levy, N., Eiler, S., Pradeau-Aubret, K., Maillot, B., Stricher, F. & Ruff, M. (2016) Production of unstable proteins through the formation of stable core complexes. *Nature Communications*, **7**, 10932.
- Li, M., Jurado, K.A., Lin, S., Engelman, A. & Craigie, R. (2014) Engineered hyperactive integrase for concerted HIV-1 DNA integration. *PLoS One*, **9**, e105078.
- Lin, Y.J., Umehara, T., Inoue, M., Saito, K., Kigawa, T., Jang, M.K., Ozato, K., Yokoyama, S., Padmanabhan, B. & Guntert, P. (2008) Solution structure of the extraterminal domain of the bromodomain-containing protein BRD4. *Protein Sci*, **17**, 2174-2179.
- Linding, R., Russell, R.B., Neduva, V. & Gibson, T.J. (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res*, **31**, 3701-3708.

- Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, **46**, 3-26.
- Lipinski, C.A. (2004) Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol*, **1**, 337-341.
- Llano, M., Delgado, S., Vanegas, M. & Poeschla, E.M. (2004) Lens epithelium-derived growth factor/p75 prevents proteasomal degradation of HIV-1 integrase. *J Biol Chem*, **279**, 55570-55577.
- Lodi, P.J., Ernst, J.A., Kuszewski, J., Hickman, A.B., Engelman, A., Craigie, R., Clore, G.M. & Gronenborn, A.M. (1995) Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry*, **34**, 9826-9833.
- Lopata, K., Wojdas, E., Nowak, R., Lopata, P. & Mazurek, U. (2018) Porcine Endogenous Retrovirus (PERV) - Molecular Structure and Replication Strategy in the Context of Retroviral Infection Risk of Human Cells. *Front Microbiol*, **9**, 730.
- Lu, D.Y., Wu, H.Y., Yarla, N.S., Xu, B., Ding, J. & Lu, T.R. (2018) HAART in HIV/AIDS Treatments: Future Trends. *Infect Disord Drug Targets*, **18**, 15-22.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251-260.
- Luke, K., Apiyo, D. & Wittung-Stafshede, P. (2005) Dissecting Homo-Heptamer Thermodynamics by Isothermal Titration Calorimetry: Entropy-Driven Assembly of Co-Chaperonin Protein 10. *Biophysical Journal*, **89**, 3332-3336.

M

- Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z. & Engelborghs, Y. (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J Biol Chem*, **278**, 33528-33539.
- 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.
- Magdeldin, S. & Moser, A. (2012) Affinity Chromatography: Principles and Applications. In.
- Maignan, S., Guilloteau, J.P., Zhou-Liu, Q., Clement-Mella, C. & Mikol, V. (1998) Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *J Mol Biol*, **282**, 359-368.
- Maillot, B., Lévy, N., Eiler, S., Crucifix, C., Granger, F., Richert, L., Didier, P., Godet, J., Pradeau-Aubretton, K., Emiliani, S., Nazabal, A., Lesbats, P., Parissi, V., Mely, Y., Moras, D., Schultz, P. & Ruff, M. (2013) Structural and Functional Role of INI1 and LEDGF in the HIV-1 Preintegration Complex. *PLoS ONE*, **8**, e60734.
- Maldarelli, F., Wu, X., Su, L., Simonetti, F.R., Shao, W., Hill, S., Spindler, J., Ferris, A.L., Mellors, J.W., Kearney, M.F., Coffin, J.M. & Hughes, S.H. (2014) HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science*, **345**, 179-183.

- Maskell, D.P., Renault, L., Serrao, E., Lesbats, P., Matadeen, R., Hare, S., Lindemann, D., Engelman, A.N., Costa, A. & Cherepanov, P. (2015) Structural basis for retroviral integration into nucleosomes. *Nature*, **523**, 366-369.
- Masuda, T. (2011) Non-Enzymatic Functions of Retroviral Integrase: The Next Target for Novel Anti-HIV Drug Development. *Front Microbiol*, **2**, 210.
- Matsuda, T., Kigawa, T., Koshiba, S., Inoue, M., Aoki, M., Yamasaki, K., Seki, M., Shinozaki, K. & Yokoyama, S. (2006) Cell-free synthesis of zinc-binding proteins. *J Struct Funct Genomics*, **7**, 93-100.
- Metifiot, M., Marchand, C. & Pommier, Y. (2013) HIV integrase inhibitors: 20-year landmark and challenges. *Adv Pharmacol*, **67**, 75-105.
- Miller, M.D., Farnet, C.M. & Bushman, F.D. (1997) Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol*, **71**, 5382-5390.
- 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., Felix, 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., Felix, 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.
- Moalic, Y., Felix, H., Jestin, A. & Blanchard, Y. (2011) Risque rétroviral endogène porcin (PERV) en xénotransplantation (French). *Xenotransplantation and porcine endogenous retroviruses (English)*, **15**, 63-72.
- Murray, J.M., Emery, S., Kelleher, A.D., Law, M., Chen, J., Hazuda, D.J., Nguyen, B.Y., Teppler, H. & Cooper, D.A. (2007) Antiretroviral therapy with the integrase inhibitor raltegravir alters decay kinetics of HIV, significantly reducing the second phase. *Aids*, **21**, 2315-2321.

N

- Nelson, P.N., Carnegie, P.R., Martin, J., Davari Ejtehad, H., Hooley, P., Roden, D., Rowland-Jones, S., Warren, P., Astley, J. & Murray, P.G. (2003) Demystified. Human endogenous retroviruses. *Mol Pathol*, **56**, 11-18.
- Nishizawa, Y., Usukura, J., Singh, D.P., Chylack, L.T., Jr. & Shinohara, T. (2001) Spatial and temporal dynamics of two alternatively spliced regulatory factors, lens epithelium-derived growth factor (ledgf/p75) and p52, in the nucleus. *Cell Tissue Res*, **305**, 107-114.
- 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.

P

- Passos, D.O., Li, M., Yang, R., Rebensburg, S.V., Ghirlando, R., Jeon, Y., Shkriabai, N., Kvaratskhelia, M., Craigie, R. & Lyumkis, D. (2017) Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science*, **355**, 89-92.
- Patience, C., Takeuchi, Y. & Weiss, R.A. (1997) Infection of human cells by an endogenous retrovirus of pigs. *Nat Med*, **3**, 282-286.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. & Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, **25**, 1605-1612.
- Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D. & Gallo, R.C. (1980) Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci U S A*, **77**, 7415-7419.
- Porath, J., Carlsson, J., Olsson, I. & Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature*, **258**, 598-599.
- Primrose, S.B. & Twyman, R.M. (2006) *Principles of Gene Manipulation and Genomics*. Blackwell Publishing.
- Putnam, C.D., Hammel, M., Hura, G.L. & Tainer, J.A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys*, **40**, 191-285.

Q

- Qi, X., Sun, Y. & Xiong, S. (2015) A single freeze-thawing cycle for highly efficient solubilization of inclusion body proteins and its refolding into bioactive form. *Microbial Cell Factories*, **14**, 24.

R

- Raghavendra, N.K., Shkriabai, N., Graham, R., Hess, S., Kvaratskhelia, M. & Wu, L. (2010) Identification of host proteins associated with HIV-1 preintegration complexes isolated from infected CD4⁺ cells. *Retrovirology*, **7**, 66.
- Robert, X. & Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*, **42**, W320-324.
- Rosano, G.L. & Ceccarelli, E.A. (2014) Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in microbiology*, **5**, 172-172.
- Rous, P. (1911) A Sarcoma Of The Fowl Transmissible By An Agent Separable From The Tumor Cells. *The Journal of experimental medicine*, **13**, 397-411.

S

- Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, **74**, 5463-5467.

- Sato, M., Motomura, T., Aramaki, H., Matsuda, T., Yamashita, M., Ito, Y., Kawakami, H., Matsuzaki, Y., Watanabe, W., Yamataka, K., Ikeda, S., Kodama, E., Matsuoka, M. & Shinkai, H. (2006) Novel HIV-1 integrase inhibitors derived from quinolone antibiotics. *J Med Chem*, **49**, 1506-1508.
- Scobie, L., Padler-Karavani, V., Le Bas-Bernardet, S., Crossan, C., Blaha, J., Matouskova, M., Hector, R.D., Cozzi, E., Vanhove, B., Charreau, B., Blancho, G., Bourdais, L., Tallacchini, M., Ribes, J.M., Yu, H., Chen, X., Kracikova, J., Broz, L., Hejnar, J., Vesely, P., Takeuchi, Y., Varki, A. & Soulillou, J.P. (2013) Long-term IgG response to porcine Neu5Gc antigens without transmission of PERV in burn patients treated with porcine skin xenografts. *J Immunol*, **191**, 2907-2915.
- Serrao, E., Ballandras-Colas, A., Cherepanov, P., Maertens, G.N. & Engelman, A.N. (2015) Key determinants of target DNA recognition by retroviral intasomes. *Retrovirology*, **12**, 39.
- Serrao, E. & Engelman, A.N. (2016) Sites of Retroviral DNA Integration: From Basic Research to Clinical Applications. *Critical reviews in biochemistry and molecular biology*, **51**, 26-42.
- Sharma, A., Larue, R.C., Plumb, M.R., Malani, N., Male, F., Slaughter, A., Kessler, J.J., Shkriabai, N., Coward, E., Aiyer, S.S., Green, P.L., Wu, L., Roth, M.J., Bushman, F.D. & Kvaratskhelia, M. (2013) BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc Natl Acad Sci U S A*, **110**, 12036-12041.
- Skalka, A.M. & Katz, R.A. (2005) Retroviral DNA integration and the DNA damage response. *Cell Death Differ*, **12 Suppl 1**, 971-978.
- Smith, D.B. & Johnson, K.S. (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene*, **67**, 31-40.
- Sterling, T. & Irwin, J.J. (2015) ZINC 15--Ligand Discovery for Everyone. *J Chem Inf Model*, **55**, 2324-2337.
- Studamire, B. & Goff, S.P. (2008) Host proteins interacting with the Moloney murine leukemia virus integrase: multiple transcriptional regulators and chromatin binding factors. *Retrovirology*, **5**, 48.
- Studier, F.W. & Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*, **189**, 113-130.
- Sundquist, W.I. & Krausslich, H.G. (2012) HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med*, **2**, a006924.
- Suzuki, Y. & Craigie, R. (2007) The road to chromatin - nuclear entry of retroviruses. *Nat Rev Microbiol*, **5**, 187-196.
- Svergun, D.I. & Koch, M.H.J. (2003) Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*, **66**, 1735-1782.

T

- Takeuchi, Y., Patience, C., Magre, S., Weiss, R.A., Banerjee, P.T., Le Tissier, P. & Stoye, J.P. (1998) Host range and interference studies of three classes of pig endogenous retrovirus. *J Virol*, **72**, 9986-9991.
- Tegel, H., Ottosson, J. & Hober, S. (2011) Enhancing the protein production levels in Escherichia coli with a strong promoter. *Febs j*, **278**, 729-739.

- Tekeste, S.S., Wilkinson, T.A., Weiner, E.M., Xu, X., Miller, J.T., Le Grice, S.F., Clubb, R.T. & Chow, S.A. (2015) Interaction between Reverse Transcriptase and Integrase Is Required for Reverse Transcription during HIV-1 Replication. *J Virol*, **89**, 12058-12069.
- Temin, H.M. & Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, **226**, 1211-1213.
- Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol*, **60**, 523-533.
- Trott, O. & Olson, A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*, **31**, 455-461.

U

- Uchiyama, T., Yodoi, J., Sagawa, K., Takatsuki, K. & Uchino, H. (1977) Adult T-cell leukemia: clinical and hematologic features of 16 cases. *Blood*, **50**, 481-492.
- Umehara, T., Nakamura, Y., Jang, M.K., Nakano, K., Tanaka, A., Ozato, K., Padmanabhan, B. & Yokoyama, S. (2010) Structural basis for acetylated histone H4 recognition by the human BRD2 bromodomain. *J Biol Chem*, **285**, 7610-7618.

V

- van Nuland, R., van Schaik, F.M., Simonis, M., van Heesch, S., Cuppen, E., Boelens, R., Timmers, H.M. & van Ingen, H. (2013) Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. *Epigenetics Chromatin*, **6**, 12.
- Velazquez-Campoy, A., Leavitt, S.A. & Freire, E. (2004) Characterization of protein-protein interactions by isothermal titration calorimetry. *Methods Mol Biol*, **261**, 35-54.
- Viejo-Borbolla, A., Ottinger, M., Bruning, E., Burger, A., Konig, R., Kati, E., Sheldon, J.A. & Schulz, T.F. (2005) Brd2/RING3 interacts with a chromatin-binding domain in the Kaposi's Sarcoma-associated herpesvirus latency-associated nuclear antigen 1 (LANA-1) that is required for multiple functions of LANA-1. *J Virol*, **79**, 13618-13629.
- Vincent, K.A., York-Higgins, D., Quiroga, M. & Brown, P.O. (1990) Host sequences flanking the HIV provirus. *Nucleic Acids Res*, **18**, 6045-6047.
- Vogt, V.M. (1997) Retroviral Virions and Genomes. In Coffin, J.M., Hughes, S.H. & Varmus, H.E. (eds.), *Retroviruses*. Cold Spring Harbor Laboratory Press
Cold Spring Harbor Laboratory Press., Cold Spring Harbor (NY).

W

- Wang, F., Liu, H., Blanton, W.P., Belkina, A., Lebrasseur, N.K. & Denis, G.V. (2009) Brd2 disruption in mice causes severe obesity without Type 2 diabetes. *Biochem J*, **425**, 71-83.
- Wang, J.Y., Ling, H., Yang, W. & Craigie, R. (2001) Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *Embo j*, **20**, 7333-7343.

- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R. & Schwede, T. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*, **46**, W296-w303.
- Weidner-Glunde, M., Ottinger, M. & Schulz, T.F. (2010) WHAT do viruses BET on? *Front Biosci (Landmark Ed)*, **15**, 537-549.
- Wilens, C.B., Tilton, J.C. & Doms, R.W. (2012) Molecular mechanisms of HIV entry. *Adv Exp Med Biol*, **726**, 223-242.
- Winans, S., Larue, R.C., Abraham, C.M., Shkriabai, N., Skopp, A., Winkler, D., Kvaratskhelia, M. & Beemon, K.L. (2017) The FACT Complex Promotes Avian Leukosis Virus DNA Integration. *J Virol*, **91**.
- Wu, S.Y. & Chiang, C.M. (2007) The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. *J Biol Chem*, **282**, 13141-13145.

Y

- Yajjou-Hamalian, H. (2017) 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. Université Claude Bernard Lyon 1.
- Yamashita, M. & Emerman, M. (2006) Retroviral infection of non-dividing cells: old and new perspectives. *Virology*, **344**, 88-93.
- Yang, Z.-N., Mueser, T.C., Bushman, F.D. & Hyde, C.C. (2000) Crystal structure of an active two-domain derivative of rous sarcoma virus integrase1. *Journal of Molecular Biology*, **296**, 535-548.
- Yin, Z., Shi, K., Banerjee, S., Pandey, K.K., Bera, S., Grandgenett, D.P. & Aihara, H. (2016) Crystal structure of the Rous sarcoma virus intasome. *Nature*, **530**, 362-366.
- Yoder, K.E. & Bushman, F.D. (2000) Repair of Gaps in Retroviral DNA Integration Intermediates. *Journal of Virology*, **74**, 11191-11200.

Z

- Zhang, Q., Zeng, L., Shen, C., Ju, Y., Konuma, T., Zhao, C., Vakoc, C.R. & Zhou, M.M. (2016) Structural Mechanism of Transcriptional Regulator NSD3 Recognition by the ET Domain of BRD4. *Structure*, **24**, 1201-1208.