N°d'ordre NNT : 2017LYSE1245

THESE de DOCTORAT DE L'UNIVERSITE DE LYON

opérée au sein de

l'Université Claude Bernard Lyon 1

Ecole Doctorale N°205

EDISS – Ecole Doctorale InterDisciplinaire Sciences-Santé

Spécialité de doctorat : Biochimie

Soutenue publiquement le 30/11/2017, par :

Christelle FOLIO

Études fonctionnelle et structurale de deux protéines rétrovirales d'intérêt thérapeutique : la protéine Tax du virus HTLV et la protéine de capside du FIV.

Devant le jury composé de :

JAULT, Jean-Michel	DR	CNRS, Lyon	Examinateur
BERNACCHI, Serena	CR	CNRS, Strasbourg	Rapporteure
MORERA, Solange	DR	CNRS, Saclay	Rapporteure
MOUREY, Lionel	DR	IPBS, Toulouse	Rapporteur
MIELE, Adriana	PR	UCBL, Lyon	Examinatrice
GUILLON, Christophe	CR	CNRS, Lyon	Directeur de thèse

UNIVERSITE CLAUDE BERNARD - LYON 1

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

COMPOSANTES SANTE

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

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

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

REMERCIEMENTS

Alors que mon aventure doctorale touche à sa fin, je tiens à remercier toutes les personnes qui ont contribué à l'aboutissement de ce projet. Sans vous, rien n'aurait été possible.

Je souhaite tout d'abord 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, qui m'ont accueillie au sein de l'Institut pendant ces quatre années.

J'adresse également mes remerciements aux membres de mon jury, les Dr. Serena Bernacchi, Jean-Michel Jault, Solange Morera et Lionel Mourey ainsi que le Pr Adriana Miele qui m'ont fait l'honneur de juger mon travail de thèse.

Mes plus sincères remerciements vont ensuite à mon directeur de thèse, le Dr. Christophe Guillon, qui m'a accordé sa confiance depuis mon premier stage. Je souhaiterais le remercier pour ses précieux conseils, sa présence et la patience dont il a su faire preuve tout au long de ce parcours. Je n'oublierai jamais son humour et sa grande culture générale (surtout ses lectures bizarres) qui m'a toujours impressionnée. Aussi, je le remercie pour ses dons surnaturels en biomol sans lesquels ce travail de thèse n'aurait pas abouti.

Pour les nombreux équipements que j'ai pu utiliser au cours de ce projet doctoral, je souhaite remercier les membres de la plateforme *Protein Science Facility* (*PSF*, SFR Biosciences, UMS 3444) et plus particulièrement Virginie, Eric et Isabelle pour leur aide et leurs conseils.

Je remercie également les gestionnaires Dorothée, Souad et Christine pour leur gentillesse.

A mes remerciements j'associe l'organisme Ecos-Sud (Action n°U14S01), qui m'a permis d'effectuer un séjour de recherche inoubliable en Uruguay.

Je tiens aussi à exprimer ma gratitude au Dr. Guzman Alvarez qui m'a accueillie au sein de son équipe *« Laboratorio de Moléculas Bioactivas »*, à Paysandù en Uruguay. Je remercie les membres de son équipe pour leur accueil et leur gentillesse à mon égard.

Mes remerciements s'adressent également à tous les membres de l'équipe « Biocristallographie et Biologie Structurale des Cibles Thérapeutiques » du Dr. Nushin Aghajari. Lionel, je te remercie pour toutes tes blagues douteuses, ta folie et les bonnes parties de rigolade que l'on a pu avoir. Merci à Seb pour ta bonne humeur. J'espère que tu finiras par regarder Breaking Bad un jour… Fred, mi remerci a ou pou tou lé ti sega nou la danse ensamb et tout' le band soiré ou sa nou la boir et ri ensamb. Enfin, merci à Carrique (même si tu ne fais plus partie de l'équipe, RIP) pour tous tes conseils scientifiques ou non et tes blagues pourries (tu fais bien la paire avec Lionel !). Tu as été le meilleur thésard en chef qui puisse exister !

Je souhaite également remercier les membres de mon équipe pour tout ce que nous avons pu partager au fil de ces années. Merci au chef, le Pr Patrice Gouet, qui a toujours été bienveillant envers moi. X, merci pour ton soutien, tes relectures et les discussions que nous avons eu sur l'avenir et la vie (et pour l'horoscope quotidien, bien sûr !). Merci à Marie pour ta gentillesse. Francesca, je te remercie pour tes conseils d'après-thèse, tes pâtisseries et de m'avoir prêté ton petit kiwi. Je ne pouvais pas écrire de remerciements sans y inclure Romain, maître canin des techniques de qualité. Wouaf.

Margaux et Halima, je ne vous remercierai jamais assez pour tous ces bons moments que l'on a passés ensemble. Je suis heureuse de vous avoir rencontrées et de vous compter parmi mes amies. J'espère que notre belle amitié durera encore longtemps. Merci Margaux pour les pauses au soleil, les discussions films, les pauses jeux et *blindtest* et surtout ta folie Diiisneyylaaand. J'attends avec impatience le prochain concert des Arctic Monkeys ! Enfin, merci à Halima, my love. Je n'aurais jamais pensé qu'on se lierait autant d'amitié toutes les deux et je suis heureuse d'avoir vécu cette aventure à tes côtés. Merci pour ton oreille attentive et tes conseils avisés, pour toutes les chansons que tu m'as chantées et pour m'avoir rappelé d'aller sur Showroom tous les matins.

Mes derniers remerciements IBCPiens s'adressent à mes collègues et amis thésards et ingénieurs qui ont rendu cette aventure plus qu'agréable. Merci donc à Juju Crayon (je crois que ça s'écrit comme ça), Marine, Célia, Sylvain, Maxime, Stéphanie, JB et Thomas (Charles, l'américain) pour l'inoubliable retraire MMSB et toutes ces soirées à rigoler et refaire le monde. Merci également à Sylvain et à Frank pour tous les bons moments passés ensemble.

Enfin, merci à Arthur pour la jolie et transcendantale histoire que nous avons commencé à écrire tous les deux.

Au fil des années, j'ai également pu tisser des liens avec des personnes formidables qui ont rendu ma vie plus belle. A mes plus vieux amis, Adrien, Jessy et Antoine, merci pour tous ces moments passés ensemble. Malgré la distance, je sais maintenant que nous nous retrouverons toujours, comme si l'on ne s'était jamais quitté. Je remercie aussi mes amis rencontrés à la fac, Hanane, Marie, Laurine et Alain. Merci aux vrais Arthur, Emon et Michou pour tous ces apéros bongos et ces soirées de folie.

Je tiens plus particulièrement à remercier ma meilleure amie (#bff), Marie-Charlotte. Je chéris toutes les aventures que nous avons vécues ensemble et j'espère que les prochaines seront encore plus belles. Merci pour ton soutien et ta présence à mes côtés. Merci pour ton amitié. Tu es ma sœur, ma famille.

Une pensée également pour toute ma famille à La Réunion qui me manque tous les jours.

Enfin, je terminerai en remerciant ma maman, Jessie, et ma sœur, Dorinda, que j'aime plus que tout. Merci de m'avoir toujours soutenue et de croire en moi. Do, malgré nos cinq années d'écart, nous avons réussi à nous rapprocher et à trouver une jolie complicité. Je ne saurai en être plus heureuse. Maman, en dépit de ce que tu peux penser, tu es la meilleure maman qui puisse exister. Tu es une femme forte et courageuse et tu nous as tout donné depuis le premier jour, merci. Je n'aurais jamais accompli tout ce chemin sans toi. J'espère, un jour, être à ta hauteur. Je vous aime.

Merci à tous pour tout ce que vous m'avez apporté durant ces dernières années.

J'oubliais : je souhaiterais remercier Doriane de la Roche Pallud pour son soutien, les séances ronron (presque) quotidiennes et certaines activités moins agréables que nous avons partagées depuis presque dix ans déjà.

« Pourquoi manger des fruits quand on peut les boire ? »

TABLE DES MATIERES

REMERCIEMENTS 5

TABLE DES MATIERES 11

LISTE DES ABREVIATIONS 17

INTRODUCTION 21

1. LES RETROVIRUS 23

1.1. GENERALITES 23

1.2. ORGANISATIONS GENOMIQUE ET STRUCTURALE DES RETROVIRUS 24

1.2.1. La particule rétrovirale 24

1.2.2. Le génome rétroviral 25

1.2.3. Cas des rétrovirus complexes 27

1.2.3.1. *Les deltarétrovirus : le virus humain T-lymphotrope HTLV* 27

1.2.3.2. *Les lentivirus : les virus de l'immunodéficience humaine et féline, VIH et FIV* 28

1.3. LE CYCLE REPLICATIF RETROVIRAL 30

2. LES VIRUS HTLV 32

2.1. HTLV, PREMIER RETROVIRUS ONCOGENE HUMAIN 32

2.1.1. Découverte des HTLV 32

2.1.2. Epidémiologie des HTLV 32

2.1.2.1. *Prévalence des virus HTLV dans le monde* 32

2.1.2.2. *Modes de transmission* 33

2.1.2.3. *Moyens de prévention et traitements* 34

2.2. HTLV, DES VIRUS PATHOGENES PLUS OU MOINS VIRULENTS 34

2.2.1. Pathologies associées à HTLV-1 34

2.2.1.1. *La leucémie à cellules T de l'adulte, ATLL* 34

2.2.1.2. *La myélopathie associée à HTLV/paraparésie spastique tropicale, HAM/TSP* 35

2.2.2. Pathologies associées à HTLV-2 36

2.3. L'ONCOPROTEINE TAX 36

2.3.1. Tax, une protéine multifonctionnelle 36

2.3.1.1. *Régulation de l'expression des gènes viraux* 36

2.3.1.2. *Régulation de l'expression des gènes cellulaires* 38

2.3.2. Tax, une protéine multi-domaines 42
2.3.3. Tax-1 et Tax-2 : petites différences, grandes conséquences ? 43
2.3.3.1. *Tax-1 et Tax-2, des modifications post-traductionnelles variées* 43
2.3.3.2. *Tax-1 et l'activation de la voie non canonique du NF-kB* 44
2.3.3.3. *Tax-1 et la liaison aux protéines à domaine PDZ* 44
3. LE VIRUS FIV 45
3.1. FIV, VIRUS DE L'IMMUNODEFICIENCE FELINE 45
3.1.1. Découverte du FIV 45
3.1.2. Epidémiologie du FIV 45
3.1.2.1. *Prévalence du FIV dans le monde* 45
3.1.2.2. *Modes de transmission* 46
3.1.2.3. *Moyens de prévention et traitements* 46
3.2. FIV, AGENT DU SYNDROME D'IMMUNODEFICIENCE ACQUISE FELINE 47
3.3. LA PROTEINE DE CAPSIDE 48
3.3.1. La CA, moteur de l'assemblage de la particule virale 48
3.3.1.1. *Gag, précurseur de la CA* 48
3.3.1.2. *Assemblage de la particule virale* 50
3.3.1.3. *Maturation de la particule virale* 50
3.3.2. La CA, une organisation structurale conservée chez les lentivirus 51
3.3.2.1. *Le domaine amino-terminal* 51
3.3.2.2. *Le domaine carboxy-terminal* 52
3.3.3. La CA, oligomérisation et formation de la capside virale 53
3.4. DEVELOPPEMENT D'INHIBITEURS DIRIGES CONTRE L'ASSEMBLAGE DE LA CA 56

LE PROJET DOCTORAL 59

MATERIELS ET METHODES 63

1. CONSTRUCTIONS PLASMIDIQUES 65
1.1. CONSTRUCTIONS PLASMIDIQUES DE LA PROTEINE TAX D'HTLV-2 (TAX-2) 65
1.1.1. Stratégie de clonage 65
1.1.2. Amplification en chaîne par polymérase 69
1.1.3. Clonage dans le vecteur pCR®2.1 selon la méthode du TA Cloning® 70
1.1.3.1. *Ligation dans le vecteur pCR®2.1* 70
1.1.3.2. *Transformation bactérienne et extraction d'ADN plasmidique* 71
1.1.3.3. *Sélection des transformants positifs* 72

1.1.4. Sous-clonage des séquences codantes dans le vecteur d'expression 72
1.2. CONSTRUCTION PLASMIDIQUE DE LA PROTEINE DE CAPSIDE DU FIV 74
2. SURPRODUCTION ET PURIFICATION DES PROTEINES D'INTERET 75
2.1. GENERALITES 75
2.1.1. Surproduction en système procaryote 75
2.1.2. Purification par chromatographie d'affinité pour le nickel 75
2.2. LA PROTEINE TAX-2 DU VIRUS HTLV-2 76
2.2.1. Tests de surproduction de la protéine Tax-2 en système bactérien 76
2.2.2. Lyse bactérienne et extraction de la protéine Tax-2 79
2.2.3. Solubilisation des protéines Tax-2 80
2.2.3.1. *Dénaturation par le chlorure de guanidinium* 80
2.2.3.2. *Dénaturation par l'urée* 80
2.2.4. Purification des protéines Tax-2 81
2.2.4.1. *Précipitation au sulfate d'ammonium* 81
2.2.4.2. *Purification par chromatographie d'affinité pour le nickel* 82
2.3. LA PROTEINE DE CAPSIDE DU FIV 83
2.3.1. Surproduction de la CA en système bactérien 83
2.3.2. Purification de la CA 83
2.3.2.1. *Lyse bactérienne* 83
2.3.2.2. *Purification de la CA par chromatographie d'affinité pour le nickel* 84
2.3.3. Concentration de la CA et échange de tampon 85
2.3.4. Clivage de l'étiquette d'affinité de la CA du FIV 85
3. ANALYSE QUALITATIVE ET QUANTITATIVE DES PROTEINES 86
3.1. ÉLECTROPHORESE SUR GEL DISCONTINU DE POLYACRYLAMIDE EN CONDITIONS DENATURANTES 86
3.2. *WESTERN BLOT* 86
3.3. CHROMATOGRAPHIE D'EXCLUSION STERIQUE 87
3.4. DIFFUSION DYNAMIQUE DE LA LUMIERE 88
4. SUIVI DE L'ASSEMBLAGE DE LA CA DU FIV *IN VITRO* ET IDENTIFICATION D'INHIBITEURS 89
4.1. SUIVI DE L'ASSEMBLAGE DE LA CA *IN VITRO* PAR DLS 89
4.2. IDENTIFICATION D'INHIBITEURS DE L'ASSEMBLAGE DE LA CA *IN VITRO* PAR SPECTROPHOTOMETRIE UV 89
5. ÉTUDE STRUCTURALE DES PROTEINES D'INTERET 91
5.1. ÉTUDE PAR DIFFUSION DES RAYONS X AUX PETITS ANGLES 91
5.1.1. Principe de la diffusion des rayons X aux petits angles 91
5.1.2. Acquisition et analyse des données 93

5.2. ÉTUDE STRUCTURALE PAR CRISTALLOGRAPHIE AUX RAYONS X 94
5.2.1. Cristallogenèse 94
5.2.1.1. *Principe de la cristallogenèse* 94
5.2.1.2. *Criblage des conditions de cristallisation* 96
5.2.1.3. *Technique de co-cristallisation* 96
5.2.2. Principe de la diffraction aux rayons X 97
5.2.3. Analyse des données de diffraction 99
5.2.3.1. *Traitement des données de diffraction* 99
5.2.3.2. *Solution du problème de phase par remplacement moléculaire* 100
5.2.3.3. *Affinement de la structure* 101

RESULTATS ET DISCUSSIONS 103

1. ÉTUDE DE LA PROTEINE TAX-2 DU VIRUS HTLV-2 105
1.1. CONSTRUCTIONS PLASMIDIQUES 105
1.2. EXPRESSION ET PURIFICATION DES PROTEINES TAX-2 SEULES 107
1.2.1. Tests d'expression 107
1.2.1.1. *Tax-2 en fusion avec une étiquette His_6* 109
1.2.1.2. *Tax-2 en fusion avec une étiquette de solubilité* 110
1.2.2. Optimisation des tests d'expression 111
1.2.2.1. *Expression en présence de $ZnSO_4$* 112
1.2.2.2. *Expression avec ajout de tétracycline : arrêt de la traduction* 112
1.2.2.3. *Expression avec ajout de rifampicine : arrêt de la transcription* 113
1.2.3. Solubilisation des protéines Tax-2 114
1.2.3.1. *Dénaturation par le chlorure de guanidinium* 114
1.2.3.2. *Dénaturation par l'urée* 115
1.2.4. Précipitation au sulfate d'ammonium et purification par Ni-NTA de Tax-2 116
1.3. CO-EXPRESSION DE LA PROTEINE TAX-2 ENTIERE AVEC CREB 117
1.3.1. Expression avec ajout de rifampicine 118
1.3.2. Purification par Ni-NTA 119
1.4. DISCUSSION 120
2. ÉTUDES FONCTIONNELLES ET STRUCTURALES DE LA PROTEINE DE CAPSIDE DU VIRUS FIV 122
2.1. SURPRODUCTION, PURIFICATION ET CLIVAGE DE L'ETIQUETTE D'AFFINITE 122
2.1.1. Purification de la CA 122
2.1.2. Clivage de l'étiquette d'affinité 123

2.2. ETUDES FONCTIONNELLES DE LA CA DU FIV 123
2.2.1. Etat d'oligomérisation de la CA en solution 124
2.2.2. Suivi de l'assemblage *in vitro* par DLS 124
2.3. ETUDES STRUCTURALES DE LA CA DU FIV 125
2.3.1. Etude structurale par cristallographie aux rayons X 125
2.3.1.1. *Cristallogenèse* 125
2.3.1.2. *Résolution de la première structure de la CA du FIV* 128
2.3.2. Etude structurale par SAXS 134
2.4. IDENTIFICATION D'INHIBITEURS DE L'ASSEMBLAGE DE LA CA DU FIV 136
2.5. DISCUSSION 138

CONCLUSION ET PERSPECTIVES 145

BIBLIOGRAPHIE 151

PUBLICATIONS 165

RESUMES 189

LISTE DES ABREVIATIONS

ARNm	Acide ribonucléique messager
ATF	*Activating Transcription Factor*
ATLL	*Adult T-cell Leukemia/Lymphoma* | Leucémie à cellules T de l'adulte
AZT	Zidovudine
BCIP	5-Bromo-4-Chloro-3-Indolyl Phosphate
BMV	Bevirimat
CA	Protéine de capside
CA_{CTD}	*Capsid C-terminal domain* | Domaine C-terminal de la CA
CA_{NTD}	*Capsid N-terminal domain* | Domaine N-terminal de la CA
CAI	*Capsid Assembly Inhibitor*
CAPS	*3-(cyclohexylamino)-1-propanesulfonic acid*
CBP	*CREB Binding Protein*
CCR5	*C-C Chemokine Receptor type 5*
CD4	*Cluster of Differenciation 4*
CD134	*Cluster of Differenciation 134*
Cdk	*Cyclin-dependent kinase*
CHAPS	3-[(3-cholamidopropyl)diméthylammonio]-1-propanesulfonate
CIAP	Phosphatase alcaline intestinale de veau
CREB	*Cyclic AMP Response Element Binding-protein*
CV	*Column Volume* | Volume colonne
CXCR4	*Chemokine (C-X-C motif) Receptor type 4*
CypA	*Cyclophilin A*
CypA-BL	*Cyclophilin A-Binding Loop*
DLS	*Dynamic Light Scattering* | Diffusion dynamique de la lumière
DMSO	Diméthylsulfoxide
dNTP	désoxynucléotide triphosphate
DTT	Dithiothréitol
E. coli	*Escherichia coli*
EDTA	*Ethylene Diamine Tetraacetic Acid*
EIAV	*Equine Infectious Anemia Virus*
env	*envelope*
ESRF	*European Synchrotron Radiation Facility*
FIV	*Feline Immunodeficiency Virus* | Virus de l'immunodéficience féline
gag	*group-specific antigen*
GLUT-1	*GLUcose Transporter 1*
HAM	*HTLV-1–Associated Myelopathy*
HB	*Hyper Broth*
HBZ	*HTLV-1 bZIP factor*

HTLV *Human T-Lymphotropic Virus* | Virus humain T-lymphotrope
HSPG *Heparan Sulfate ProteoGlycans*
IFN-α Interféron α
IκB *Inhibitory protein kappaB*
IKK *IκB Kinase*
IN Intégrase
IPTG IsoPropyl β-D-1-ThioGalactopyranoside
kb kilobase
kDa kiloDalton
LB *Lysogeny Broth*
LDS *Lithium Dodecyl Sulfate*
LTR *Long Terminal Repeat*
LZR *Leucine Zipper-like Region*
MA Protéine de matrice
MCS *Multiple Cloning Site* | Site multiple de clonage
MES *2-(N-Morpholino)EthaneSulfonic acid*
NBT *NitroBlue Tetrazolium*
NC Protéine de nucléocapside
Nef *Negative regulatory factor*
NES *Nuclear Export Signal* | Signal d'export nucléaire
NF-κB *Nuclear Factor-kappaB*
NIK *NF-κB Inducing kinase*
NLS *Nuclear Localization Signal* | Signal de localisation nucléaire
NRP-1 *Neuropilin-1*
NTA *NitriloTriacetic Acid*
o/n *Overnight* | Pendant la nuit
ORF *Open Reading Frame* | Phase ouverte de lecture
PBS *Phosphate Buffer Saline*
PCR *Polymerase Chain Reaction* | Amplification en chaîne par polymérase
PBM *PDZ Binding Motif* | Motif de liaison aux protéines à domaine PDZ
PEG Polyéthylène glycol
PMSF *Phenylmethylsulfonyl fluoride* | Fluorure de phénylméthylsulfonyle
pol *polymerase*
PR Protéase
Relik *Rabbit endogenous lentivirus type K*
Rev *Regulator of expression of viral proteins*
Rex *Regulator of expression of pX viral proteins*
RT *Reverse Transcriptase* | Transcriptase inverse
SAXS *Small Angle X-ray Scattering* | Diffusion des rayons X aux petits angles
SDS *Sodium Dodecyl Sulfate*

SDS-PAGE	*SDS - PolyAcrylamide Gel Electrophoresis* | Electrophorèse en conditions dénaturantes
SIDA	Syndrome d'immunodéficience acquise
SIV	*Simian Immunodeficiency Virus* | Virus de l'immunodéficience simienne
SOC	*Super Optimal broth with Catabolic repression*
SU	Glycoprotéine de surface
STLV	*Simian T-cell Lymphotropic Virus* | Virus simien T-lymphotrope
TAE	*Tris-Acétate/EDTA*
Tat	*Transactivator of transcription*
Tax	*Transactivator of transcription of pX*
TB	*Turbo Broth*
TED	Tris(carboxyméthyl)ethylène diamine
TerB	*Terrific Broth*
TM	Glycoprotéine TransMembranaire
TSP	*Tropical Spastic Paraparesis*
TxRE	*Tax Responsive Element* | Elément de réponse à Tax
vCRE	*viral cAMP Response Element* | Elément viral de réponse à l'AMPc
Vif	*Viral infectivity factor*
VIH	Virus de l'immunodéficience humaine
Vpr	*Viral protein r*
Vpu	*Viral protein u*
X-Gal	5-bromo-4-chloro-3-indolyl-β- D -galactopyranoside

INTRODUCTION

1. Les rétrovirus

1.1. Généralités

Les rétrovirus sont des virus enveloppés de la famille des *Retroviridae.* Ces virus, de forme sphérique, possèdent un génome ARN sous forme de deux brins monocaténaires à polarité positive. Leur particularité éponyme réside en la présence, au sein de la particule rétrovirale, d'une enzyme particulière appelée transcriptase inverse qui catalyse la rétrotranscription, c'est à dire la synthèse d'un ADN double brin à partir du génome viral ARN simple brin, en vue de son intégration dans le génome de la cellule infectée.

D'après la classification du Comité International de Taxonomie des Virus (ICTV, *International Committee on Taxonomy of Viruses*) (King *et al.*, 2011), basé sur l'analyse phylogénétique des régions conservées du gène viral *pol* (**Figure 1**), la famille des *Retroviridae* est divisée en deux sous-familles : les *Spumaretroviridae* et les *Orthoretroviridae*, divisée elle-même en plusieurs genres. Au sein des *Orthoretroviridae,* six genres sont distingués, parmi lesquels les *Deltarétrovirus*, dotés d'un potentiel oncogénique et les *Lentivirus*, capables d'induire des maladies chroniques à évolution lente chez leur hôte.

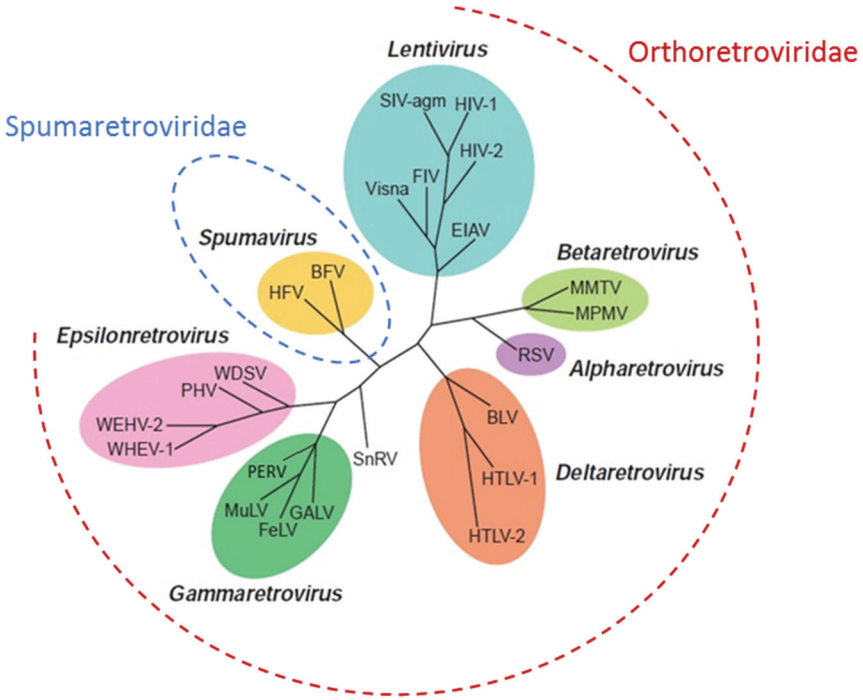

Figure 1 | Classification des rétrovirus basée sur l'analyse génétique des régions conservées du gène rétroviral *pol* codant la transcriptase inverse (King *et al.*, 2011).

1.2. Organisations génomique et structurale des rétrovirus

1.2.1. La particule rétrovirale

La particule rétrovirale mature (**Figure 2**) se présente sous forme sphérique d'environ 80 à 100 nm de diamètre. L'enveloppe protège le cœur central constitué d'une capside virale et d'un génome composé de deux molécules identiques d'ARN simple brin.

L'enveloppe virale est constituée d'une bicouche phospholipidique, d'origine cellulaire, dans laquelle est insérée la glycoprotéine transmembranaire (TM) elle-même liée à la glycoprotéine de surface (SU). Ces glycoprotéines d'enveloppe, d'origine virale, permettent aux virus de pénétrer les cellules cibles.

Sous l'enveloppe se trouvent les protéines de matrice (MA) qui renferment la capside virale, issue de l'assemblage des protéines de capside (CA). Au cœur de cette capside se trouvent la transcriptase inverse (RT) et l'intégrase (IN), ainsi que les deux copies du génome viral associées à la protéine de nucléocapside (NC) (**Figure 2**).

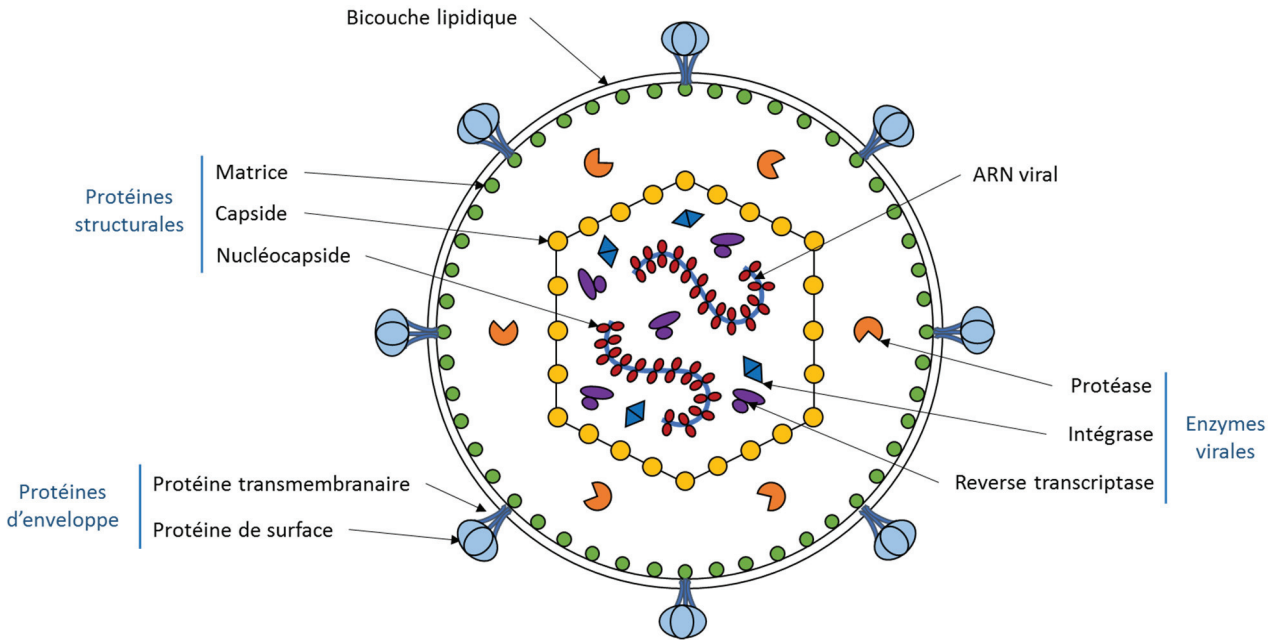

Figure 2 | Modèle d'une particule rétrovirale mature. Adapté de (Verdonck *et al.*, 2007).

1.2.2. Le génome rétroviral

Pour tous les membres de la sous-famille des *Orthoretroviridae*, l'organisation génomique ainsi que les différentes protéines codées par leur génome sont sensiblement similaires. En effet, trois gènes communs à tous les rétrovirus infectieux, *gag*, *pol* et *env*, codent les différentes protéines et enzymes virales (**Figure 3**).

Le génome rétroviral est constitué de deux molécules d'ARN simple brin à polarité positive, de 7 à 13 kb, coiffées et poly-adénylées (King *et al.*, 2011).

Sous forme d'ADN proviral, c'est-à-dire intégré au génome de l'hôte, le génome rétroviral est flanqué de deux séquences terminales répétées non codantes, appelées LTR (*Long Terminal Repeat*) (**Figure 3**), générées lors de la transcription inverse, et impliquées dans l'intégration et la régulation de l'expression du génome viral.

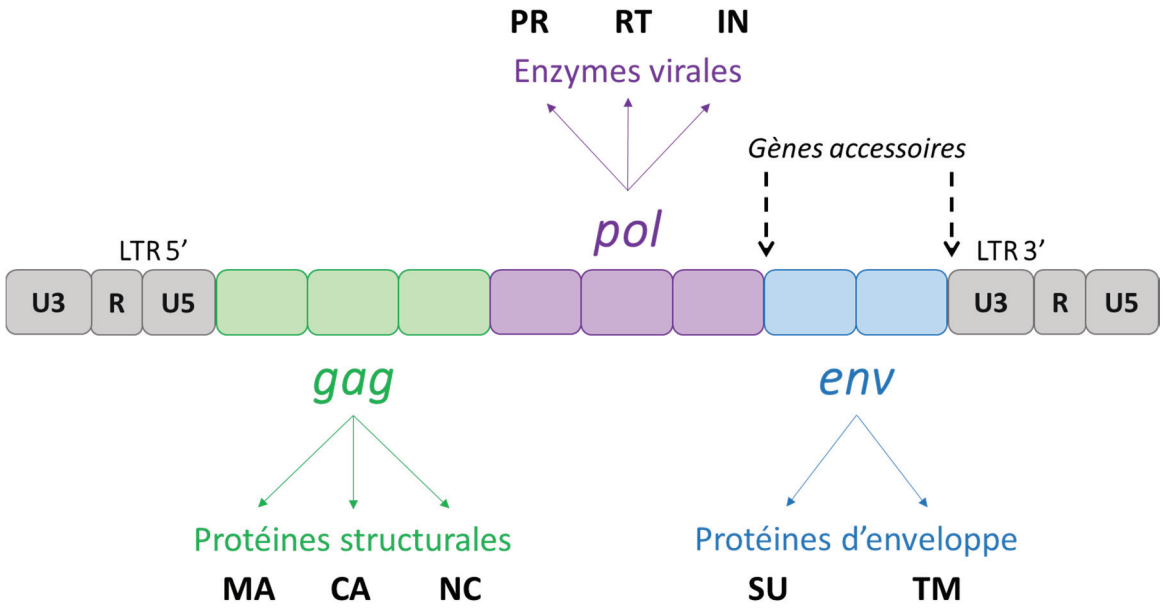

Figure 3 | Représentation schématique de l'organisation génomique des rétrovirus. Adapté de (King *et al.*, 2011).

Dans sa partie LTR 5', la séquence du provirus contient un promoteur permettant de transcrire les ARN messagers (ARNm) codant les protéines structurales à partir du gène *gag,* les enzymes virales à partir du gène *pol* ou encore les glycoprotéines d'enveloppe à partir du gène *env* (**Figure 3**) (Vogt, 1997).

Le gène *gag* code un précurseur polyprotéique, nommé Gag, dont le clivage par la protéase virale (PR) est à l'origine des différentes protéines impliquées dans l'organisation structurale du virus et la formation de la particule virale (Freed, 1998) :

- La protéine de matrice (MA), ancrée dans la membrane virale, participe notamment à l'étape d'assemblage des nouvelles particules virales ;
- La protéine de capside (CA), en plus de son rôle dans l'assemblage et la maturation des virions, s'assemble pour former une structure, nommée capside, protégeant le génome viral (**Figure 2**) ;
- La protéine de nucléocapside (NC), impliquée dans l'interaction avec l'ARN viral, participe à l'encapsidation du génome viral.

Le gène *pol* code également un précurseur polyprotéique, qui sera lui-même clivé par la protéase virale. Il code les différentes enzymes virales impliquées dans la réplication et l'intégration du génome viral (Vogt, 1997). Les enzymes communes à tous les rétrovirus sont les suivantes :

- La protéase (PR) impliquée dans le clivage de la polyprotéine Gag ;
- La transcriptase inverse (RT) catalysant la réaction de transcription inverse de l'ARN viral en ADN ;
- L'intégrase (IN) permettant l'intégration de l'ADN viral rétro-transcrit.

Enfin, le gène *env* code les glycoprotéines d'enveloppe virales impliquées dans le processus d'entrée des virus dans la cellule hôte (Vogt, 1997) :

- Les glycoprotéines de surface (SU) engagées dans la reconnaissance et l'interaction avec les récepteurs et co-récepteurs cellulaires ;
- Les glycoprotéines transmembranaires (TM) permettant l'ancrage des virions à la membrane virale pour la fusion et l'entrée des virus.

Certains rétrovirus, dit « complexes » (epsilon et deltarétrovirus, lentivirus et spumavirus), contiennent des gènes supplémentaires codant des protéines non structurales impliquées dans la régulation de l'expression des gènes et la réplication virale.

1.2.3. Cas des rétrovirus complexes

1.2.3.1. Les deltarétrovirus : le virus humain T-lymphotrope HTLV

Les deltarétrovirus ont été découverts chez certains groupes de mammifères dont l'Homme (HTLV, virus humain T-lymphotrope) et les primates (STLV, virus simien T-lymphotrope). L'infection par ces rétrovirus présente une longue période de latence avant l'apparition des pathologies induites. Elle est notamment associée à des cas de leucémies et lymphomes à cellules T mais aussi à des maladies neurologiques (King *et al.*, 2011).

Le génome des deltarétrovirus HTLV contient une région supplémentaire nommée *pX,* située en 3' du gène *env,* qui comporte plusieurs phases de lecture ouvertes (ORF, *Open Reading Frame*) codant des protéines régulatrices et auxiliaires (Rende *et al.*, 2012).

Le promoteur situé dans le LTR 5' ainsi que des mécanismes d'épissage alternatif permettent de transcrire les ARNm codant les protéines régulatrices et auxiliaires (**Figure 4**).

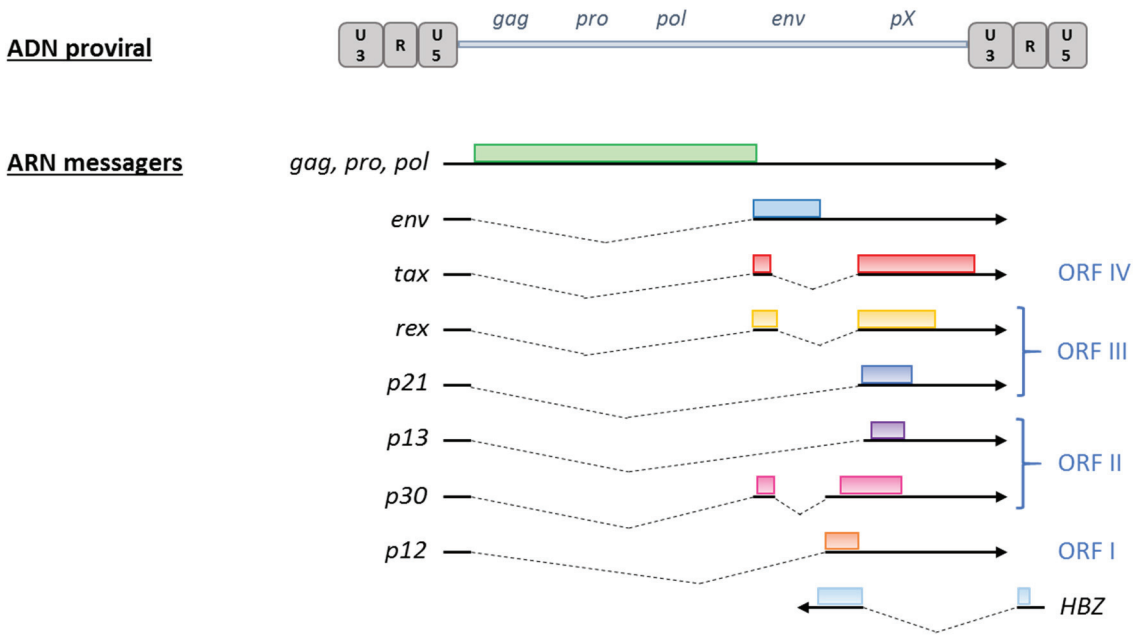

Figure 4 | Organisation génomique du rétrovirus humain HTLV-1. D'après (Matsuoka et Jeang, 2007). Les rectangles représentent les séquences codantes et les pointillés les épissages alternatifs.

Le LTR 3' des HTLV constitue également un promoteur viral qui permet de transcrire l'ARNm codant une protéine régulatrice « anti-sens » et différente selon le type d'HTLV considéré (HBZ dans le cas d'HTLV-1) (**Figure 4**). La structure génomique des HTLV-2, HTLV-3 et HTLV-4 est très proche de celle d'HTLV-1.

1.2.3.2. Les lentivirus : les virus de l'immunodéficience humaine et féline, VIH et FIV

Les lentivirus, responsables de maladies chroniques à évolution lente, ont été découverts chez l'Homme (VIH, virus de l'immunodéficience humaine) mais également chez de nombreuses espèces animales parmi lesquelles les félins (FIV, *Feline Immunodeficiency virus*), les singes (SIV, *Simian Immunodeficiency Virus*) ou encore les bovins (BIV, *Bovin Immunodeficiency Virus*). Ces virus entraînent chez leurs hôtes un syndrome d'immunodéficience dont les signes cliniques ne se manifestent qu'après de nombreux mois, voire des années après l'infection par le virus (King *et al.*, 2011). D'autres lentivirus d'ovidés et d'équidés, eux, n'induisent pas d'immunodéficience mais des maladies inflammatoires et des anémies, respectivement.

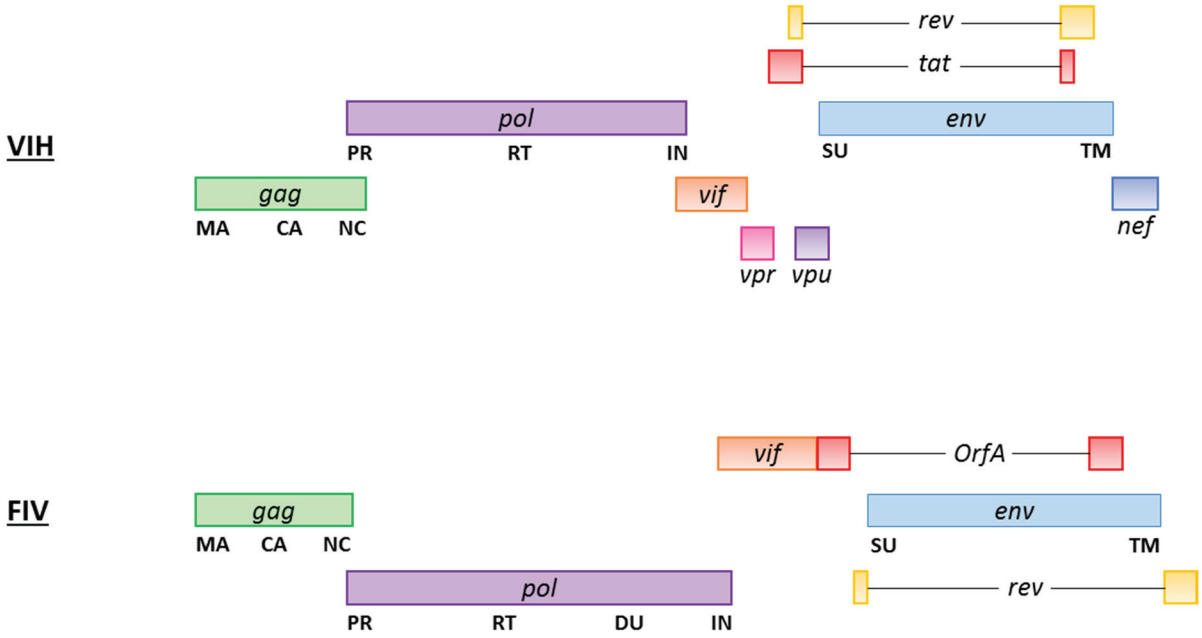

Figure 5 | Organisation génomique des lentivirus VIH et FIV. D'après (Kenyon et Lever, 2011).

A l'instar de celui des deltarétrovirus, le génome des lentivirus possède plusieurs gènes supplémentaires codant des protéines régulatrices et accessoires (**Figure 5**) (Frankel et Young, 1998).

Le VIH, par exemple, qui est le lentivirus le plus décrit et étudié à ce jour, code six protéines supplémentaires dont deux régulatrices et quatre accessoires.

Les protéines régulatrices précoces du VIH, indispensables à la réplication virale, sont impliquées dans la régulation de l'expression virale (Frankel et Young, 1998) :

- La protéine Tat (*Transactivator of transcription*) joue le rôle de transactivateur de la transcription des gènes viraux à partir du promoteur viral ;
- La protéine Rev (*Regulator of expression of virion proteins*) est responsable du transport des ARNm viraux non épissés du noyau vers le cytoplasme.

Les protéines accessoires, elles, sont majoritairement impliquées dans la régulation du cycle réplicatif (Trono, 1995) :

- La protéine Nef (*Negative regulatory factor*) est à l'origine de la survie des cellules infectées et participe à la persistance de l'infection ;
- La protéine Vpr (*Viral protein r*) régule le transport des complexes pré-intégratifs vers le noyau ;
- La protéine Vif (*Viral infectivity factor*) est impliquée dans le mécanisme d'évasion au système de défense de l'hôte en provoquant la dégradation de la cytidine désaminase APOBEC3G, dont le rôle est de limiter la réplication des rétrovirus ;
- La protéine Vpu (*Viral protein u*) participe au bourgeonnement des virions à partir de la membrane de la cellule hôte.

Le génome du FIV, quant à lui, possède une sous-unité supplémentaire dans son gène *pol* codant une enzyme nommée dUTPase (désoxyuridine triphosphatase) (DU) (**Figure 5**). Cette enzyme, essentielle à la réplication du FIV, hydrolyse le dUTP (désoxyuridine triphosphate) en dUMP (désoxyuridine monophosphate) et en pyrophosphate. Elle permet ainsi de limiter la mauvaise incorporation d'uracile dans l'ADN lors de la transcription inverse (Hizi et Herzig, 2015).

De plus, le FIV ne code que trois protéines supplémentaires, à savoir Rev, Vif et OrfA (**Figure 5**). Cette dernière pourrait jouer un rôle dans l'activation de la transcription et la dissémination des virions (Kenyon et Lever, 2011).

1.3. Le cycle réplicatif rétroviral

Lors de l'infection de la cellule hôte, le rétrovirus passe par plusieurs étapes lui permettant de se répliquer en vue de former de nouvelles particules virales, qui pourront à leur tour infecter une nouvelle cellule. L'ensemble de ces étapes est nommé cycle réplicatif rétroviral (**Figure 6**). De même que pour l'organisation de la particule virale, les principales étapes de ce cycle sont communes à tous les rétrovirus (Freed et Mouland, 2006).

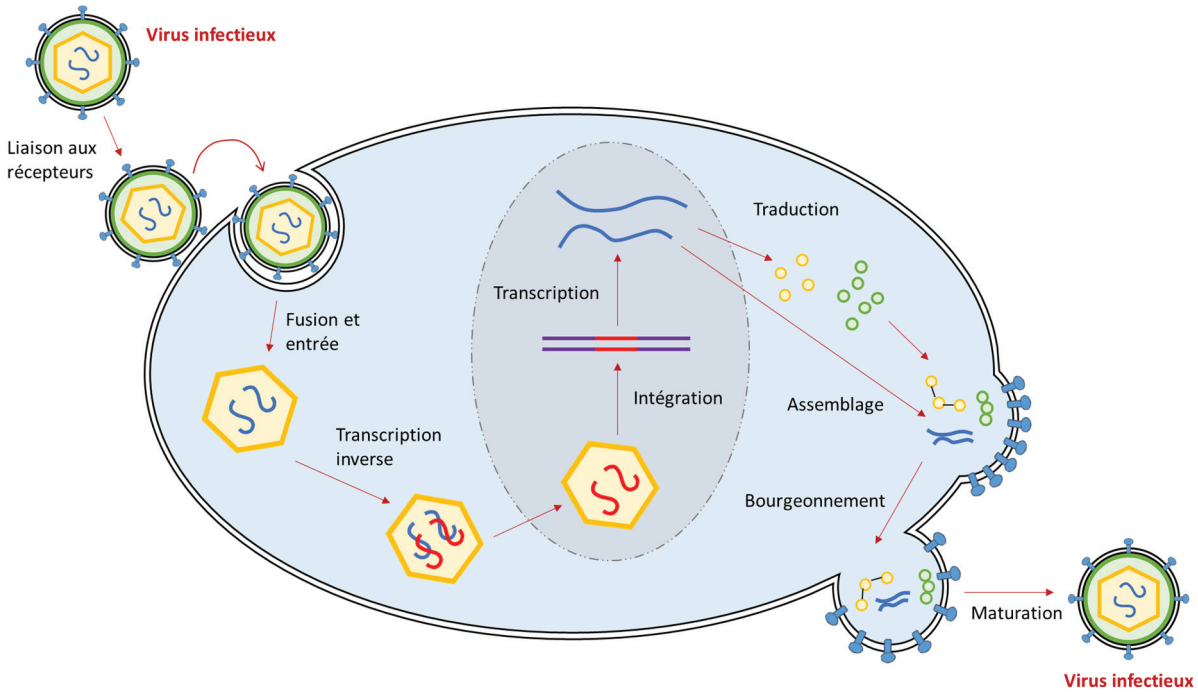

Figure 6 | Le cycle réplicatif rétroviral. Adapté de (Lairmore *et al.*, 2012).

La première étape de ce cycle consiste en l'entrée du virus à l'intérieur de la cellule cible grâce à l'interaction spécifique entre les glycoprotéines d'enveloppe (SU) du rétrovirus et des récepteurs membranaires (et éventuels co-récepteurs) cellulaires (**Tableau 1**). Cette interaction entraîne un changement conformationnel des protéines SU, démasquant les glycoprotéines transmembranaires TM qui vont déclencher la fusion entre l'enveloppe virale et la membrane cellulaire (Kenyon et Lever, 2011).

Suite à la fusion des membranes virales et cellulaires, la capside virale est libérée dans le cytoplasme de la cellule hôte (Wilen *et al.*, 2012) où l'étape de transcription inverse peut débuter. Cette réaction, catalysée par la transcriptase inverse, permet de générer un ADN viral double brin, à partir du génome ARN simple brin du rétrovirus (**Figure 6**).

Rétrovirus		Cellules cibles	Récepteurs cellulaires	Principaux co-récepteurs cellulaires
Oncorétrovirus	**HTLV-1**	Lymphocytes T	GLUT-1 NRP-1 HSPG	
	HTLV-2		GLUT-1 NRP-1	
Lentivirus	**VIH-1**	Lymphocytes T Monocytes Macrophages Cellules dendritiques	CD4	CCR5 CXCR4
	FIV	Lymphocytes T Macrophages Cellules dendritiques	CD134	CXCR4

Tableau 1 | Cellules cibles et récepteurs cellulaires des deltarétrovirus HTLV-1 et -2 et des lentivirus VIH et FIV. HSPG : Heparan Sulfate ProteoGlycans, GLUT-1 : GLUcose Transporter 1, NRP-1 Neuropiline 1, CD4 : Cluster Differenciation 4, CCR5 : C-C Chemokine Receptor type 5 et CXCR4 : Chemokine (C-X-C motif) Receptor type 4.

Après décapsidation, l'ADN viral double brin associé aux protéines de nucléocapside et à l'intégrase est adressé au noyau sous la forme d'un complexe nucléoprotéique dit complexe de pré-intégration. A l'intérieur du noyau, l'ADN viral est ensuite intégré au génome de la cellule cible lors de la réaction d'intégration catalysée par l'intégrase, devenant ainsi un provirus (Kenyon et Lever, 2011).

A la suite de son intégration au génome de la cellule hôte, le provirus est transcrit et traduit grâce à la machinerie cellulaire de l'hôte, notamment grâce à l'ARN polymérase de type II qui permet de transcrire, à partir du promoteur viral, les ARNm codant les protéines virales.

Les protéines Gag exprimées s'accumulent alors à la membrane cellulaire avec le nouveau génome ARN viral produit, ce qui va permettre leur assemblage puis le bourgeonnement, à partir de la membrane plasmique, de nouveaux virions non infectieux (Sundquist et Krausslich, 2012) (**Figure 6**).

Enfin, une dernière étape dite de maturation, où les polyprotéines précurseurs sont clivées par la protéase virale, est nécessaire afin de rendre le virion infectieux (Lee *et al.*, 2012) (Freed, 2015). Cette étape peut entraîner chez certains rétrovirus, dont le VIH et le FIV, un changement de morphologie de la capside virale qui va prendre une forme particulière, conique dans le cas de ces deux virus.

2. Les virus HTLV

2.1. HTLV, premier rétrovirus oncogène humain

2.1.1. Découverte des HTLV

Isolé en 1980 par l'équipe du Dr. Gallo aux États-Unis, le virus HTLV-1 est le premier rétrovirus pathogène pour l'Homme à avoir été découvert (Poiesz *et al.*, 1980) et à être associé à un cancer, la leucémie à cellules T de l'adulte (ATLL, *Adult T-cell Leukemia/Lymphoma*) (Yoshida *et al.*, 1984). Il est également l'agent étiologique d'une maladie neurologique chronique, la myélopathie associée à HTLV-1/paraparésie spastique tropicale (HAM/TSP, *HTLV-1 Associated Myelopathy/Tropical Spastic Paraparesis*) (Gessain *et al.*, 1985).

A ce jour, la famille HTLV comporte trois autres membres en plus d'HTLV-1. HTLV-2, découvert en 1982 (Kalyanaraman *et al.*, 1982), n'est associé à aucun type de cancers mais à des cas de lymphocytoses et à des maladies neurologiques de type HAM/TSP (Roucoux et Murphy, 2004). Plus récemment, deux nouveaux virus apparentés, HTLV-3 et HTLV-4 dont le caractère pathogène n'a pas encore été déterminé, ont été découverts (Mahieux et Gessain, 2011). HTLV-1 reste encore aujourd'hui le seul rétrovirus oncogène humain jamais identifié.

2.1.2. Epidémiologie des HTLV

2.1.2.1. Prévalence des virus HTLV dans le monde

Des études épidémiologiques ont permis d'estimer que 5 à 20 millions de personnes sont infectées par HTLV-1 dans le monde dont 2 à 10% développent, au cours de leur vie, l'une des maladies associées au virus (Gessain et Cassar, 2012).

Les virus HTLV sont répartis dans le monde sous forme de foyers de forte endémie (**Figure 7**). HTLV-1 est majoritairement présent dans le sud-ouest du Japon, dans certaines régions d'Amérique du Sud (Brésil, Guyane et Colombie), en Afrique mais aussi chez quelques populations aborigènes d'Australie. La prévalence d'HTLV-1 est plus élevée chez les femmes et augmente avec l'âge (Verdonck *et al.*, 2007). HTLV-2, quant à lui, est surtout endémique dans certaines populations d'Amérindiens et d'Africains, mais est aussi retrouvé sous une forme épidémique chez les toxicomanes aux drogues intraveineuses en Europe et aux États-Unis (Roucoux et Murphy, 2004). Les virus HTLV-3 et HTLV-4, eux, ont été identifiés dans des populations africaines isolées (Mahieux et Gessain, 2011).

Cependant, la distribution planétaire des HTLV reste assez difficile à évaluer précisément puisque les données épidémiologiques de la Chine, de l'Inde, du Maghreb et de l'Afrique de l'Est sont encore peu connues (Gessain et Cassar, 2012).

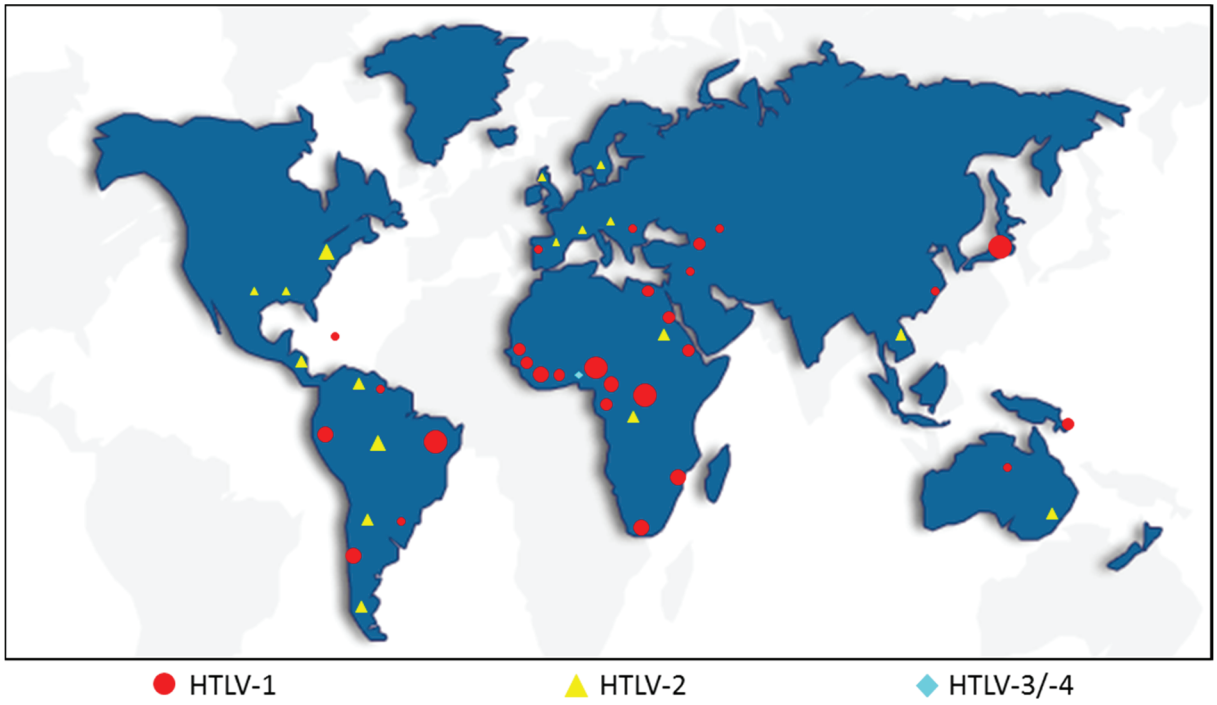

Figure 7 | Répartition des virus HTLV dans le monde sous forme de foyers de forte endémie. D'après (Gessain et Cassar, 2012) (Roucoux et Murphy, 2004).

2.1.2.2. Modes de transmission

Les virus HTLV peuvent être transmis selon trois modes : de la mère à l'enfant, par voie sexuelle ou par voie sanguine (Ferreira *et al.*, 1997). Dans chaque cas, la transmission se fait par transfert de liquides corporels infectés : le lait, la semence et le sang, respectivement.

La transmission mère-enfant se fait majoritairement *via* l'allaitement maternel avec un risque de transmission allant jusqu'à 20% pour les allaitements prolongés (Wiktor *et al.*, 1997) (Percher *et al.*, 2016).

La transmission par voie sexuelle, elle, se fait principalement dans le sens homme → femme avec 50% des transmissions se faisant dans ce sens contre 0,1% pour le cas femme → homme (Roucoux *et al.*, 2005).

Enfin, la transmission par voie sanguine intervient principalement lors d'échanges de seringues et matériels contaminés chez les toxicomanes dans les zones épidémiques.

2.1.2.3. Moyens de prévention et traitements

Les modes de transmission des HTLV étant connus, des moyens de prévention tels que l'allaitement artificiel, les campagnes d'information à l'attention des populations à risque et consommateurs de drogues ou encore l'amélioration des techniques de dépistage, participent à la diminution du risque d'infection. L'utilisation du préservatif, comme pour les autres infections sexuellement transmissibles, reste le moyen de protection le plus efficace contre ce type d'infection.

Cependant, il n'existe pas de traitement spécifique anti-HTLV. Les formes aigües de l'ATLL étant résistantes aux chimiothérapies, avec une rechute dans les mois suivants le traitement, des stratégies antirétrovirales basées sur la combinaison zidovudine/interféron α (AZT/IFN-α) sont utilisées pour lutter contre l'infection par HTLV (Bazarbachi *et al.*, 2010) (Cachat *et al.*, 2013) (Kinpara *et al.*, 2013).

2.2. HTLV, des virus pathogènes plus ou moins virulents

2.2.1. Pathologies associées à HTLV-1

2.2.1.1. La leucémie à cellules T de l'adulte, ATLL

La leucémie à cellules T de l'adulte est une forme agressive de lymphoprolifération pouvant aboutir à des cas de leucémies aigües (Iwanaga *et al.*, 2012). Elle ne concerne qu'environ 5% des personnes infectées et survient généralement en moyenne entre 40 à 60 ans après l'infection par HTLV-1, principalement lors de la transmission par allaitement.

L'ATLL est une maladie à plusieurs stades cliniques définis selon les différents symptômes présentés par le patient (**Tableau 2**) ; les formes aigües (leucémie et lymphome) étant rapidement mortelles. Chez les personnes infectées développant l'ATLL, la majorité des cas présente des formes aigües et donc mortelles de la maladie, caractérisées par une prolifération des lymphocytes T.

A ce jour, seule la combinaison AZT/IFN-α permet de soigner les formes leucémiques, chroniques et indolentes de l'ATLL avec respectivement un taux de survie à 5 ans de 80% pour la première et de 100% pour les deux dernières formes citées (Bazarbachi *et al.*, 2010).

Dans le cas de la forme lymphomateuse de la maladie, les patients sont toujours traités principalement par chimiothérapie (CHOP : doxorubicine, cyclophosphamide, vincristine et prednisone) avec un faible taux de rémission, en raison des rechutes survenant après le traitement.

<table>
<tr><th colspan="2">Formes cliniques de l'ATLL</th><th>Proportion des cas</th><th colspan="2">Symptômes présentés par le malade</th><th>Médiane de survie</th></tr>
<tr><td colspan="2">Indolente</td><td>5%</td><td colspan="2">Lymphocytose et lésions cutanées</td><td>Non reportée</td></tr>
<tr><td colspan="2">Chronique</td><td>20%</td><td colspan="2">Hyperlymphocytose et lésions cutanées</td><td>24 mois</td></tr>
<tr><td rowspan="2">Aigües</td><td>Lymphomateuse</td><td>20%</td><td rowspan="2">Hépatosplénomégalie
Hypercalcémie
Lésions cutanées</td><td>Lymphome T à cellules pléomorphes</td><td>10 mois</td></tr>
<tr><td>Leucémique</td><td>55%</td><td>Hyperlymphocytose
Cellules en « fleur »</td><td>6 mois</td></tr>
</table>

Tableau 2 | Formes cliniques de l'ATLL et symptômes associés. D'après (Verdonck *et al.*, 2007) (Gessain, 2011).

2.2.1.2. La myélopathie associée à HTLV/paraparésie spastique tropicale, HAM/TSP

La neuromyélopathie chronique HAM/TSP est également une maladie de l'adulte dont les premiers signes cliniques se déclarent aux alentours de 40-50 ans, principalement après transmission par voie sexuelle ou sanguine (Gessain et Mahieux, 2012). Les symptômes cliniques majeurs sont une paraparésie spastique (ou paraplégie), faiblesse musculaire touchant la racine des membres inférieurs, des douleurs dorsales, des troubles urinaires ainsi que des dysfonctionnements sexuels (Verdonck *et al.*, 2007). Les lésions induites au niveau de la moelle épinière, dont une démyélinisation suivie d'une dégradation axonale des substances blanche et grise, sont irréversibles (Yamano et Sato, 2012).

Les traitements apportés pour ce type de maladie, tels que les corticostéroïdes, les stéroïdes anabolisants ou encore la vitamine C, apportent uniquement de modestes améliorations des symptômes chez les patients en phase précoce (Goncalves *et al.*, 2010).

Outre l'ATLL et l'HAM/TSP, des uvéites (Takahashi *et al.*, 2000) et des dermatites infectieuses (Lee et Schwartz, 2011) sont aujourd'hui également reconnues comme maladies associées à HTLV-1 (Gessain, 2011).

2.2.2. Pathologies associées à HTLV-2

Tout comme HTLV-1, le virus HTLV-2 est associé à une encéphalomyélopathie chronique similaire au syndrome neurologique HAM/TSP, caractérisée par une faiblesse spastique au niveau des membres inférieurs (Araujo et Hall, 2004). En plus de cette neuromyélopathie chronique, HTLV-2 est également responsable d'autres troubles neurologiques proches de l'HAM avec des symptômes tels que la neuropathie et des troubles mentaux (Roucoux et Murphy, 2004).

Cependant, contrairement à HTLV-1, HTLV-2 n'a été associé à aucun type de leucémies mais seulement à des cas de lymphocytoses (Bartman *et al.*, 2008).

2.3. L'oncoprotéine Tax

Chez HTLV, les deux protéines régulatrices Tax et Rex, exprimées à partir de la région supplémentaire *pX*, jouent un rôle crucial dans l'expression des gènes viraux et sont indispensables à la réplication des HTLV (Ciminale *et al.*, 2014).

La protéine Tax, plus particulièrement, joue un rôle crucial dans l'expression des gènes viraux, en raison de son rôle d'activateur de la transcription virale *via* ses interactions avec différents facteurs cellulaires (Boxus *et al.*, 2008).

2.3.1. Tax, une protéine multifonctionnelle

2.3.1.1. Régulation de l'expression des gènes viraux

La protéine Tax est dite transactivatrice de la transcription, car elle active la transcription des gènes viraux en agissant au niveau de l'élément de réponse à Tax (TxRE, *Tax Responsive Element*) localisé dans la partie U3 de la région LTR 5' du génome d'HTLV (**Figure 8**) (Currer *et al.*, 2012).

Ce TxRE est composé de trois répétitions imparfaites de 21 paires de bases, nécessaires pour la transcription des gènes sous le contrôle du promoteur situé dans le LTR 5'.

L'activation de la transcription est médiée par le recrutement par Tax de son partenaire cellulaire, le facteur de transcription CREB (*Cyclic AMP Response Element Binding-protein*), qui va directement interagir avec l'élément viral de réponse à l'AMPc (vCRE, *viral cAMP Responsive Element*) présent dans le TxRE (Suzuki *et al.*, 1993). Pour cela, Tax forme un homodimère interagissant avec un dimère de CREB (Tie *et al.*, 1996) (**Figure 8**) *via* le domaine de liaison à CREB situé dans la région amino-terminale de Tax (Goren *et al.*, 1995) (**Figure 11**).

En recrutant CREB sur le TxRE, la protéine Tax permet l'activation de la transcription des gènes viraux à partir du promoteur viral du LTR 5'.

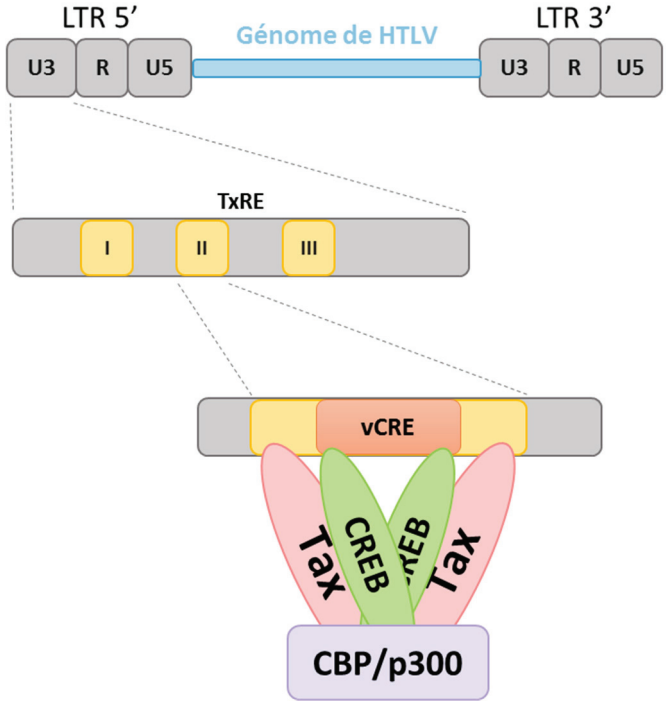

Figure 8 | Interaction entre l'élément de réponse à Tax (TxRE) du LTR 5' du génome de HTLV *via* son domaine vCRE et le complexe Tax:CREB:CBP/p300. D'après (Currer *et al.*, 2012). Les trois répétitions imparfaites de 21 paires de bases du TxRE sont représentées par des carrés jaunes (I, II et III).

Une fois le complexe Tax:CREB:TxRE stabilisé, CREB recrute des histones acétyltransférases telles que la CBP (*CREB-Binding Protein*) et son homologue p300 (Harrod *et al.*, 1998) (**Figure 8**), conduisant à l'ouverture de la chromatine en aval du promoteur.

Ainsi, en recrutant le facteur de transcription CREB permettant d'adresser les histones acétyltransférases CBP/p300 au niveau du promoteur viral pour faciliter l'ouverture de la chromatine, Tax joue un rôle clé dans le cycle réplicatif en favorisant la transcription efficace du génome viral.

2.3.1.2. Régulation de l'expression des gènes cellulaires

La protéine Tax est également directement impliquée dans la pathogénicité des virus HTLV et joue un rôle central dans l'immortalisation, la transformation et la prolifération des cellules T infectées (Feuer et Green, 2005) (Rende *et al.*, 2012). En effet, outre son rôle dans la transactivation de la transcription du promoteur viral, la protéine Tax est également capable de réguler de nombreuses voies cellulaires chez son hôte par différents mécanismes.

- Tax et la voie de CREB/ATF

En plus de moduler l'expression des gènes viraux, Tax peut également réguler l'expression de gènes cellulaires grâce à son interaction avec les facteurs de transcription de la famille CREB/ATF (*Activating Transcription Factor*). Ces facteurs de transcription jouent un rôle crucial dans la croissance et la survie cellulaire ainsi que dans l'apoptose (Currer *et al.*, 2012).

Les protéines Tax peuvent ainsi augmenter l'expression de certains gènes cellulaires comme ceux codant l'interleukine 17 ou les protéines c-fos (Alexandre et Verrier, 1991) (Dodon *et al.*, 2004) mais elles peuvent également réprimer l'expression d'autres gènes. Par exemple, grâce à son interaction avec la voie CREB/ATF, Tax-1 exerce une régulation négative de l'activité de la protéine p53, un suppresseur de tumeur impliqué dans la régulation du cycle cellulaire et l'apoptose, entraînant ainsi l'immortalisation des cellules T infectées (Jeong *et al.*, 2004) (Shirinian *et al.*, 2013).

- Tax et la voie de signalisation du NF-κB

Outre la voie CREB/ATF, Tax peut également augmenter l'expression de nombreux promoteurs hétérologues cellulaires *via* son interaction avec les facteurs de transcription cellulaire de la voie de signalisation du NF-κB (*Nuclear Factor kappa-B*).

La voie de signalisation du NF-κB est impliquée dans la réponse inflammatoire de la cellule, notamment en réponse au stress cellulaire (radiations, dommages UV, ionisation). Elle est aussi impliquée dans la réponse immunitaire lors d'infection par des pathogènes et dans la voie d'activation de la prolifération des lymphocytes en présence d'interleukine-2 (IL-2).

Dans les cellules non activées, les facteurs de transcription de la famille NF-κB sont piégés dans le cytoplasme (**Figure 9**) (Perkins, 2007).

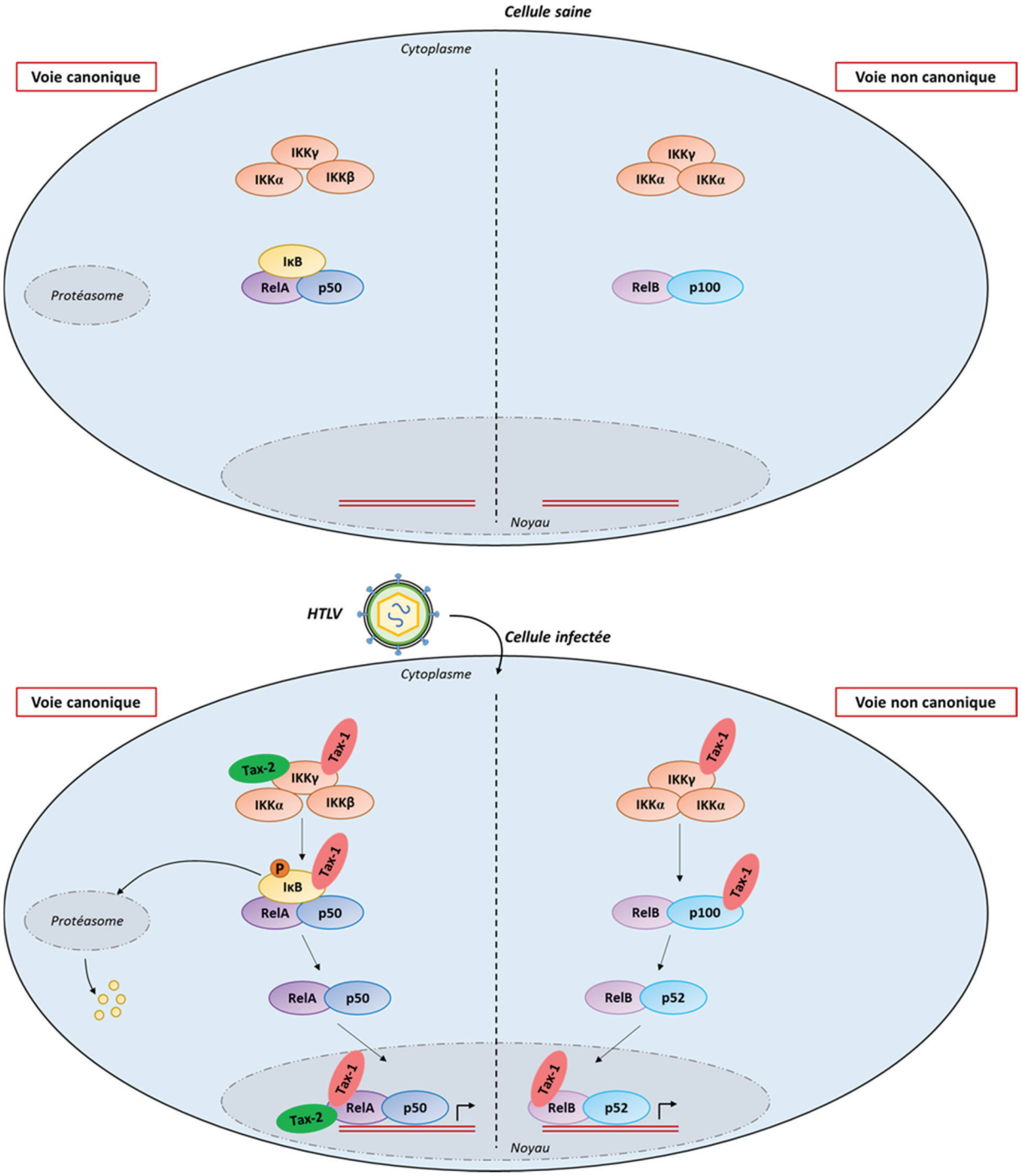

Figure 9 | Activation de la voie de signalisation du NF-κB lors de l'infection par HTLV : interaction entre la protéine Tax et les membres de la famille NF-κB. D'après (Shirinian *et al.*, 2013).

Lors de dommages aux cellules, l'activation de la voie du NF-κB peut se faire selon deux voies distinctes, dites canonique et non canonique :

- La voie canonique du NF-κB implique une activation du complexe IKK (composé de IKKα, IKKβ et IKKγ) qui va phosphoryler l'inhibiteur IκB, induisant son détachement de RelA et permettant la translocation du facteur de transcription RelA dans le noyau.
- La voie non canonique, elle, implique l'activation de IKKα par la kinase NIK (*NF-κB Inducing Kinase*) ce qui va permettre le clivage de la protéine p100 en p52 et la translocation du facteur de transcription RelB dans le noyau.

L'activation de l'expression de gènes cellulaires *via* la voie canonique du NF-κB peut se faire par les protéines Tax-1 et Tax-2, mais seule Tax-1 est capable d'activer la voie non-canonique du NF-κB (**Figure 9**).

En effet, en plus d'interagir avec les protéines de la voie canonique tels que le complexe IKK et RelA, Tax-1 peut également se lier aux protéines p100 et RelB de la voie non canonique du NF-κB. La liaison de Tax-1 à p100 induit son clivage en p52 afin de permettre la translocation de RelB au noyau (**Figure 9**).

La protéine Tax-2, quant à elle, est uniquement capable d'interagir avec IKKγ et RelA de la voie canonique mais ne reconnaît pas la protéine p100 de la voie non canonique (Shirinian *et al.*, 2013) (**Figure 9**). En effet, contrairement à la protéine Tax-1, Tax-2 ne possède pas au sein de sa séquence le domaine permettant de se lier à la protéine p100 (**Figure 11**).

La stimulation de la voie du NF-κB par Tax, lors de l'infection par HTLV, entraîne notamment une augmentation de l'expression d'un facteur de croissance des cellules T, l'interleukine-2, à l'origine de la prolifération des cellules infectées (Higuchi *et al.*, 2007) (Currer *et al.*, 2012).

- <u>Tax et la dérégulation du cycle cellulaire</u>

La protéine Tax-1 est impliquée dans la dérégulation et l'accélération du cycle cellulaire *via* ses interactions avec de nombreuses protéines impliquées dans ce cycle (**Figure 10**) (Currer *et al.*, 2012), telles que les cyclines ou les kinases cycline-dépendantes (CDK, *Cyclin-Dependant Kinase*).

Tax-1 empêche ainsi la mort cellulaire programmée des cellules infectées et promeut leur prolifération et leur immortalisation en perturbant les différents points de contrôle du cycle cellulaire.

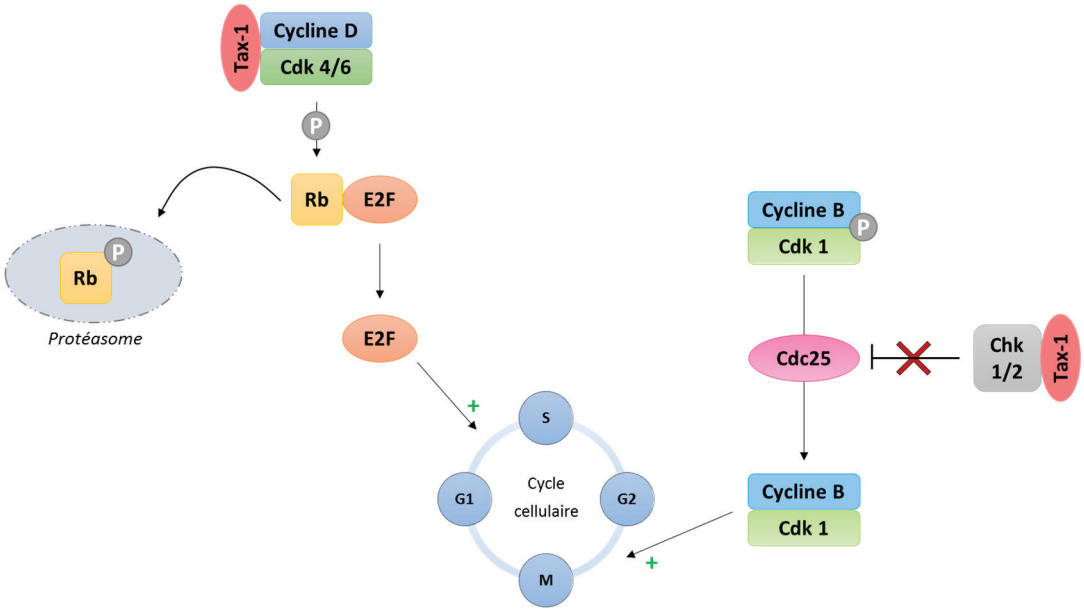

Figure 10 | Effet de Tax-1 sur la progression du cycle cellulaire. Adapté de (Boxus *et al.*, 2008). Grâce à ses nombreuses interactions avec les protéines impliquées dans la régulation des points de contrôle du cycle cellulaire, Tax-1 est capable de déréguler ces derniers et d'accélérer les transitions G1/S et G2/M.

En plus d'augmenter l'expression de certaines cyclines (D et E) et CDK (2, 4 et 6) *via* la voie du NF-κB, Tax-1 interagit directement avec les CDK4 et CDK6 et entraîne la dérégulation de la transition G1/S du cycle (**Figure 10**) (Boxus *et al.*, 2008). En effet, les protéines CDK vont se lier à la cycline D et induire la phosphorylation et l'inactivation de la protéine du rétinoblastome (Rb) du complexe Rb:E2F. La protéine E2F, régulatrice de la transition G1/S, est alors libérée ce qui a pour conséquence une accélération du cycle cellulaire et le passage en phase S des cellules infectées (Currer *et al.*, 2012).

Tax-1 est également capable de forcer l'entrée en phase M du cycle cellulaire en se liant directement aux protéines suppresseurs de tumeurs Chk (1 et 2) (Haoudi *et al.*, 2003) : la régulation négative de l'action de la phosphatase Cdc25 sur le complexe cycline B:CDK1, responsable de la transition G2/M, est alors levée ce qui induit le passage en phase M des cellules infectées (**Figure 10**) (Currer *et al.*, 2012).

2.3.2. Tax, une protéine multi-domaines

La protéine Tax est organisée en plusieurs domaines (**Figure 11**) par lesquels sont médiées ses différentes fonctions.

Elle contient un domaine de liaison à CREB et un domaine d'activation de CREB/ATF à ses extrémités amino- et carboxy-terminale, respectivement, ainsi qu'un domaine de liaison à CBP/p300. Ces domaines lui permettent d'interagir avec le complexe CREB:CBP/p300 (Harrod *et al.*, 1998) et ainsi de moduler l'expression des gènes viraux et cellulaires.

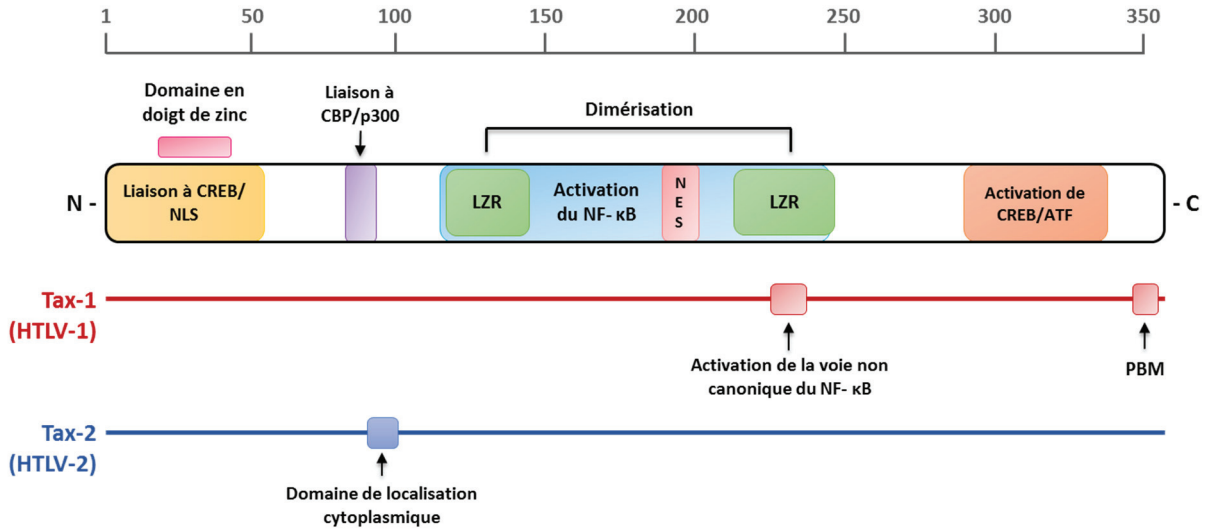

Figure 11 | Domaines fonctionnels des protéines Tax-1 et Tax-2. Adapté de (Romanelli *et al.*, 2013).

Tax comporte également un domaine en doigt de zinc, à son extrémité amino-terminale, impliqué dans la formation du dimère de Tax nécessaire à son activité. Cette dimérisation est aussi médiée par sa région centrale dite de dimérisation (Basbous *et al.*, 2003). Ses deux motifs LZR (*Leucine Zipper-like Region*) sont également impliqués dans sa dimérisation (Jin et Jeang, 1997) mais aussi dans sa liaison à l'ADN et aux différents facteurs cellulaires.

Tax possède un domaine central d'activation de la voie canonique du NF-κB, lui permettant de moduler l'expression des gènes viraux et cellulaires lors de l'infection par HTLV.

Enfin, la localisation de Tax dans le noyau ou le cytoplasme est médiée par la présence dans sa séquence de signaux de localisation nucléaire (NLS, *Nuclear Localisation Signal*) et d'export nucléaire (NES, *Nuclear Export Signal*) (Romanelli *et al.*, 2013).

Malgré des fonctions et des domaines similaires, ainsi que 80% d'identité de séquence, les protéines Tax-1 et Tax-2 présentent quelques différences. La séquence de Tax-1 (HTLV-1) est composée de 353 résidus d'acides aminés tandis que celle de Tax-2 peut varier selon le sérotype d'HTLV-2 de 331 à 356 résidus d'acides aminés.

Le principal caractère distinguant Tax-1 de Tax-2 est la présence d'un motif activant la voie de signalisation non canonique du NF-κB dans Tax-1 (**Figure 11**) (Romanelli *et al.*, 2013). Une seconde distinction entre ces deux protéines réside dans la présence d'un motif de liaison aux protéines à domaine PDZ (PBM, *PDZ Binding Motif*) dans la région carboxy-terminale de Tax-1 (**Figure 11**) (Romanelli *et al.*, 2013).

Tax-2, elle, ne présente pas les deux caractéristiques précédentes mais contient un domaine de localisation supplémentaire nommé « domaine de localisation cytoplasmique » (**Figure 11**) qui augmente sa localisation dans le cytoplasme comparativement à Tax-1 (Meertens *et al.*, 2004).

2.3.3. Tax-1 et Tax-2 : petites différences, grandes conséquences ?

2.3.3.1. Tax-1 et Tax-2, des modifications post-traductionnelles variées

Les protéines Tax subissent de nombreuses modifications post-traductionnelles, qui jouent un rôle important dans leur localisation cellulaire, leur fonction et leurs interactions protéines-protéines (Shirinian *et al.*, 2013).

La phosphorylation et l'acétylation de Tax sont impliquées dans sa fonction d'activation des promoteurs cellulaires par les voies CREB/ATF et NF-κB (Bertazzoni *et al.*, 2011).

La sumoylation et l'ubiquitinylation jouent un rôle important dans la localisation cellulaire et les interactions protéines-protéines de Tax. L'ubiquitinylation est impliquée dans l'interaction de Tax avec la voie du NF-κB au sein du cytoplasme. La sumoylation, elle, est impliquée dans le recrutement des facteurs de transcription du NF-κB (RelA et RelB) à l'intérieur du noyau, où Tax-1 s'accumule préférentiellement sous forme de corps nucléaires (Turci *et al.*, 2009) (Bertazzoni *et al.*, 2011). La protéine Tax-2, elle, bien qu'elle soit également retrouvée dans le noyau (Bertazzoni *et al.*, 2011), est majoritairement localisée dans le cytoplasme (Meertens *et al.*, 2004), en raison de son domaine de localisation cytoplasmique.

2.3.3.2. Tax-1 et l'activation de la voie non canonique du NF-kB

Le caractère principal distinguant Tax-1 de Tax-2 est la présence d'un motif activant la voie de signalisation non canonique du NF-κB, motif localisé entre les résidus d'acides aminés 225 et 232 de Tax-1 (**Figure 11**).

L'implication de Tax-1 dans la voie de signalisation non canonique du NF-κB serait responsable de la modulation de l'expression de gènes impliqués dans l'apoptose et la prolifération cellulaire, ce qui entraînerait la transformation des cellules T à l'origine de l'ATLL (Higuchi *et al.*, 2007). L'activation de cette voie du NF-κB par Tax-1 est, par conséquent, essentielle pour la transformation, la prolifération et la survie des cellules T lors de l'ATLL induite par le HTLV-1 (Shirinian *et al.*, 2013). De plus, cette interaction serait également responsable d'inflammations telles que les uvéites observées chez les patients infectés par HTLV-1.

L'absence d'interaction entre Tax-2 et la voie non canonique du NF-κB pourrait être à l'origine de la différence de pathogénicité entre les virus HTLV-1 et HTLV-2 (Higuchi et Fujii, 2009).

2.3.3.3. Tax-1 et la liaison aux protéines à domaine PDZ

Une seconde distinction entre ces deux protéines réside dans la présence d'un domaine PBM dans la région carboxy-terminale de Tax-1 (**Figure 11**), lui permettant d'interagir avec d'autres facteurs cellulaires contenant un domaine PDZ.

Parmi ces facteurs, on retrouve des protéines régulant la progression du cycle cellulaire et la prolifération cellulaire comme certains suppresseurs de tumeurs dont hDLG (*human Disc Large Protein*) (Suzuki *et al.*, 1999). En se liant à hDLG, Tax empêche l'arrêt du cycle cellulaire des cellules infectées par HTLV induit par hDLG, entraînant ainsi leur prolifération.

Ainsi, la protéine Tax, grâce à ses nombreuses interactions avec les facteurs de transcription des voies de signalisation cellulaires (CREB/ATF, NF-κB) ou encore les protéines régulatrices du cycle cellulaire, joue un rôle clé dans la réplication et la pathogénicité des virus HTLV. L'étude biochimique et structurale de cette protéine, dont la structure reste encore inconnue, permettrait de mieux comprendre les mécanismes impliqués dans ses différentes fonctions mais également conduire à la découverte de nouvelles cibles thérapeutiques contre ce virus.

3. Le virus FIV

3.1. FIV, virus de l'immunodéficience féline

Outre le virus de l'immunodéficience humaine (VIH), responsable du syndrome d'immunodéficience acquise (SIDA) chez l'Homme, les infections par les lentivirus concernent également une partie du règne animal et constituent donc un problème de santé publique majeur dans le domaine vétérinaire.

3.1.1. Découverte du FIV

A l'origine nommé lentivirus T-lymphotropique félin, le virus de l'immunodéficience féline (FIV) a été découvert quatre ans après le VIH, en 1986 à Petaluma en Californie, chez des chats présentant une immunodéficience (Pedersen *et al.*, 1987) (Yamamoto *et al.*, 1988).

Ce virus cause, chez les chats domestiques et les félins sauvages, un syndrome semblable à celui du SIDA chez l'humain, appelé plus communément « SIDA du chat ». Les félins infectés développent un dérèglement progressif de leur fonction immunitaire caractérisé par une déplétion des cellules T CD4+ (Elder *et al.*, 2010).

Parmi les lentivirus n'infectant pas les primates, le FIV est le seul virus capable d'induire un syndrome d'immunodéficience acquise chez son hôte naturel.

3.1.2. Epidémiologie du FIV

3.1.2.1. Prévalence du FIV dans le monde

Il existe 5 sous-types de FIV (de A à E, basés sur l'analyse phylogénétique des séquences des gènes *env* et *gag* (Hayward et Rodrigo, 2010)) infectant les chats dans le monde, avec une prévalence différente selon le continent touché (**Figure 12**) (Hosie et Beatty, 2007). Les sous-types A et B sont prédominants en Amérique du Nord, en Europe, au Japon et en Australie. Le sous-type A est également retrouvé en Afrique du Sud alors que le sous-type B est prévalent en Argentine. Le sous-type C a été identifié au Japon et au Canada tandis que les sous-types D et E, les moins prévalents, sont retrouvés au Japon et en Argentine, respectivement (Hosie et Beatty, 2007) (Yamamoto *et al.*, 2007) (**Figure 12**).

A ce jour, plus de 50 millions de chats domestiques sont infectés par ce virus dans le monde avec une prévalence pouvant atteindre 30% des chats domestiques dans certaines zones géographiques (Bendinelli *et al.*, 1995).

Le FIV a également été observé chez les félidés sauvages pour lesquels la prévalence peut atteindre 70% chez les lions en Afrique, 25% chez les pumas en Amérique et 30% chez les chats sauvages en Eurasie (Troyer *et al.*, 2005).

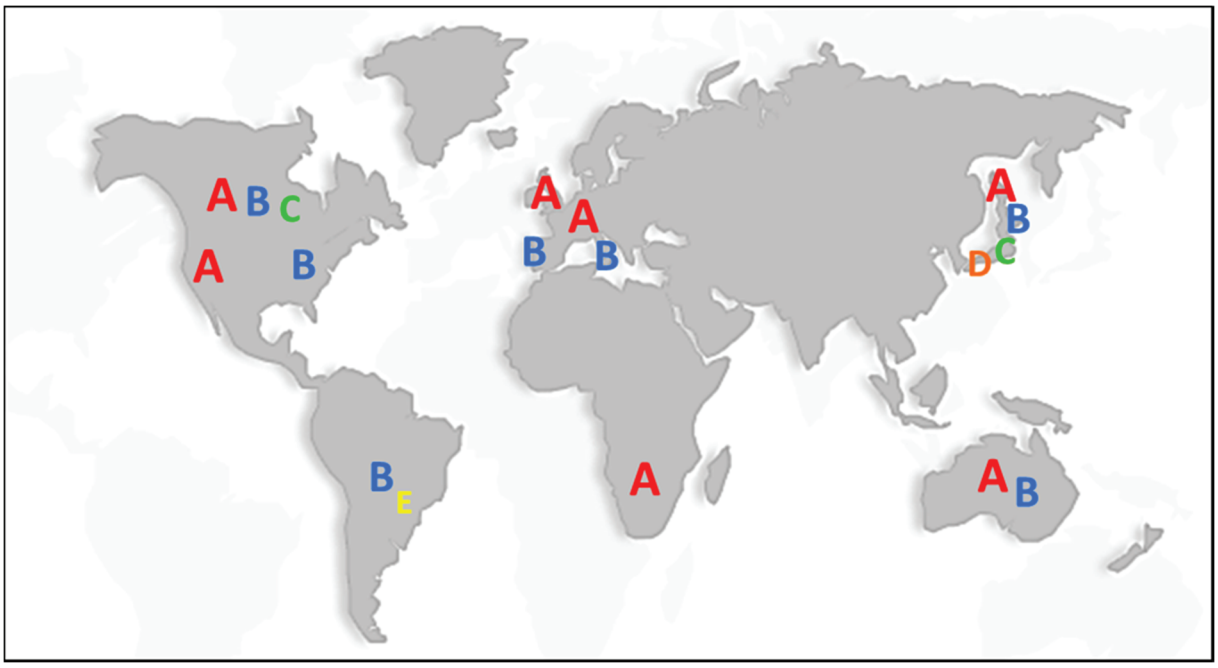

Figure 12 | Répartition sous-types du FIV (A, B, C, D et E) chez les chats domestiques dans le monde. D'après (Hosie et Beatty, 2007) (Yamamoto *et al.*, 2007).

3.1.2.2. Modes de transmission

Le FIV se transmet principalement par la salive, notamment par les morsures entre chats lors de bagarres ou plus rarement lors du toilettage. Il peut également se transmettre, dans de rares cas, au cours de la gestation, lors de la mise-bas et par l'allaitement.

3.1.2.3. Moyens de prévention et traitements

Afin de prévenir l'infection par le FIV, il est recommandé d'éviter tout contact entre chats sains et infectés mais également de faire stériliser les chats afin de diminuer le risque de bagarres.

Dans le cadre de la lutte contre le FIV, de nombreuses thérapies antirétrovirales ont été testées, notamment celles utilisées dans le cas de l'infection par le VIH chez l'Homme. Des inhibiteurs des enzymes virales indispensables au cycle de réplication, tels que l'AZT pour la transcriptase inverse (Gomez *et al.*, 2012), ont été testés chez les chats présentant le syndrome du SIDA. Même si l'efficacité de ces traitements et notamment de trithérapies a été démontré

in vitro (Bisset *et al.*, 2002), les phénomènes de résistance des souches à leur action thérapeutique sont de plus en plus importants.

Un vaccin nommé Fel-O-Vax®FIV (Fort Dodge), contenant les souches Petaluma (sous-type A) et Shizuoka (sous-type D) inactivées est actuellement commercialisé aux États-Unis, au Canada et en Australie. Il protégerait les félins contre les sous-types A, B et D du FIV avec un taux d'efficacité allant de 40% à 100% selon les sous-types (Hayward et Rodrigo, 2010) (Coleman *et al.*, 2014).

3.2. FIV, agent du syndrome d'immunodéficience acquise féline

Le FIV induit chez son hôte un syndrome d'immunodéficience acquise, le « SIDA du chat ». À l'image du SIDA humain, cette maladie immunitaire est caractérisée par plusieurs stades cliniques déterminés par les symptômes observés et leur sévérité (**Tableau 3**) (Moraillon, 2000). Les stades 4 et 5 (pré-SIDA et SIDA) sont associés à des infections opportunistes secondaires graves avec une durée de vie inférieure à un an après le passage en phase 5-SIDA.

<table>
<tr><th colspan="2">Stades</th><th>Durée estimée</th><th colspan="2">Symptômes</th></tr>
<tr><td>1</td><td>Primo-infection</td><td>2 mois</td><td colspan="2">Syndrome mononucléosique</td></tr>
<tr><td>2</td><td>Asymptomatique</td><td>5 à 10 ans</td><td colspan="2">Retour à l'état asymptomatique</td></tr>
<tr><td>3</td><td>LPG</td><td>Quelques mois à années</td><td colspan="2">Perte de poids
Changements comportementaux
Lymphadénopathie généralisée</td></tr>
<tr><td>4</td><td>Pré-SIDA</td><td>Quelques mois à années</td><td rowspan="2">Gingivites
Amaigrissement
Rhinites et conjonctivites
Abcès et dermites
Anémie
Leucopénie</td><td>Infections bactériennes</td></tr>
<tr><td>5</td><td>SIDA</td><td>1 à 6 mois</td><td>Infections opportunistes et signes généraux graves</td></tr>
</table>

Tableau 3 | Stades cliniques de l'infection par le FIV et symptômes associés. D'après (Moraillon, 2000).

3.3. La protéine de capside

Tout comme les autres lentivirus, FIV s'assemble à la surface interne de la membrane plasmique cellulaire puis bourgeonne pour former de nouveaux virions. Les déterminants de cet assemblage sont contenus dans la polyprotéine Gag, indispensable à l'assemblage et à la maturation des nouvelles particules virales (Kenyon et Lever, 2011).

La protéine de capside (CA), sous-unité de la polyprotéine Gag, en plus de sa fonction de protection du génome viral (**Figure 2**), joue un rôle important dans l'étape de formation des virions matures et dans le potentiel infectieux du virus. En effet, elle est impliquée dans les étapes tardives du cycle de réplication viral comme l'assemblage, le bourgeonnement et la maturation (Freed, 2015).

Les sous-unités de la protéine Gag du VIH-1 étant les plus étudiées à ce jour d'un point de vue structural, elles seront utilisées comme références dans cette partie bibliographique : les spécificités de celles du FIV seront mentionnées le cas échéant.

3.3.1. La CA, moteur de l'assemblage de la particule virale

3.3.1.1. Gag, précurseur de la CA

La polyprotéine Gag du VIH est codée par le gène *gag* (**Figure 3**). Cette protéine, naturellement myristoylée, est composée de trois sous-unités majeures, la matrice (MA), la capside (CA) et la nucléocapside (NC), ainsi que de deux peptides espaceurs, sp1 et sp2, et d'un peptide final p6 (**Figure 13A**). Ces différents domaines sont directement impliqués dans la formation de nouveaux virions (Luttge et Freed, 2010) (Freed, 2015) (**Figure 13B** et **Figure 13C**) :

- La MA permet l'adressage de la polyprotéine Gag à la membrane plasmique et son interaction avec celle-ci grâce à son extrémité amino-terminale N-myristoylée.
- La CA participe à l'assemblage de Gag et à la formation de la capside conique des lentivirus.
- Le peptide espaceur sp1 est également impliqué dans l'assemblage de Gag.
- La NC participe à l'encapsidation du génome viral en servant de chaperon à l'ARN.
- Le peptide espaceur sp2 semble important pour le clivage de la jonction NC-p6, lui-même essentiel à la régulation de la maturation virale (de Marco *et al.*, 2012).
- Le peptide carboxy-terminal p6 est impliqué dans le bourgeonnement de la particule virale.

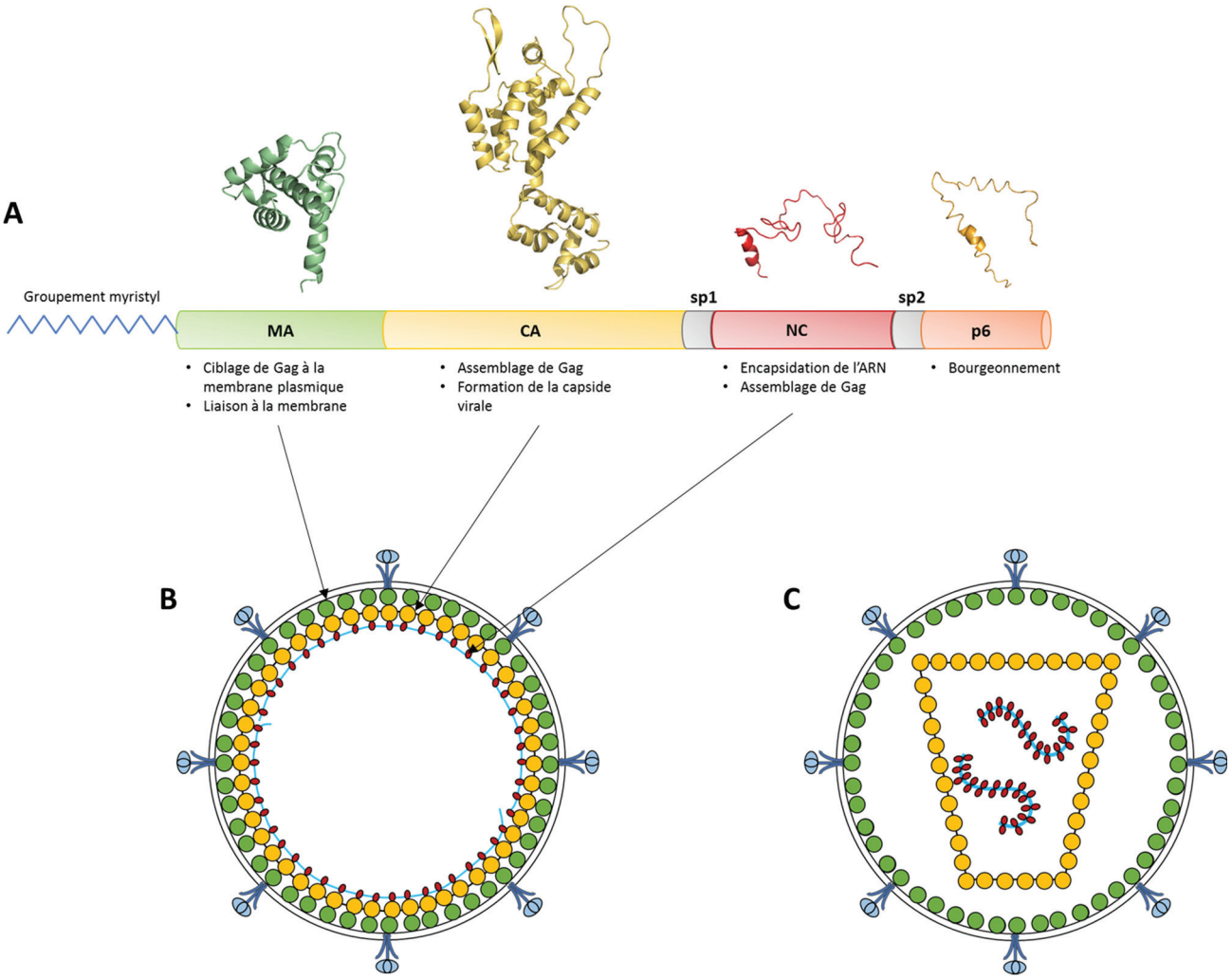

Figure 13 | La polyprotéine Gag : assemblage et maturation de la particule virale du VIH. (A) Structure et fonctions de Gag. D'après (Freed, 2015). Les structures à haute résolution des différents domaines composant Gag sont présentées au-dessus de sa représentation schématique. Les identifiants PDB suivants ont été utilisés : 1HIW (MA) (Hill *et al.*, 1996), 3NTE (CA) (Du *et al.*, 2011), 1F6U (NC) (Amarasinghe *et al.*, 2000) et 2C55 (protéine p6) (Fossen *et al.*, 2005). **(B) et (C) Structures schématisées d'un virion immature (B) et d'une particule virale mature (C).** D'après (Cortines *et al.*, 2011).

Le domaine carboxy-terminal p6 du VIH n'est pas présent au sein de la polyprotéine Gag du FIV : son rôle est assuré par son peptide carboxy-terminal sp2 (Luttge *et al.*, 2008), fonctionnellement équivalent à la protéine p6 du VIH. De plus, la protéine de nucléocapside (NC) du FIV, avec une masse moléculaire d'environ 13 kDa, apparaît être plus grande que celle du VIH.

3.3.1.2. Assemblage de la particule virale

Lors de l'assemblage viral, la polyprotéine précurseur Gag et le génome se condensent à la surface membranaire. Les protéines Gag, ancrées à la membrane plasmique *via* l'extrémité N-myristoylée de leur domaine MA (Dick et Vogt, 2014), interagissent alors entre elles et s'assemblent sous forme de multimères (**Figure 13B**).

Cette multimérisation des protéines Gag est dépendante de la présence de l'ARN génomique et de son interaction avec la protéine de nucléocapside (Mailler *et al.*, 2016), aussi bien dans le cas du VIH que dans celui du FIV (Alfadhli *et al.*, 2005) (Affranchino et Gonzalez, 2010).

Les interactions Gag-Gag font intervenir la protéine de capside dont les domaines amino-terminal (CA_{NTD}) et carboxy-terminal (CA_{CTD}) contribuent à la multimérisation de Gag, conduisant ainsi à la formation de virus immatures (Abdusetir Cerfoglio *et al.*, 2014). En effet, la CA possède une interface de dimérisation CA-CA dans son domaine carboxy-terminal, impliquée dans les interactions Gag-Gag, ainsi qu'une région d'homologie majeure (MHR, *Major Homology Region*) importante pour l'assemblage viral (Dorfman *et al.*, 1994).

3.3.1.3. Maturation de la particule virale

Après assemblage, les virions immatures qui ont bourgeonné passent par une étape de maturation afin d'acquérir leur pouvoir infectieux.

Cette étape fait intervenir de nombreux réarrangements morphologiques et structuraux déclenchés notamment par le clivage de la polyprotéine Gag en ses principaux constituants : MA, CA et NC (Elder *et al.*, 1993) (**Figure 13C**).

Alors que la MA reste ancrée à la membrane par son extrémité N-myristoylée, les protéines CA et NC se condensent autour du génome viral. La CA s'arrange alors spontanément en une capside virale de forme conique, caractéristique du VIH et du FIV, autour du complexe ribonucléique, formé par l'association entre la NC et l'ARN viral (Briggs *et al.*, 2003) (Kenyon et Lever, 2011).

Cette capside virale conique dite mature est composée de plus de 1000 monomères de CA (Pornillos *et al.*, 2011). Elle renferme ainsi l'ARN viral et les protéines structurales des rétrovirus ; ce qui lui confère son rôle de protecteur du génome viral.

3.3.2. La CA, une organisation structurale conservée chez les lentivirus

Les protéines de capside lentivirales comportent environ 230 résidus d'acides aminés (231 dans le cas du VIH) et sont composées de deux domaines repliés en hélices α, le domaine amino-terminal (CA_{NTD}) et le domaine carboxy-terminal (CA_{CTD}), séparés par une boucle flexible (**Figure 14**).

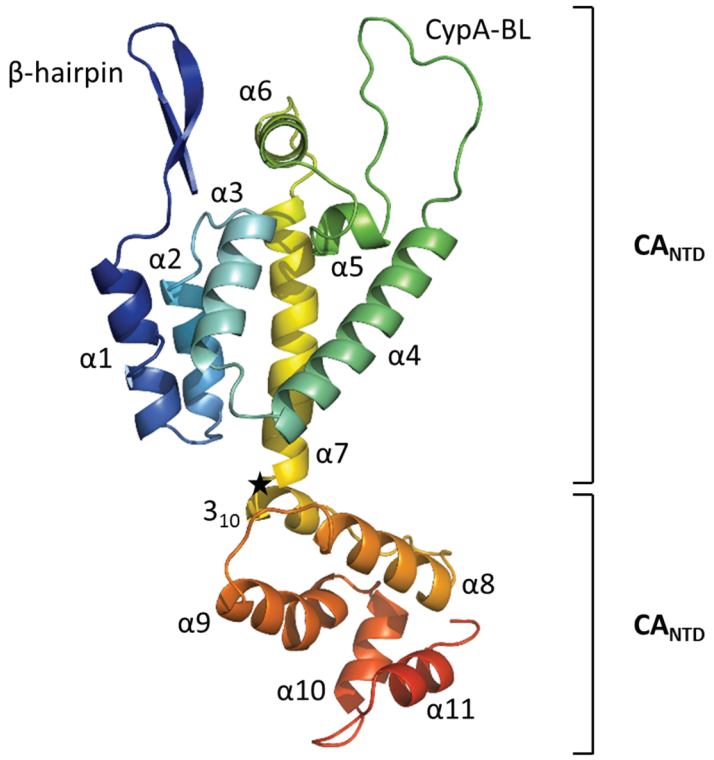

Figure 14 | Structure cristallographique de la protéine de capside du VIH-1 (PDB ID : 3NTE) (Du *et al.*, 2011). La protéine de capside du VIH-1 est composée de deux domaines repliés en hélices α, CA_{NTD} et CA_{CTD}, séparés par une boucle flexible (étoile noire).

3.3.2.1. Le domaine amino-terminal

Le domaine amino-terminal de la CA, composé d'environ 150 résidus d'acides aminés, joue un rôle important dans la formation du cœur protégeant le génome viral et dans la décapsidation de ce génome pour l'étape de transcription inverse.

Ce domaine comprend un motif en épingle-à-cheveux (*β-hairpin*) à son extrémité amino-terminale dans le cas du VIH, 7 hélices α (α1 à α7) ainsi qu'une boucle de liaison à la cyclophiline A (CypA-BL, *Cyclophilin A-Binding Loop*) (**Figure 14**) (Du *et al.*, 2011).

Le motif *β-hairpin*, observé à l'extrémité amino-terminale du CA_{NTD} du VIH-1, est formé par l'établissement d'un pont salin entre le groupement NH^{2+} du résidu prolyl 1, conservé à l'extrémité amino-terminale des CA lentivirales, et la chaîne latérale du résidu aspartyl 51. La Pro1 a par ailleurs été décrite comme essentielle à la formation de la capside virale (Cortines *et al.*, 2011). Le motif *β-hairpin* joue un rôle important dans l'assemblage capsidique lors de la maturation des virions, et plus particulièrement dans la stabilisation de l'orientation des protéines de capside lors du clivage de la polyprotéine Gag (Cortines *et al.*, 2011).

La boucle CypA-BL, elle, est conservée chez les lentivirus et permet à la capside *via* son résidu Pro90 d'interagir avec la cyclophiline A (Goldstone *et al.*, 2010), une *cis-trans* prolyl isomérase permettant l'isomérisation de la liaison peptidique entre une proline et le résidu d'acide aminé voisin. L'isomérisation de *trans* à *cis* de cette liaison peptidique entrainerait une déstabilisation de la CypA-BL et, par conséquent, de la capside virale, ce qui favoriserait la décapsidation du génome virale qui pourra ensuite être rétro-transcrit (Fassati, 2012) (Qi *et al.*, 2008).

3.3.2.2. Le domaine carboxy-terminal

Le domaine carboxy-terminal de la CA, composé d'environ 80 résidus d'acides aminés, est directement impliqué dans l'oligomérisation de la protéine de capside et, par conséquent, dans l'assemblage des nouvelles particules virales.

Dans le cas du VIH-1, ce domaine contient 4 hélices α (α8 à α11) ainsi qu'une hélice 3_{10}, faisant suite à la boucle flexible le séparant du CA_{NTD} (**Figure 14**). Il contient également la région MHR conservée chez tous les rétrovirus.

Cette région MHR, comprise dans l'hélice α8, est importante pour les étapes d'assemblage et de maturation des particules virales. Elle est notamment impliquée dans le repliement correct du domaine CA_{CTD} (Bocanegra *et al.*, 2015) et le réarrangement des CA en une capside virale de forme conique (Mammano *et al.*, 1994), étape essentielle à l'acquisition du pouvoir infectieux des virus VIH et FIV (Dorfman *et al.*, 1994).

Le CA_{CTD} est capable de se dimériser *via* son interface dite de dimérisation CA-CA située au niveau de l'hélice α9 (Gamble *et al.*, 1997). Cette interface est directement impliquée dans la mulimérisation des protéines Gag nécessaires à l'assemblage viral lors du bourgeonnement. Elle fait notamment intervenir deux résidus de cystéines conservés chez les lentivirus (Cys198/Cys218 pour le VIH, Cys190/Cys210 pour le FIV), apparaissant comme

indispensables pour l'assemblage capsidique (Nath et Peterson, 2001). Ces deux cystéines sont impliquées dans un pont disulfure intramoléculaire stabilisant la conformation de la CA.

Ainsi, les domaines CA_{NTD} et CA_{CTD} présentent toutes les interfaces d'interaction nécessaires à l'assemblage de la capside virale immature et aux réarrangements structuraux ayant lieu lors de la formation des particules virales matures.

3.3.3. La CA, oligomérisation et formation de la capside virale

La formation de la capside virale est une étape essentielle à la maturation des lentivirus. Cette capside est formée, comme pour tous les rétrovirus, d'un réseau de protéines CA hexamériques dans lequel sont incorporées des CA sous forme de pentamères (**Figure 15**) (Zhao *et al.*, 2013), qui vont permettre de refermer la capside selon une forme donnée. La forme adoptée par la capside est dépendante du nombre de pentamères de CA incorporés à l'intérieur de celle-ci (Nguyen *et al.*, 2006) : ainsi, l'insertion de 12 pentamères de CA dans la capside du VIH est à l'origine de son architecture conique (**Figure 15A**).

Au sein de la capside virale du VIH, les domaines CA_{NTD} se trouve à la face externe du cône tandis que les domaines CA_{CTD} sont orientés vers l'intérieur de celui, quelle que soit la forme oligomérique de la CA (**Figure 15**). Au sein de ces deux formes oligomériques (hexamères et pentamères) les domaines CA_{NTD} forment un anneau compact central (**Figure 15B** et **Figure 15C** en bleu) bordé par un anneau étendu de domaines CA_{CTD} (**Figure 15B** et **Figure 15C** en rouge) (Pornillos *et al.*, 2011).

Ce sont les interactions entres les domaines CA_{NTD} et CA_{CTD} lors de l'oligomérisation des CA qui permettent de stabiliser la formation de la capside virale. Ces interactions mettent notamment en jeu des interfaces intra-hexamères et intra-pentamères CA_{NTD}-CA_{NTD} et CA_{NTD}-CA_{CTD} entre les monomères de CA adjacents et des interfaces inter-hexamères CA_{CTD}-CA_{CTD} entre deux hexamères adjacents (Gres *et al.*, 2015).

Les interactions CA_{NTD}-CA_{NTD}, au sein des hexamères et pentamères, mettent en jeu de nombreux contacts hydrophobes entre l'hélice α2 d'un monomère et les hélices α1 et α3 du monomère adjacent. Ces interactions conduisent au rapprochement des CA_{NTD} à l'intérieur des oligomères (**Figure 15B** et **Figure 15C**) (Gres *et al.*, 2015), favorisant la formation de l'anneau central de CA_{NTD} (Pornillos *et al.*, 2011).

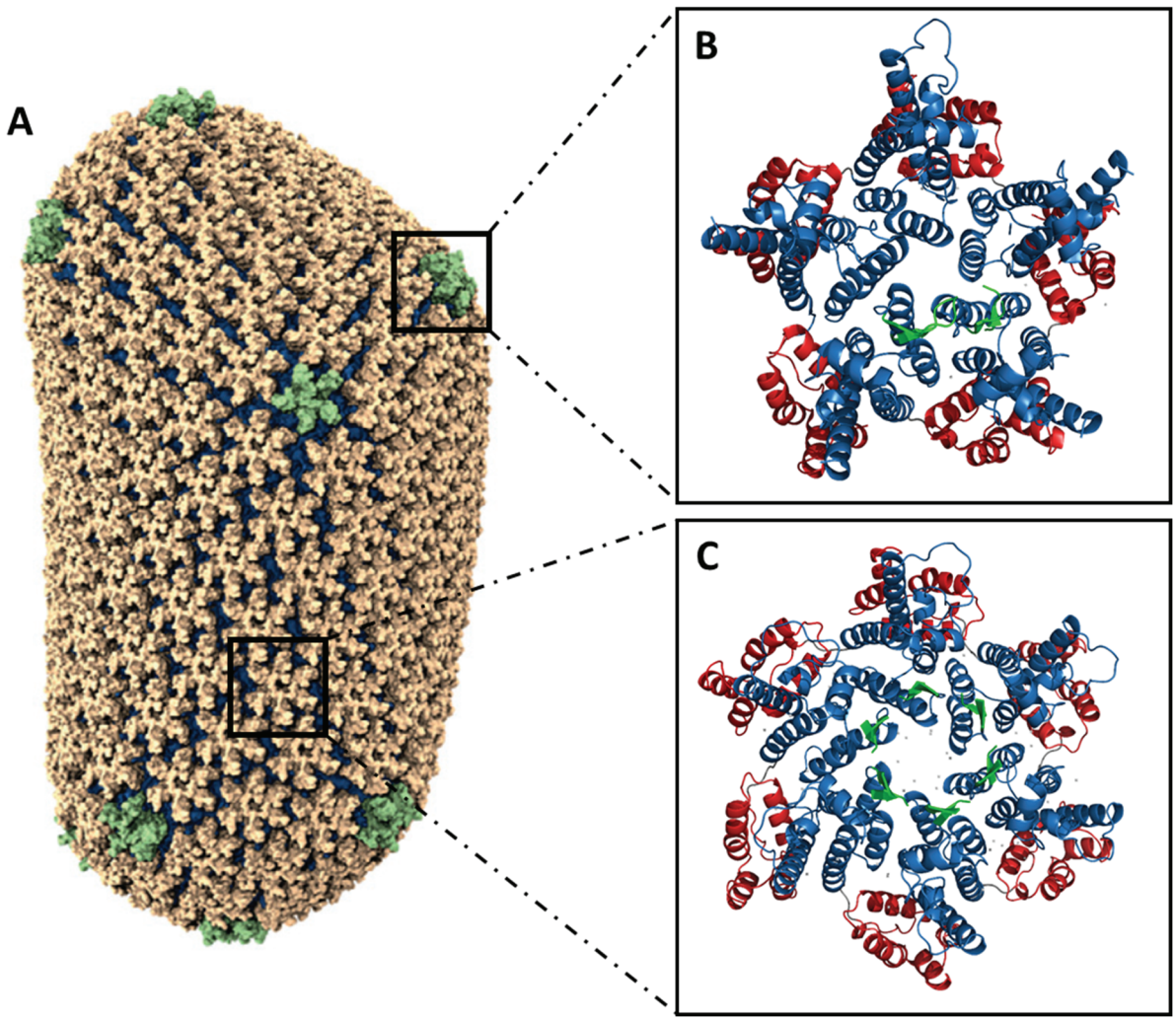

Figure 15 | Capside virale et oligomérisation de la protéine de capside (CA) du VIH-1.

(A) Modèle pseudo-atomique de la capside du VIH-1 composée de 288 hexamères (beige) et de 12 pentamères (vert) de protéine de capside (Perilla et Gronenborn, 2016). Les domaines CA_{NTD} et CA_{CTD} des **(B)** pentamères (PDB ID : 3P05) (Pornillos *et al.*, 2011) et **(C)** hexamères (PDB ID : 4XFX) (Gres *et al.*, 2015) sont colorés en bleu et rouge, respectivement, et leurs motifs *β-hairpin* sont représentés en vert.

Dans le cas des hexamères, l'anneau central de CA_{NTD} forme un pore par lequel sont importés les nucléotides nécessaires à la synthèse de l'ADN lors de l'étape de transcription inverse (Jacques *et al.*, 2016). En effet, le motif *β-hairpin* situé à l'extrémité amino-terminale de la CA peut adopter deux conformations alternées dites « ouverte » et « fermée » qui induisent la formation d'un pore au centre de l'hexamère (**Figure 16**). Ces différentes conformations sont notamment induites par le mouvement des résidus Pro1 et His12 du *β-hairpin*.

De plus, le pH de la solution de cristallisation est également impliqué dans le mouvement du *β-hairpin* et la formation du pore : un pH acide favorisant la conformation ouverte du *β-hairpin* et un pH basique sa conformation fermée (Jacques *et al.*, 2016).

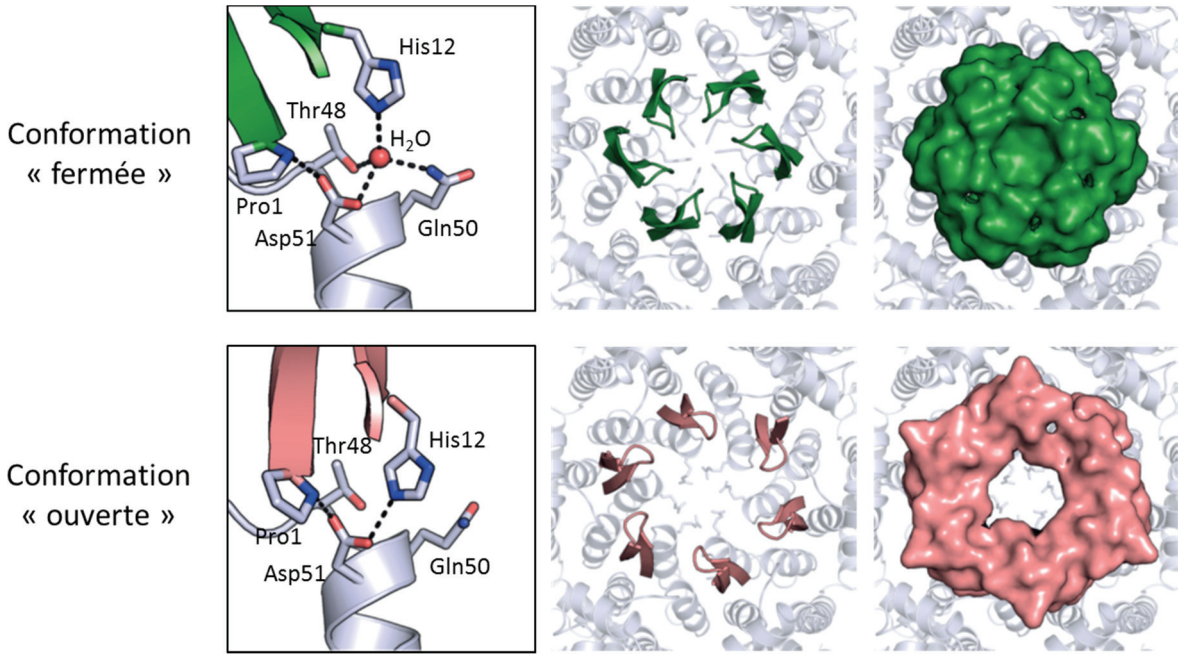

Figure 16 | Conformations « fermée » (vert) et « ouverte » (rose) du *β-hairpin* permettant la formation d'un pore au centre de la CA hexamérique (Jacques *et al.*, 2016).

Les contacts intra-hexamères et intra-pentamères CA_{NTD}-CA_{CTD}, eux, permettent de maintenir les CA_{CTD} de l'anneau externe en contact avec l'anneau central. En outre, grâce à sa flexibilité, cette interface est à l'origine des courbures adoptées par la capside lors de son assemblage conique (Pornillos *et al.*, 2011).

Enfin, les interactions inter-hexamères CA_{CTD}-CA_{CTD} font intervenir le domaine de dimérisation du CA_{CTD}, principalement l'hélice α9, (Pornillos *et al.*, 2010) de deux hexamères adjacents (**Figure 17**). De plus, la présence de molécules d'eau et de contacts hydrophiles à l'interface inter-hexamérique contribue à la stabilisation de la capside virale (Gres *et al.*, 2015).

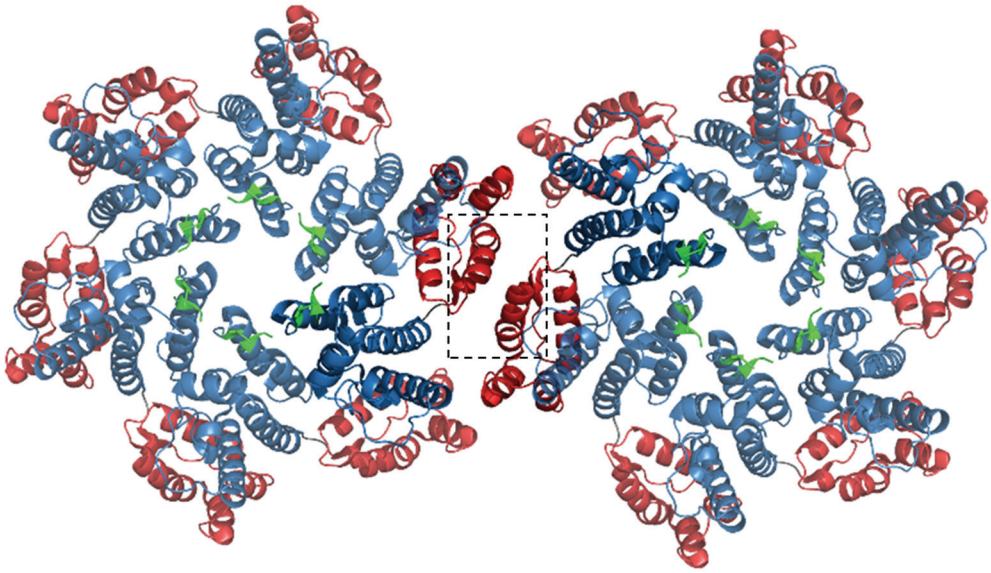

Figure 17 | Interface inter-hexamère CA_{CTD}-CA_{CTD} impliquant les hélices α9 de la CA du VIH (PDB ID : 4XFX) (Gres *et al.*, 2015) . Les domaines CA_{NTD} et CA_{CTD} de la CA sont colorés en bleu et rouge, respectivement, et les motifs *β-hairpin* sont représentés en vert. Les hélices α9 sont encadrées en pointillés noirs.

Bien que les effets physiques/biochimiques permettant la transition entre hexamères et pentamères n'aient pas encore été démontrés, il est probable que de subtils changements conformationnels interviennent au niveau des trois interfaces décrites précédemment, autorisant ainsi la transition entre les deux états oligomériques présents dans la capside virale (Perilla et Gronenborn, 2016).

3.4. Développement d'inhibiteurs dirigés contre l'assemblage de la CA

Pour lutter contre les infections rétrovirales, les efforts actuels portent sur la conception de composés antirétroviraux visant les protéines virales telles que la transcriptase inverse, la protéase ou encore les protéines d'enveloppe. Cependant, les phénomènes de résistance à ce type de thérapie deviennent de plus en plus notables et de nouvelles cibles thérapeutiques doivent être identifiées afin de compléter les multi-thérapies existantes (Uhl *et al.*, 2002) (Hosie et Beatty, 2007).

La protéine de capside, essentielle à la formation de particules virales infectieuses, constitue une cible intéressante pour le développement de nouvelles stratégies antirétrovirales. De plus, les nombreuses interfaces entre les CA formant la capside virale (CA_{NTD}-CA_{NTD}, CA_{NTD}-CA_{CTD} et CA_{CTD}-CA_{CTD}), essentielles pour l'architecture d'une capside mature, représentent des cibles d'intérêt pour le développement de composés inhibant l'assemblage viral (Spearman, 2016).

Des méthodes de criblage *in vitro* de molécules inhibitrices de l'assemblage la CA du VIH et de criblage *in silico* de molécules liant la CA, à partir de sa structure 3D, ont permis d'identifier de nombreux composés potentiellement inhibiteurs de l'assemblage capsidique du VIH (**Tableau 4**) (Thenin-Houssier et Valente, 2016).

Composés	Cibles	Effets	Étape du cycle réplicatif ciblée
CAP-1	CA_{NTD}	Inhibition des interactions CA_{NTD}-CA_{CTD}	Assemblage
BD-1		Interfèrent avec la formation des interfaces CA_{NTD}-CA_{CTD}	Assemblage
BM-1			Maturation
CAI	CA_{CTD}	Inhibent les interfaces CA_{NTD}-CA_{CTD}	Assemblage et maturation
NYAD-1*			
CAC1		Mime l'hélice α9 et inhibe les interactions CA_{CTD}-CA_{CTD}	Assemblage
PF74*	CA_{NTD} et CA_{CTD}	Interférence avec les interactions CA_{NTD}-CA_{CTD}	Assemblage
BMV	Site de clivage CA-sp1 de la polyprotéine Gag	Inhibition du clivage de Gag par la protéase virale	Maturation

Tableau 4 | Molécules ciblant la CA du VIH pour l'inhibition de l'assemblage et de la maturation des particules virales (Thenin-Houssier et Valente, 2016). Les astérisques (*) indiquent les molécules intervenant également dans l'inhibition d'une étape précoce du cycle réplicatif, à savoir la décapsidation virale. BD-1 : benzodiazepine-1 et BM-1 : benzimidazole-1.

Parmi les composés ciblant la CA du VIH, les plus étudiés actuellement sont le composé PF74, le peptide CAI (*Capsid Assembly Inhibitor*), et le composé BVM (Bevirimat) (Thenin-Houssier et Valente, 2016) :

- Le PF74, qui est l'inhibiteur de capside rétrovirale le plus étudié, cible l'assemblage de la capside virale. En se fixant à l'interface CA_{NTD}-CA_{CTD} (hélices α3, α4, α5 et α7 du CA_{NTD} et hélices α8 et α9 du CA_{CTD} (Gres *et al.*, 2015) située entre deux monomères de CA adjacents au sein d'un hexamère, le PF74 déstabilise la formation des particules virales.

- Le peptide CAI, contenant 12 résidus d'acides aminés (ITFEDLLDYYGP), inhibe efficacement l'assemblage capsidique *in vitro* en se fixant dans la poche hydrophobe formée par les hélices α8, α9 et α11 du CA_{CTD}. Cette poche étant impliquée dans l'interaction avec l'hélice α4 du CA_{NTD}, il semblerait que le CAI inhibe les interfaces CA_{NTD}-CA_{CTD} formées par la CA, déstabilisant ainsi la formation des hexamères.
- Le Bevirimat, un dérivé naturel de plantes, est le seul inhibiteur capsidique impliqué dans des tests de phase clinique II. Il inhibe la maturation du VIH avec un mode d'action différent de celui des autres inhibiteurs capsidique. En effet, ce composé cible la jonction CA-sp1 de la polyprotéine Gag, empêchant ainsi la dernière étape de son clivage par la protéase virale. Cette action résulte en l'accumulation de la sous-unité CA-sp1 et stabilise la forme immature de la capside virale.

Le FIV étant un enjeu vétérinaire important, il serait intéressant de développer des inhibiteurs spécifiques anti-FIV, afin d'améliorer le traitement des chats et des espèces félines en danger. La connaissance de la structure tridimensionnelle de la CA du FIV permettrait de développer, de façon rationnelle, des inhibiteurs spécifiques contre l'assemblage du FIV.

LE PROJET DOCTORAL

Les rétrovirus sont caractérisés par leur capacité à développer des résistances aux thérapies antirétrovirales, notamment grâce à l'accumulation de mutations dans leur génome. La compréhension des déterminants structuraux sous-jacents à la fonction de ces protéines virales constitue une étape essentielle dans le développement de stratégies antirétrovirales efficaces.

Ce projet doctoral porte sur l'étude des bases structurales des mécanismes moléculaires impliqués dans les fonctions clés des rétrovirus que sont i) la régulation de l'expression des protéines de rétrovirus complexes et ii) l'assemblage des particules virales, à travers l'étude de deux protéines rétrovirales d'intérêt thérapeutique.

La première partie de ce projet porte sur l'étude fonctionnelle et structurale de la protéine transactivatrice de la transcription Tax d'HTLV-2 (Tax-2). En effet, Tax étant un élément crucial pour la réplication et la pathogénicité des HTLV, une meilleure connaissance de sa structure, jusqu'alors inconnue, pourrait conduire à la découverte de nouvelles cibles thérapeutiques contre ces virus. L'apport d'information structurale sur cette protéine pourrait ainsi guider la conception d'antirétroviraux anti-HTLV. Par ailleurs, le recrutement du facteur de transcription CREB par Tax étant crucial pour sa fonction de transactivation de la transcription, l'étude du complexe Tax-2:CREB permettrait de mieux comprendre les mécanismes impliqués dans cette fonction.

La seconde partie du projet doctoral concerne la caractérisation structurale de la protéine de capside du FIV, essentielle aux étapes d'assemblage et de maturation des virions nouvellement formés. Actuellement, peu d'informations sont recensées concernant la CA du FIV, contrairement à celle du VIH. La résolution de sa structure permettrait donc de mieux comprendre les mécanismes moléculaires mis en jeu lors de l'assemblage du FIV et ses spécificités par rapport aux autres rétrovirus. Parallèlement à l'étude structurale de la CA, une collaboration a été initiée avec l'équipe du Dr. G Alvarez (Paysandú, Uruguay) pour l'identification de molécules thérapeutiques ciblant l'assemblage de la CA. L'identification d'un tel composé mènerait à la conception d'un composé antirétroviral bloquant la formation des particules de FIV en inhibant l'assemblage de la capside.

MATERIELS ET METHODES

1. Constructions plasmidiques

1.1. Constructions plasmidiques de la protéine Tax d'HTLV-2 (Tax-2)

1.1.1. Stratégie de clonage

Avant mon arrivée au laboratoire, la protéine Tax-2 entière, dont la structure est encore inconnue, a subi une expérience de protéolyse ménagée par la thermolysine afin de déterminer les frontières de ses domaines globulaires. Les différents fragments issus de cette protéolyse ont été analysés par chromatographie liquide couplée à la spectrométrie de masse en tandem et un fragment protéique, composé des résidus d'acides aminés 34 à 200 de Tax-2, a été identifié comme potentiellement globulaire. Cependant, celui-ci ne possède pas l'intégralité du domaine en doigt de zinc de la protéine nécessaire à sa dimérisation.

Une étude complémentaire *in silico* de la séquence de la protéine Tax-2 par le programme en ligne « Psipred » (http://bioinf.cs.ucl.ac.uk/psipred/) (**Figure 18**), prédisant la composition en structures secondaires des protéines à partir de leur séquence, a alors été réalisée.

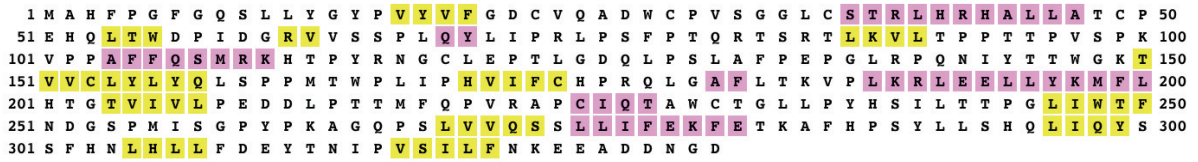

Figure 18 | Prédiction de la composition en structures secondaires de la protéine Tax-2 par le programme en ligne « Psipred ».

En jaune et en rose sont annotés respectivement les résidus d'acides aminés repliés sous forme de brins β et d'hélices α. Les acides aminés non colorés représentent des zones dont la structure secondaire n'est pas prédite.

A partir de ces prédictions, plusieurs fragments de la protéine Tax-2, contenant différents domaines fonctionnels et domaines prédits comme structurés, ont été construits (**Figure 19**).

En plus des fragments de Tax-2 seuls, la protéine a également été clonée sous forme entière (**Figure 19**) pour tenter d'obtenir Tax-2 en complexe avec son partenaire cellulaire CREB. En effet, le recrutement du facteur de transcription CREB par la protéine Tax est essentiel à son activité de transactivation de la transcription.

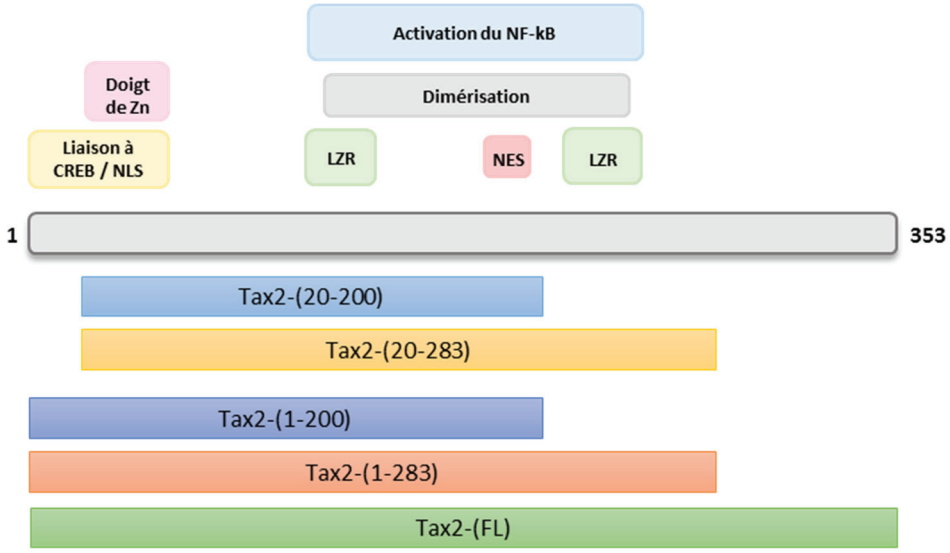

Figure 19 | Constructions de Tax-2 et domaines fonctionnels associés.

En plus de la protéine Tax-2 entière (Tax2-(FL)), 4 fragments ont été construits : Tax2-(20-200), Tax2-(20-283), Tax2-(1-200) et Tax2-(1-283). Ils seront clonés dans différents vecteurs d'expression en fusion avec des étiquettes d'affinité/solubilité aux extrémités amino- et/ou carboxy-terminales.

Enfin, une optimisation de codons a été réalisée par la société GenScript afin d'améliorer la production et la solubilité de la protéine Tax-2 entière en système bactérien. En effet, il existe plusieurs combinaisons de trois nucléotides qui codent le même résidu d'acide-aminé ; et chaque organisme utilise préférentiellement certains codons afin de coder les acides aminés correspondants. La séquence de la protéine Tax-2, issue d'un virus, a donc été optimisée pour l'expression en système bactérien en remplaçant les codons rares de la séquence par des codons utilisés plus fréquemment chez *E. coli*. Cette optimisation a permis la création d'une seconde construction Tax-2 entière nommée Tax2-(FL)-opti.

Le clonage des gènes correspondants, amplifiés par PCR (**Matériels et méthodes 1.1.2**), a été réalisé dans le vecteur intermédiaire pCR®2.1 (Thermofisher) (**Figure 20**), permettant d'amplifier et de caractériser les gènes d'intérêt, puis dans le vecteur d'expression choisi.

Pour les protéines seules, les vecteurs d'expression pET-30a (Novagen) et pET-M30 (**Figure 20**) ont été choisis afin de générer des protéines de fusion avec une étiquette d'affinité (poly-histidine (His_6)) et/ou une étiquette de solubilité (GST ou SUMO).

Le vecteur d'expression pETDuet-1 (**Figure 20**), lui, a été choisi pour la co-expression des protéines Tax-2 et CREB sous forme de complexe puisqu'il permet l'expression simultanée de deux protéines sous le contrôle de deux promoteurs identiques, inductibles simultanément.

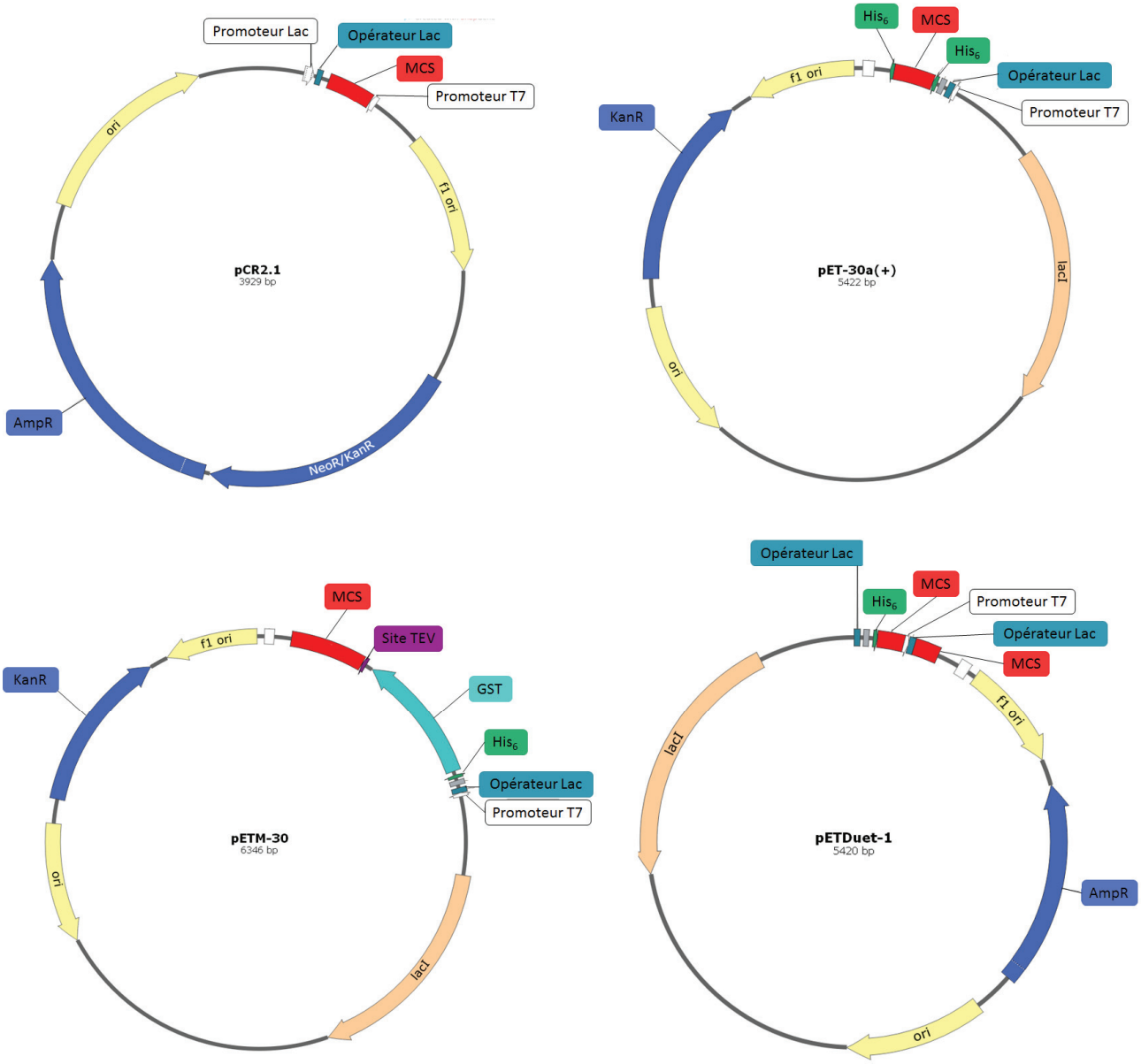

Figure 20 | Carte des plasmides pCR®2.1, pET-30a, pETM-30 et pETDuet-1 utilisés pour les différentes constructions de Tax-2. Les gènes d'intérêt ont été clonés dans le site multiple de clonage (MCS, *Multiple Cloning Site*, en rouge) des vecteurs pET-30a, pETM-30 et pETDuet-1. Figure réalisée avec le logiciel SnapGene.

Au total, 10 constructions distinctes de Tax-2 (**Tableau 5**) ont été générées en fusion avec une ou plusieurs étiquette(s) positionnée(s) aux extrémités amino- et/ou carboxy-terminales de la protéine.

Constructions		Étiquettes et position		Amorces utilisées pour l'amplification		Vecteurs d'expression	Enzymes utilisées pour le clonage
				Sens	***Anti-sens***		
Expression Tax-2 seules							
1	Tax2-(20-200)-6H-C	*His_6*	C	F1	R2	pET-30a	NdeI/HindIII
2	Tax2-(20-200)-6H-N		N	F3 F4	R2		
3	Tax2-(20-283)-6H-C	*His_6*	N	F1	R1	pET-30a	NdeI/HindIII
4	Tax2-(20-283)-6H-N		C	F3 F4	R3		
5	Tax2-(20-283)-6H-GST-N	*His_6-GST*	N	F5	R3	pETM-30	NcoI/HindIII
6	Tax2-(1-200)-6H-N	*His_6*	N	F2 F4	R2	pET-30a	NdeI/HindIII
7	Tax2-(1-283)-6H-N	*His_6*	N	F2 F4	R3	pET-30a	NdeI/HindIII
8	Tax2-(1-283)-6H-N-Sumo-C	*His_6* *Sumo*	N C	F2 F6 F4	R4 R5		
9	Tax2-(FL)-opti-6H-C	*His_6*	C			pET-30a	NdeI/HindIII
Co-expression Tax-2/CREB							
10	Tax2-(FL)-CREB	*His_6 (Tax)* *Flag (CREB)*	N N	F7 F8 F9	R6 R7	pETDuet-1	NcoI/HindIII NdeI/SalI

Tableau 5 | Constructions plasmidiques pour la protéine Tax-2 : étiquettes, amorces, vecteurs et enzymes de restriction utilisés. N et C désignent les extrémités amino- et carboxy-terminales de la protéine respectivement, et His_6 l'étiquette poly-histidine. Les amorces utilisées et leurs séquences sont reportées dans le tableau 6. Les enzymes de restriction ont permis le clonage des différents fragments d'ADN dans les sites multiples de clonage des vecteurs présentés en figure 19.

1.1.2. Amplification en chaîne par polymérase

La création et l'amplification des différents gènes *tax-2* et *creb* ont été réalisées par PCR *(Polymerase Chain Reaction)*. L'amplification des séquences codantes de Tax-2 et CREB a été réalisée à partir des plasmides recombinants pET-30a-Tax2 et pCF-CREB contenant l'ADN codant la protéine Tax-2 entière du sérotype HTLV-2B et l'ADN codant la protéine CREB, respectivement (plasmides fournis par le Pr. R. Mahieux, CIRI - Centre International de Recherche en Infectiologie).

Les réactions de PCR ont été réalisées avec le kit Phusion® High-Fidelity PCR (New England BioLabs) dans un mélange réactionnel composé de Tampon High Fidelity 1X final, 200 µM de désoxyribonucléotides triphosphates (dNTPs), 0,5 µM d'amorces et 0,02 $U.\mu L^{-1}$ de Phusion® ADN Polymerase. Les amplifications ont été réalisées dans un thermocycleur GeneAmp® PCR System 9700 (Applied Biosystems) selon la programmation suivante : dénaturation à 95°C pendant 5 minutes suivie de 30 cycles d'amplification (dénaturation à 95°C pendant 30 secondes, hybridation à 50°C pendant 30 secondes et élongation à 72°C pendant 1 minute) et d'une élongation finale de 7 minutes à 72°C.

Les différentes amorces utilisées (**Tableau 5** et **Tableau 6**) permettent l'addition de sites de restriction aux extrémités 5' et 3' des gènes en vue de leur clonage dans le vecteur d'expression choisi (**Tableau 5**) ainsi que, pour certaines constructions, l'ajout d'une étiquette d'affinité et/ou de solubilité à l'extrémité amino- ou carboxy-terminale des protéines.

En vue du clonage des gènes d'intérêt dans le vecteur pCR®2.1, les PCR ont été suivies d'une étape permettant d'ajouter une adénine à l'extrémité 3' des fragments d'ADN grâce à l'activité terminale transférase aspécifique de la *Taq* polymérase. Cette extension en 3' a été réalisée en ajoutant 0,5 µL de *Taq* polymérase (5 $U.\mu L^{-1}$) (New England BioLabs) au produit de PCR puis en incubant le mélange à 72°C pendant 15 minutes.

Les produits issus de la réaction de PCR ont ensuite été soumis à une migration dans un gel d'agarose 1% (m/v)-TAE (Tris-Acétate/EDTA (*Ethylene DiamineTetraacetic Acid*) (40 mM Tris-acétate pH 7,8, 1 mM EDTA) en présence de bromure d'éthidium, un agent intercalant de l'ADN. Une solution de marqueurs de taille « 100bp DNA Ladder » (New England BioLabs) a été utilisée. Après révélation aux ultraviolets, les bandes correspondant aux produits d'amplification ont été extraites et purifiées en utilisant le kit NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel) selon le protocole du fournisseur.

La concentration de l'ADN purifié est ensuite estimée grâce à un spectrophotomètre NanoDrop (ND-100, Thermo Scientific) par mesure de l'absorbance à 260 nm.

Noms des amorces		Séquences
Sens	F1	AACAT**ATG**GGCGATTGTGTACAAGCCGATTGGTG
	F2	CTCGTTCCTCGTGTTTCTGCCCATTTCCCAGGATTTG
	F3	CTCGTTCCTCGTGTTTCTGGCGATTGTGTACAAGCCG
	F4	ACAT**ATG**CACCATCACCATCACCATCTCGTTCCTCGTGTTTC
	F5	CCCC**ATG**GGCGATTGTGTACAAGCCGATTGG
	F6	GAAAAATTCCAAACCAAAGGGTCTGACCAGGAGGCTAAAC
	F7	AAATCTTTATTTTCAAGGTGCCCATTTCCCAGGATTTGG
	F8	AGTCC**ATG**GGCCACCATCACCATCACCATGAAAATCTTTATTTTCA
	F9	ATACAT**ATG**GACTACAAGGACGACGATGACAAG
Anti-sens	R1	AAAAGCTTGGTTTGGAATTTCTCGAAGATTAATAGAGAG
	R2	TAAGCTTTCATCATAAGAACATTTTGTATAG
	R3	TAAGCTTTCAGGTTTGGAATTTCTCGAAGATTAATAG
	R4	GTTTAGCCTCCTGGTCAGACCCTTTGGTTTGGAATTTTTC
	R5	AAAAGCTTCCCCCCGTTTGTTCCTGATAAAC
	R6	GCAAAGCTTCTAGTCGCCATTGTCATCCGCCTC
	R7	TCTGTCGACTTAATCTGATTTGTGGCAGTAAAG

Tableau 6 | Constructions Tax-2 : Séquences des amorces utilisées pour l'amplification des fragments d'ADN. Les sites de restriction (NdeI : CATATG, HindIII : AAGCTT, NcoI : CCATGG et SalI : GTCGAC) sont soulignés. Les codons d'initiation de la traduction (ATG) sont en gras et les codons stop en rouge. Les étiquettes His_6 et Flag-tag sont indiquées en vert et en jaune respectivement, le site de coupure au facteur X en bleu et le site de coupure à la TEV en mauve.

1.1.3. Clonage dans le vecteur pCR®2.1 selon la méthode du TA Cloning®

1.1.3.1. Ligation dans le vecteur pCR®2.1

La méthode du TA Cloning® permet le clonage direct d'un produit de PCR dans le vecteur de clonage pCR®2.1, sans avoir recours aux enzymes de restriction, afin d'amplifier un gène d'intérêt dans des bactéries de clonage. Elle repose sur la capacité des bases complémentaires adénine (A) et thymine (T) à s'hybrider. Le vecteur linéaire pCR®2.1 fourni dans le kit TA Cloning® (Thermofisher) possède à ses extrémités 3' une thymidine (T) pouvant s'apparier

avec l'adénine de l'extrémité 3' de l'insert ajoutée par la *Taq* polymérase. L'insertion des ADN d'intérêt dans le vecteur se fait grâce à l'action de l'ADN ligase du bactériophage T4 qui va former une liaison phosphodiester entre les extrémités 5'-phosphate et 3'-OH des ADN.

Les ligations ont été effectuées selon le protocole kit de TA cloning avec une incubation à 14°C pendant la nuit puis une inactivation de la ligase à 65°C pendant 15 minutes.

1.1.3.2. Transformation bactérienne et extraction d'ADN plasmidique

Les plasmides recombinants obtenus après ligation ont ensuite été amplifiés par transformation de 50 µL bactéries compétentes *Escherichia coli* (*E. coli*) DH5α (Lucigen) avec la moitié du produit de ligation. Pour cela, le mélange ADN/bactéries est incubé dans la glace pendant 30 minutes puis soumis à un choc thermique à 42°C pendant 45 secondes qui va déstabiliser les membranes bactériennes et les rendre momentanément perméables à l'ADN plasmidique. Les bactéries sont supplémentées par 300 µL de milieu optimal (SOC, *Super Optimal broth with Catabolic repression*) et incubées une heure à 37°C sous agitation. Le produit de transformation obtenu est mis en culture sur un milieu LB-agar (Sigma), supplémenté par l'antibiotique approprié, puis est incubé pendant la nuit à 37°C.

Dans le cas des plasmides recombinants pCR®2.1, les boîtes de Petri LB-agar, sur lesquelles ont été étalées les bactéries transformées par le produit de ligation, ont été additionnées de kanamycine (50 $\mu g.mL^{-1}$ final) et de X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, Euromedex) à 100 $\mu g.mL^{-1}$ final. Le vecteur pCR®2.1 utilisé, contenant un gène de résistance à la kanamycine, permet la sélection des bactéries ayant inséré un plasmide par application de cet agent sélectif. De plus, un second système de sélection repose sur la présence dans le vecteur du gène *lacZ,* codant la β-galactosidase, une enzyme capable d'hydrolyser le substrat chromogène X-Gal. L'insertion de l'ADN d'intérêt dans le vecteur se faisant à l'intérieur du gène *lacZ*, un décalage du cadre de lecture de la β-galactosidase apparaît. La souche *E. coli* utilisée induit naturellement l'expression du gène *lacZ* codant la β-galactosidase contenu dans le vecteur. Par conséquent, les bactéries contenant un plasmide recombinant ayant inséré l'ADN d'intérêt n'exprimeront pas une β-galactosidase fonctionnelle et seront blanches contrairement aux bactéries ayant incorporé un plasmide vide qui, exprimant l'enzyme, seront bleues en raison de l'hydrolyse du X-Gal.

1.1.3.3. Sélection des transformants positifs

Des colonies positives isolées sont ensuite repiquées et mises en culture dans 3 mL de milieu nutritif LB (Sigma) additionné de l'antibiotique approprié à 37°C pendant la nuit. À partir des culots bactériens obtenus après centrifugation des cultures bactériennes, l'ADN plasmidique est extrait grâce au kit NucleoSpin Plasmid (Macherey-Nagel), selon le protocole du fournisseur.

Une vérification de la présence et de l'intégrité des fragments d'ADN d'intérêt dans le plasmide est effectuée. Pour cela, les ADN plasmidiques extraits sont digérés par des enzymes de restriction et déposés sur gel d'agarose 1% (m/v)-TAE. Les plasmides présentant un insert à la taille attendue sont ensuite séquencés selon la méthode de Sanger (séquençage commercial, GATC Biotech) en utilisant des amorces spécifiques au plasmide fournie par la société de séquençage ou par nos soins.

Pour le clonage en pCR®2.1, l'enzyme de restriction *Eco*RI et les amorces M13-RP et M13-FP ont été utilisées pour la sélection de clones positifs.

Enfin, la concentration en ADN des clones positifs ainsi que sa pureté sont estimées grâce au spectrophotomètre NanoDrop par mesure de l'absorbance à 260 nm.

1.1.4. Sous-clonage des séquences codantes dans le vecteur d'expression

Les gènes d'intérêt sont clonés dans le vecteur d'expression choisi (pET-30a, pETM-30 ou pETDuet-1) afin d'exprimer les protéines d'intérêt en système procaryote. Ce sous-clonage s'effectue à l'aide des sites de restriction introduits dans les différentes constructions et également présents dans le site multiple de clonage (MCS, *Multiple Cloning Site*) du vecteur d'expression choisi (**Tableau 5**).

Les ADN d'intérêt présents dans le vecteur pCR®2.1 ainsi que le vecteur d'expression subissent une digestion enzymatique par les enzymes de restriction correspondants à chaque construction. Pour cela, 2,5 µg de plasmides sont incubés avec 0,5 µL de chaque enzyme de restriction pendant 2h à 37°C dans le tampon préconisé par le fournisseur. Le vecteur d'expression est ensuite soumis à l'action de la phosphatase alcaline intestinale de veau (CIAP, Invitrogen) permettant l'élimination des groupements phosphates à l'extrémité 5' du vecteur et évitant ainsi qu'il ne se referme sur lui-même pendant la ligation.

En pratique, le vecteur digéré est traité par 0,01 $U.\mu L^{-1}$ de CIAP pendant 15 minutes à 37°C puis l'enzyme est inactivée 10 minutes à 70°C. Les ADN digérés sont ensuite purifiés après migration dans un gel d'agarose 1% (m/v)-TAE en présence de bromure d'éthidium selon le protocole décrit précédemment (**Matériels et méthodes 1.1.2**).

Enfin, la ligation est effectuée entre les fragments d'ADN et le vecteur d'expression purifiés grâce à l'action de 1 U de T4 DNA ligase (Lucigen) à 15°C pendant la nuit, à raison d'un rapport molaire 5:1 (insert:vecteur).

Les plasmides recombinants obtenus sont utilisés pour transformer des bactéries compétentes *E. coli* DH5α selon le protocole décrit précédemment (**Matériels et méthodes 1.1.3.2**) et les ADN plasmidiques sont ensuite été extraits grâce au kit NucleoSpin Plasmid (Macherey-Nagel), selon le protocole du fournisseur.

Après digestion par les enzymes de restriction spécifiques de chaque construction (**Tableau 5**), les ADN sont déposés sur gel d'agarose 1% (m/v)-TAE et les plasmides présentant un insert à la taille attendue sont séquencés selon la méthode de Sanger (GATC Biotech). Enfin, la concentration en ADN de l'échantillon ainsi que sa pureté sont estimées grâce au spectrophotomètre NanoDrop par mesure de l'absorbance à 260 nm.

1.2. Construction plasmidique de la protéine de capside du FIV

La construction plasmidique de la CA du FIV a été réalisée avant mon arrivée au laboratoire. Le gènes de la CA entière a été amplifié par PCR à partir du plasmide p34TF10 contenant le génome du virus FIV de la souche Petaluma (Serriere *et al.*, 2013).

Une troncature de 11 résidus d'acides aminés, accompagnée d'une mutation de la Pro1 en Thr1, ont été réalisées afin d'obtenir la construction de CA utilisée pour ce projet : CAΔC-PT, clonée dans le vecteur d'expression pRSET-B (Invitrogen) (**Figure 21**) en fusion avec une étiquette poly-histidine à son extrémité amino-terminale.

Ces modifications ont été introduites afin de faciliter la croissance des cristaux de CA. En effet, les 11 résidus de l'extrémité carboxy-terminale de la protéine sont prédits comme étant flexibles, ce qui peut défavoriser le processus de cristallisation. Avant mon arrivée, environ 1000 conditions de cristallisation ont été testées pour la CA entière native, sans succès.

Le premier résidu d'acide aminé de la CA (Pro1), quant à lui, a été muté pour permettre le clivage de l'étiquette poly-histidine par la protéase Facteur X. Cette protéase a été choisie car elle permet de retirer tous les résidus d'acides aminés du site de clivage et, par conséquent, de récupérer la protéine d'intérêt avec une extrémité amino-terminale intacte. Cependant, cette protéase n'est pas active lorsque le premier résidu d'acide aminé après le site de clivage est une proline, ce qui est le cas pour la protéine CA native du FIV.

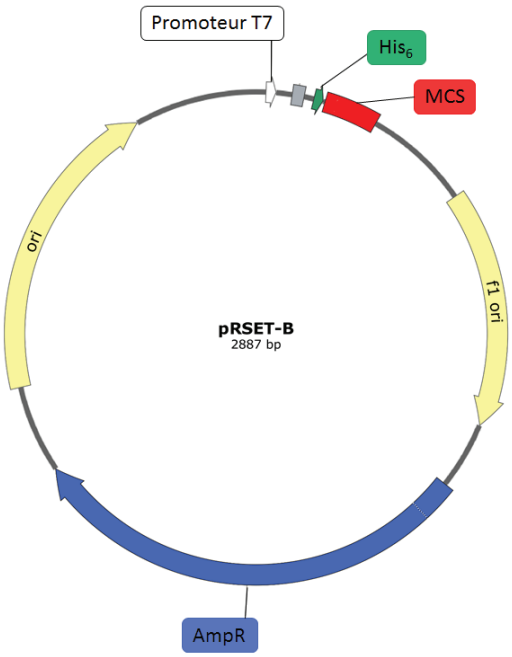

Figure 21 | Carte du plasmide pRSET-B utilisé pour la construction CAΔC-PT. Le gène d'intérêt a été cloné dans le site multiple de clonage (MCS, en rouge) du vecteur. Figure réalisée avec le logiciel SnapGene.

2. Surproduction et purification des protéines d'intérêt

2.1. Généralités

2.1.1. Surproduction en système procaryote

L'expression des gènes d'intérêt (*tax-2*, *creb* et *caΔc-pt*), en système procaryote, est sous le contrôle du promoteur du phage T7, présent dans les vecteurs d'expression utilisés et spécifiquement reconnu par l'ARN polymérase T7.

Les souches bactériennes *E. coli* BL21 (DE3) pLysS (Lucigen), C41 (DE3) / C43 (DE3) (Novagen) et RosettaGami2 (DE3) (Novagen) ont été utilisées pour la surproduction des protéines d'intérêt au cours de ce projet doctoral. Ces bactéries possèdent un système de répression de l'expression génique, basé sur la répression du promoteur *lacUV5* par le répresseur LacI (Dubendorff et Studier, 1991). L'ajout d'IsoPropyl-β-D-1-ThioGalactopyranoside (IPTG, Euromedex) dans le milieu de culture va lever cette répression et permettre la surexpression du gène cloné sous le contrôle du promoteur reconnu par l'ARN polymérase T7. Les bactéries *E. coli* BL21 (DE3) pLysS, elles, possèdent un second système de répression, basé sur l'action du lysozyme du phage T7 contenu dans le plasmide pLysS, qui permet de réprimer l'expression basale du gène d'intérêt (Studier, 1991) afin d'éviter toute fuite du promoteur T7.

2.1.2. Purification par chromatographie d'affinité pour le nickel

Après surproduction des protéines recombinantes, celles portant une étiquette poly-histidine (His_6) à leur extrémité amino- ou carboxy-termimale (**Tableau 5**) sont purifiées par chromatographie d'affinité pour le nickel. Ce type de purification utilise la capacité des cycles imidazoles des histidines à chélater certains ions métalliques (**Figure 22**).

En pratique, deux types de résines ont été utilisées pour la purification des protéines d'intérêt au cours de ce projet : une résine Ni-NTA (Qiagen) pour les protéines Tax-2 et une résine Ni-TED (Macherey Nagel) pour la protéine de capside CA. Ces résines sont constituées d'atomes de nickel immobilisés sur un support solide (billes d'agarose ou de silice, respectivement) par l'intermédiaire d'un bras espaceur, l'acide nitrilotriacétique (*NitriloTriacetic Acid*, NTA) (**Figure 22A**) ou le tris(carboxyméthyl)ethylène diamine (TED) (**Figure 22B**). La différence entre ces deux résines réside dans leur capacité de fixation, c'est-à-dire le nombre de cycles histidine pouvant être fixés à une même bille, à savoir une pour la Ni-TED et deux pour la Ni-

NTA (**Figure 22**). L'élution des protéines d'intérêt se fait ensuite à l'aide d'un compétiteur pour le nickel comme l'imidazole libre.

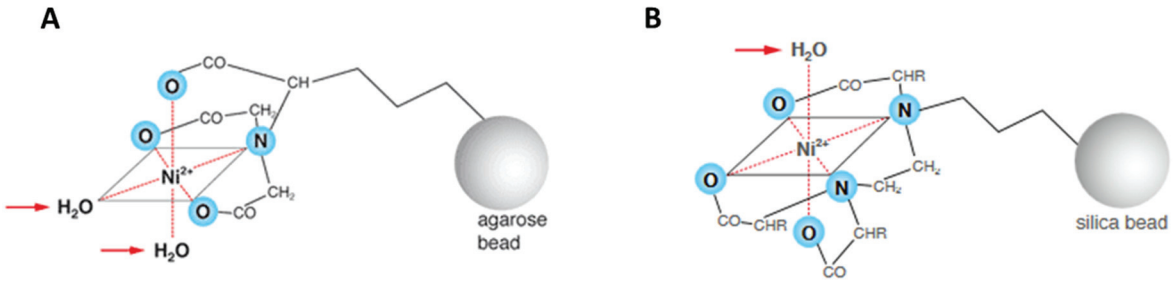

Figure 22 | Structures des résines A) Ni-NTA et B) Ni-TED utilisées lors de l'étape de purification par chromatographie d'affinité pour le nickel (Macherey-Nagel) des protéines Tax-2 et CA, respectivement.

2.2. La protéine Tax-2 du virus HTLV-2

2.2.1. Tests de surproduction de la protéine Tax-2 en système bactérien

Afin de déterminer la meilleure condition d'expression des gènes *tax-2* sous forme soluble, différents tests d'expression en petit volume ont été effectués.

Pour chaque test, 50 µL de bactéries compétentes commerciales *E. coli* sont transformées avec 50 à 100 ng de plasmide recombinant, puis étalées sur une boîte LB-agar contenant l'antibiotique approprié et du chloramphénicol (35 µg.mL^{-1}). Les bactéries ayant poussé étant celles transformées par le plasmide recombinant, elles sont repiquées afin de réaliser une préculture de 3 mL en milieu LB liquide additionné d'antibiotiques, incubée à 37°C pendant la nuit. Chaque nouvelle culture est réalisée à partir de bactéries fraîchement transformées afin de garantir une expression optimale du gène d'intérêt.

Une fraction de préculture est ensuite utilisée pour ensemencer au centième 40 mL des différents milieux de culture testés (additionnés de l'antibiotique approprié) : LB (Sigma), Turbo et Hyper Broth (AthenaES®) (TB et HB, respectivement), Terrific Broth (TerB) (Sigma), 2YT (Sigma). L'expression des gènes est ensuite induite par ajout d'IPTG dans le milieu de culture, dans le cas des milieux LB, TB, HB, TerB et 2YT.

Un dernier milieu dit d'auto-induction (AI) a également été testé, ce dernier permettant la production des protéines d'intérêt sans ajout d'IPTG. En effet, le glucose contenu dans le milieu AI (**Tableau 7**) va prévenir l'induction de l'expression des protéines par le lactose, le glucose étant métabolisé préférentiellement par les bactéries (Studier, 2005).

Après épuisement du glucose, la production des protéines se fait grâce à l'action du lactose, également présent dans le milieu, sur le promoteur T7 : c'est ce que l'on appelle l'auto-induction. Ainsi, ce milieu AI permet de produire des protéines recombinantes sans avoir recours au suivi de la croissance bactérienne pour l'ajout du composé inducteur au moment précis de la phase exponentielle de croissance.

Composition	LB	Na_2HPO_4	KH_2PO_4	NH_4Cl	$MgSO_4$	Glycérol	Glucose	Lactose
Concentration	-	25 mM	25 mM	50 mM	2 mM	0,5%	0,05%	0,2%
Quantité pour 1L de milieu	20 g	3,55 g	3,4 g	2,67 g	0,24 g	5 mL	0,5 g	2 g

Tableau 7 | Composition du milieu d'auto-induction (AI). La composition du milieu AI a été adaptée de (Studier, 2005) pour la production de protéines recombinantes en système procaryote sans ajout d'IPTG.

Enfin, après incubation, les cultures sont centrifugées à 4 500 × g à température ambiante pendant 5 minutes afin de sédimenter les bactéries. Le culot bactérien obtenu est ensuite conservé à -20°C en attendant l'étape de lyse.

Afin d'optimiser la production des protéines Tax-2 d'intérêt, différentes conditions d'expression et d'induction ont été testées en faisant varier les bactéries d'expression utilisées (*E. coli* (DE3) pLysS, C41 (DE3) / C43 (DE3) ou RosettaGami2 (DE3)), la concentration finale en IPTG dans le milieu (0,05 à 2 mM), les températures d'incubation (15°C à 37°C) ou encore les temps d'induction (3h à 48h) (**Tableau 8**).

D'autres protocoles d'expression adaptés à la production de protéines difficiles (**Tableau 8**) ont également été testés dont :

- L'ajout d'additif dans le milieu de culture : la protéine Tax contenant un domaine en doigt de zinc à son extrémité amino-terminale, des tests en présence de Zn^{2+} dans le milieu de culture, ont été réalisés (Matsuda *et al.*, 2006).
- Le contrôle de la traduction par ajout de 200 $\mu g.mL^{-1}$ de tétracycline après 1h d'induction : la tétracycline bloquant la traduction, aucun des ARNm produits ne sera traduit. Ainsi, les protéines chaperonnes ne seraient sollicitées que pour le repliement d'un pool limité de protéines d'intérêt, ce qui pourrait favoriser leur repliement dans une conformation adéquate (Carrio et Villaverde, 2001).

– Le contrôle de la transcription par ajout de 100 µg.mL^{-1} de rifampicine dans le milieu après 1h d'induction : la rifampicine bloquant la transcription et donc la croissance bactérienne, seule la traduction des ARNm déjà produits est assurée par la machinerie cellulaire pendant l'incubation (Fiorini *et al.*, 2008). Si ces ARNm codent une protéine nocive à la croissance de la bactérie, cette protéine ne pourra pas jouer son rôle délétère sur une bactérie transcriptionnellement inactive.

Paramètres variables		**Conditions testées**	**Commentaires**
Expression	**Milieux**	LB, TB, HB, TerB, 2YT, AI	Certains milieux étant tamponés ou supplémentés en sels minéraux, acides aminés ou oligo-éléments, ils peuvent permettre une meilleure expression des protéines d'intérêt.
	Bactéries *E. coli*	BL21 (DE3) pLysS	Bactéries les plus utilisées, déficientes en protéases Lon et OmpT.
		C41 (DE3) C43 (DE3)	Bactéries utilisées pour l'expression de protéines toxiques ou membranaires (Dumon-Seignovert *et al.*, 2004).
		RosettaGami2 (DE3)	Bactéries contenant le plasmide pRAR pour l'expression des codons rares et permettant la formation de ponts disulfures.
	Ajout de Zn dans le milieu de culture		Pour favoriser le repliement des protéines.
Induction	**[IPTG] finale**	0,05 mM, 0,5 mM, 1 mM ou 2 mM	Une induction avec une très faible concentration en IPTG permet aux bactéries d'avoir un meilleur contrôle sur leur métabolisme en réponse à l'expression d'une protéine toxique.
	Ajout de tétracycline 1h après l'IPTG et incubation pendant		Blocage de la traduction pour favoriser le repliement protéique.
	Ajout de rifampicine 1h après l'IPTG et incubation pendant 3h		Blocage de la transcription, seule la traduction a lieu pendant l'incubation.
	Températures	15°C, 20°C, 25°C, 30°C ou 37°C	La diminution de la température d'induction permet d'améliorer la production de protéines solubles correctement repliées et de minimiser la formation de corps d'inclusion, en ralentissant le taux de transcription et de traduction.
	Temps	3h, 5h, pendant la nuit ou 48h	Selon la température d'induction, le temps d'incubation est allongé afin de produire le plus de protéines d'intérêt possible.

Tableau 8 | Conditions d'expression et d'induction testées pour la production de la protéine Tax-2 en système procaryote.

2.2.2. Lyse bactérienne et extraction de la protéine Tax-2

Afin d'extraire les protéines produites, les culots bactériens, préalablement décongelés, sont soumis à une étape de lyse combinant l'action de détergents et la sonication par ultrasons.

Différentes conditions de lyse ont été testées afin d'améliorer la solubilité des protéines produites en faisant varier la composition du tampon de lyse (**Tableau 9**).

En pratique, les culots bactériens sont resuspendus et incubés pendant 30 minutes dans 750 µL de tampon de lyse en présence d'un mélange d'anti-protéases sans EDTA (Complete EDTA-free, Roche) et de DNase I (Sigma). Le mélange d'anti-protéases sans EDTA a été choisi afin de ne pas interférer avec la purification ultérieure des protéines d'intérêt par chromatographie d'affinité, basée sur l'utilisation de cations divalents (Ni^{2+}) qui seraient chélatés par l'EDTA.

Tampons de lyse	**Composition**		
	Composés	***Concentration***	***pH***
L1	Tris-HCl KCl CHAPS DTT	50 mM 100 mM 0,5% (m/v) 1 mM	7,5
L2	PBS NaCl EDTA β-mercaptoéthanol Glycérol $MgCl_2$	1X 150 mM 0,1 mM 2 mM 10 % 1 mM	7,5
L3	PBS (Euromedex) NaCl PMSF β-mercaptoéthanol Imidazole	1X 300 mM 0,25 mM 0,5 mM 10 mM	8,0
L4	B-PER (*Bacterial Protein Reagent*, Pierce)	Pierce	

Tableau 9 | Composition des tampons testés pour la lyse des bactéries ayant exprimées Tax-2.

PBS : Phosphate Buffer Saline, CHAPS : 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate, DTT : Dithiothréitol, PMSF : fluorure de phénylméthylsulfonyle.

Après incubation 30 minutes à 4°C, les cellules sont lysées par sonication (sonicateur Branson Sonifier 250) avec 3 cycles de 30 secondes de pulsations à 40% de puissance, séparés par des temps de refroidissement dans la glace. Le lysat cellulaire obtenu est ensuite centrifugé à

7 500 × g pendant 30 minutes à 4°C : le surnageant (fraction soluble) est conservé et le culot (fraction insoluble) est repris dans un volume de 750 µL de tampon de lyse.

Des aliquotes correspondant aux fractions totale (après sonication), soluble et insoluble (respectivement surnageant et culot après centrifugation du lysat) sont conservées en vue d'une analyse par électrophorèse sur gel discontinu de polyacrylamide en conditions dénaturantes (« SDS - PolyAcrylamide Gel Electrophoresis » ou SDS-PAGE).

2.2.3. Solubilisation des protéines Tax-2

L'expression de gènes recombinants hétérologues en système procaryote peut conduire, dans certains cas, à la production de protéines recombinantes agrégées sous forme de corps d'inclusion. Afin de solubiliser ces protéines *in vitro* (Vallejo et Rinas, 2004), deux méthodes de solubilisation utilisant deux agents chaotropiques différents ont été testées.

2.2.3.1. Dénaturation par le chlorure de guanidinium

Afin d'extraire les protéines des corps d'inclusion, un protocole utilisant du chlorure de guanidinium à forte concentration, déjà mis en place au laboratoire, a été utilisé.

En pratique, les culots bactériens sont lysés dans un tampon 20 mM Tris-HCl pH 7,5 contenant 20% de saccharose. Après un choc thermique à 4°C, l'échantillon est centrifugé et le culot est repris dans un tampon 1X PBS, 0,5 mM PMSF et centrifugé de nouveau. Les corps d'inclusion ainsi culottés sont repris dans un tampon 1X PBS, 1% Triton X-100 et centrifugés. Enfin, le culot est repris dans une solution contenant 6 M de chlorure de guanidinium, 25 mM DTT et 50 mM Tris-HCl à pH 8 puis soniqué deux fois 30 secondes à 40% de puissance et incubé 1h à 4°C. Après une dernière centrifugation, les fractions culot et surnageant sont récupérées pour une analyse par SDS-PAGE de la solubilisation des protéines d'intérêt.

2.2.3.2. Dénaturation par l'urée

Un second protocole de solubilisation des protéines par dénaturation par l'urée selon la méthode du « single freeze-thawing cycle » (Qi *et al.*, 2015) a également été testé.

Pour cela, le culot bactérien est lysé dans un tampon 20 mM Tris-HCl pH 8,5, 100 mM NaCl, 1% Triton X-100, 1 µg.mL^{-1} DNase I et un mélange d'anti-protéases. Il est ensuite soniqué quatre fois 30 secondes à 40% de puissance et centrifugé 15 minutes à 12 000 × g à 4°C.

Les corps d'inclusion sont ensuite repris dans un tampon 50 mM Tris-HCl pH 8,5, 300 mM NaCl et centrifugés de nouveau. Le culot est repris dans un tampon 20 mM Tris-HCl, 20 mM phosphate de potassium à différents pH (8, 9 et 10) contenant entre 2 M et 6 M d'urée. L'échantillon est ensuite congelé pendant la nuit à -20°C. Après décongélation, le lendemain, l'échantillon est centrifugé et les fractions culot et surnageant sont analysés par SDS-PAGE.

2.2.4. Purification des protéines Tax-2

2.2.4.1. Précipitation au sulfate d'ammonium

La méthode de précipitation au sulfate d'ammonium permet de séparer les protéines selon leur degré de solubilité en présence d'une plus ou moins forte concentration en sels, ici le sulfate d'ammonium (Duong-Ly et Gabelli, 2014). Elle constitue une première étape de purification intéressante pour s'affranchir de nombreux contaminants avant l'étape de purification par chromatographie d'affinité.

En pratique, les culots bactériens sont repris dans le tampon de lyse (**Tableau 9**) comme décrit au chapitre **Matériels et méthodes 2.2.2**. Après centrifugation, le surnageant est séparé en plusieurs fractions dans lesquelles sont ajoutées des concentrations croissantes de sulfate d'ammonium (de 20 à 100% (m/v)). Ces différentes fractions sont incubées à 4°C sous agitation pendant 30 minutes puis centrifugées pendant 10 minutes à 10 000 × g. Les fractions solubles et insolubles obtenues sont déposées sur SDS-PAGE afin de déterminer les pourcentages de $(NH_4)_2SO_4$ pour lesquels la protéine d'intérêt est soluble et insoluble.

Lorsque ces pourcentages sont déterminés, deux étapes de précipitation au sulfate d'ammonium permettent d'obtenir la protéine d'intérêt en présence d'un minimum de contaminants :

- La première étape consiste en la précipitation des contaminants qui précipitent à un pourcentage de sels inférieur à celui de la protéine d'intérêt ;
- La seconde étape consiste en la précipitation de la protéine d'intérêt par les sels, laissant ainsi les contaminants non précipités dans la fraction soluble.

Après précipitation, la protéine d'intérêt est reprise dans le tampon de lyse en vue de sa purification ultérieure par chromatographie d'affinité.

2.2.4.2. Purification par chromatographie d'affinité pour le nickel

Les protocoles de purification des protéines Tax-2 par chromatographie d'affinité pour le nickel (**Matériels et méthodes 2.1.2**) sont adaptés de deux protocoles fournis par les Dr. S. Chevalier (de l'équipe de R. Mahieux - CIRI) et F. Fiorini (de notre équipe).

Le surnageant protéique obtenu après lyse des culots bactériens est mis en contact avec la résine Ni-NTA pendant 1 heure à 4°C sur roue, à raison de 500 µL de billes pour 1 litre de culture.

L'ensemble est ensuite centrifugé pendant 5 minutes à 1 000 × g afin de sédimenter les billes de résine. Une aliquote du surnageant correspondant à la fraction non retenue est conservée. Les billes, quant à elles, sont déposées sur une colonne par gravité et lavées par le passage de 2 ou 4 fois 5 mL de tampon de lavage W (**Tableau 10**), dont la composition dépend du tampon de lyse utilisé (**Matériels et méthodes 2.2.2**), afin d'éliminer les contaminants protéiques n'ayant pas été adsorbés sur la résine ou y ayant été adsorbés de façon aspécifique. Enfin, quatre à cinq élutions successives de 750 µL sont réalisées avec le tampon d'élution E correspondant (**Tableau 10**). Des fractions des différents lavages et élutions sont conservées afin de les analyser par SDS-PAGE.

Lorsque des conditions de culture et de purification permettant d'obtenir la protéine Tax-2 sous forme soluble et pure seront identifiées, celles-ci sont reproduites à plus grande échelle (1 à 2 L de milieu de culture).

Tampons de lyse	**Tampons de lavage**			**Tampons d'élution**		
	Nom	***Composition***	***pH***	***Nom***	***Composition***	***pH***
L1	**W1**	Tampon L1 + imidazole 50 mM	7,5	**E1**	Tampon L1 + imidazole 300 mM	7,5
L2	**W2**	PBS 1,5X NaCl 225 mM EDTA 0,1 mM β-mercaptoéthanol 2 mM Glycérol 10% $MgCl_2$ 1 mM Imidazole 0 ou 50 mM	7,5	**E2**	Tampon W2 + imidazole 300 mM	7,5

Tableau 10 | Composition des tampons de lavage et d'élution testés pour la purification des protéines Tax-2. Les compositions des tampons de lyse L1 et L2 sont décrites dans le tableau 9.

2.3. La protéine de capside du FIV

En se basant sur les travaux de (Serriere *et al.*, 2013) pour la CA entière, la CA recombinante (CAΔC-PT) a été produite en système procaryote puis purifiée par chromatographie d'affinité pour le nickel.

2.3.1. Surproduction de la CA en système bactérien

Pour la surproduction de la CA recombinante, des bactéries d'expression BL21 (DE3) pLysS sont transformées par le plasmide pRSET-CAEΔC-PT. Les clones positifs sont inoculées dans du milieu LB additionné d'ampicilline (50 $\mu g.\mu L^{-1}$) et incubés toute la nuit à 37°C sous agitation.

L'ensemencement d'1L de milieu de culture LB liquide supplémenté en ampicilline (50 $\mu g.\mu L^{-1}$) se fait par ajout de 10 mL de préculture. La culture est ensuite incubée à 37°C sous agitation jusqu'à induction. Lorsque la DO_{600nm} atteint une valeur comprise entre 0,4 et 0,6, la culture est induite par ajout d'IPTG à une concentration finale de 1 mM et la température est abaissée à 25°C. Après une nuit d'incubation, les bactéries sont sédimentées par centrifugation à 6 000 × g pendant 10 minutes et conservées à -20°C en attendant l'étape de purification.

2.3.2. Purification de la CA

2.3.2.1. Lyse bactérienne

L'extraction des protéines recombinantes des bactéries s'est faite par lyse des culots bactériens dans du tampon LEW (*Lysis-Equilibration-Wash*) (**Tableau 11**) contenant 50 mM NaH_2PO_4 et 300 mM NaCl à pH 8, à raison de 5 mL de tampon par gramme de culot. Le tampon de lyse est également additionné de lysozyme (1 $mg.mL^{-1}$, Sigma) et de DNase I (2 $U.mL^{-1}$, Sigma). Après 1h d'incubation à 4°C, les bactéries sont lysées par sonication (ultrasons) avec 3 cycles de 3 minutes de pulsations à 40% de puissance.

Le lysat obtenu est ensuite centrifugé pendant 45 minutes à 11 000 × g à 10°C et le surnageant, correspondant à la fraction soluble, est filtré à l'aide d'un filtre de 0,45 µm de porosité (Millipore).

2.3.2.2. Purification de la CA par chromatographie d'affinité pour le nickel

Afin d'obtenir une protéine de CA pure, le surnageant obtenu est soumis à une purification par chromatographie d'affinité pour le nickel (**Matériels et méthodes 2.1.2**) qui permet de récupérer uniquement la protéine de capside CA fusionnée à son étiquette d'affinité His_6.

En pratique, suite à la filtration, le surnageant est mis en contact avec la résine Ni-TED pendant 1h à 4°C sur roue, à raison de 0,15 g de résine sèche (correspondant à 0,3 mL de colonne hydratée) par gramme de culot bactérien. Le mélange est ensuite transféré, par gravité, dans une colonne pour l'étape de purification. Après passage de la fraction non retenue (FT, *Flow-Through*), les billes sont lavées trois fois par 10 fois le volume de la colonne hydratée (CV, *Column Volume*) de tampon LEW. Enfin les protéines sont éluées par 50 mM d'imidazole grâce au passage de 4 élutions de 1,5 CV et 2 élutions finales de 3 CV de tampon d'élution (**Tableau 11**).

Les différentes fractions de lavage et d'élution sont analysées sur SDS-PAGE afin de vérifier la pureté de l'échantillon obtenu. Enfin, la quantité de protéines obtenues dans les fractions identifiées comme pures par SDS-PAGE est estimée par mesure de l'absorbance de l'échantillon à 280 nm à l'aide d'un spectrophotomètre NanoDrop.

Tampons		Composition	pH
Purification	Tampon LEW	50 mM NaH_2PO_4 300 mM NaCl	8
	Tampon d'élution	50 mM NaH_2PO_4 300 mM NaCl 50 mM imidazole	8
Stockage	Tampon HEPES	50 mM HEPES 100 mM NaCl	6,5
	Tampon Phosphate	50 mM NaH_2PO_4 100 mM NaCl	7,4

Tableau 11 | Composition des tampons utilisés pour la purification et le stockage de la CA.

2.3.3. Concentration de la CA et échange de tampon

La protéine de capside pure est ensuite concentrée selon le principe d'ultrafiltration à l'aide d'une membrane semi-perméable dont le seuil de coupure est choisi de telle sorte qu'elle reste dans le rétentat tandis que le tampon passe à travers le filtre.

Cette étape d'ultrafiltration permet également de modifier le tampon dans lequel se trouve la protéine afin d'éliminer l'imidazole de la solution protéique. Pour cela, le tampon de stockage choisi (HEPES ou Phosphate, **Tableau 11**) est ajouté dans le réservoir du concentrateur afin de diluer le tampon initial contenant de l'imidazole.

En pratique, une membrane en PolyEther Sulfone (PES) avec un seuil de coupure de 10 kDa (Sartorius) a été utilisée pour la concentration de la CA. Plusieurs centrifugations successives à 6 000 × g ont été réalisées jusqu'à atteindre une dilution finale d'au moins 100 fois du tampon initial (E) et la concentration protéique désirée.

2.3.4. Clivage de l'étiquette d'affinité de la CA du FIV

Afin d'éviter une quelconque interférence de l'étiquette d'affinité His_6 lors de l'étape de cristallogenèse, celle-ci est clivée grâce à la protéase Facteur Xa, qui reconnaît spécifiquement la séquence Ile-Glu-Gly-Arg introduite entre la protéine et l'étiquette d'affinité. Pour cela, 16 $U.mL^{-1}$ de Facteur Xa est ajouté à la solution protéique à 1 $mg.mL^{-1}$ en tampon Phosphate ou HEPES et le mélange est incubé à 19°C pendant la nuit.

La protéine est ensuite purifiée sur une résine composée d'anticorps anti-Facteur X (Sigma) selon le protocole du fournisseur afin d'éliminer la protéase, puis sur une résine Ni-NTA afin d'éliminer l'étiquette His_6 clivée ainsi que les protéines toujours étiquetées.

La protéine pure et sans son étiquette est ensuite concentrée par ultrafiltration et stockée à -80°C en vue de son utilisation ultérieure.

3. Analyse qualitative et quantitative des protéines

3.1. Électrophorèse sur gel discontinu de polyacrylamide en conditions dénaturantes

La présence des protéines d'intérêt dans les différentes fractions aliquotes recueillies au cours de la lyse bactérienne et de la purification des protéines est vérifiée par SDS-PAGE.

En pratique, avant migration sur gel, les protéines sont dénaturées par chauffage (en général 5 minutes à 95°C) dans un tampon de charge contenant du bleu de bromophénol, du glycérol, un agent réducteur et un détergent anionique, en général le SDS, conférant aux protéines une charge négative qui permet leur migration en fonction de leur masse moléculaire uniquement.

Après dépôt des échantillons et d'une solution de marqueurs de masses moléculaires connues (Prestained protein ladder, Fermentas) sur le gel, la migration est effectuée dans un tampon Tris-Glycine-SDS (0,25 M Tris-HCl pH 8,3, 1,92 M glycine, 1% (m/v) SDS, Euromedex) à 90 V pendant 10 minutes puis à 120 V jusqu'à ce que le front de migration atteigne la limite du gel. Après migration, le gel est coloré au bleu de Coomassie (PageBlue® Protein Staining Solution, Fermentas) selon le protocole du fournisseur. Ce bleu ayant la propriété de se fixer aux résidus d'acides aminés basiques et aromatiques des protéines, il permet de révéler les différentes protéines contenues dans les échantillons déposés sur gel.

Dans le cas des protéines Tax, des études antérieures effectuées au sein du laboratoire ont révélé que le SDS présent dans le tampon de charge entraînait, lors de la dénaturation à 95°C, le clivage de Tax au niveau des dipeptides Asp-Pro (Lamed *et al.*, 2001). Ainsi, un autre tampon de charge équivalent contenant du DTT comme agent réducteur et du LDS (*Lithium Dodecyl Sulfate*) pour conférer aux protéines leur charge négative, est utilisé pour l'analyse des protéines Tax-2 recombinantes.

3.2. *Western blot*

Cette technique permet de détecter et d'identifier des protéines à l'aide d'anticorps spécifiques. Après leur séparation par SDS-PAGE, les protéines ne sont pas colorées mais transférées sur une membrane de nitrocellulose (Hybond ECL Nitrocellulose Membrane, GE Healthcare). Ce transfert se fait dans un tampon 10 mM CAPS (*3-(cyclohexylamino)-1-propanesulfonic acid*) pH 11, méthanol 10% (v/v) grâce à l'application d'un courant électrique sous un ampérage constant de 400 mA pendant 45 minutes.

Après transfert et rinçage de la membrane en eau distillée, un blocage des sites d'interactions non spécifiques est réalisé par saturation de la membrane avec des protéines du lait (tampon PBS 1X, 0,1% (v/v) Tween®20 (Sigma), 5% (m/v) lait écrémé) pendant 1 heure à température ambiante, afin de minimiser les interactions non spécifiques entre la membrane et les anticorps spécifiques de l'étiquette His_6

La membrane est ensuite incubée pendant 1 heure avec la solution de saturation contenant l'anticorps primaire monoclonal de souris anti-histidines (clone HIS-1, Sigma) dilué au 1/1500 qui va reconnaître l'étiquette His_6 des protéines d'intérêt. Puis la membrane est lavée plusieurs fois avec un tampon PBS 1X, 0,1% (v/v) Tween®20 afin d'éliminer l'excès d'anticorps. L'anticorps secondaire de chèvre, reconnaissant les immunoglobulines G de souris et couplé à la phosphatase alcaline (Interchim), dilué à 1/6000 dans une solution PBS/Tween®20 est ensuite ajouté et incubé avec la membrane pendant 1 heure. La membrane est de nouveau lavée plusieurs fois avec une solution de PBS/Tween®20.

Enfin, la révélation des protéines se fait grâce au système de révélation SIGMAFAST™ BCIP®/NBT (Sigma) contenant un substrat appelé BCIP (5-Bromo-4-Chloro-3-Indolyl Phosphate) et un chromogène NBT (*NitroBlue Tetrazolium*). Après incubation avec la membrane, le BCIP est hydrolysé par la phosphatase alcaline en un produit intermédiaire qui va se dimériser. Durant cette dimérisation, un relargage de protons a lieu, provoquant la réduction du NBT en un produit final insoluble ayant une couleur bleu/violet. La présence des protéines d'intérêt sur la membrane se traduit donc par l'apparition de bandes bleues/violettes.

3.3. Chromatographie d'exclusion stérique

La chromatographie d'exclusion stérique permet la séparation de protéines en fonction de leur masse moléculaire et de leur forme sur une colonne contenant des granules de gel poreux. Les protéines de haute masse moléculaire, dont le diamètre est supérieur à celui des pores, sont exclues du gel et éluées en premier dans le volume mort tandis que les protéines de plus faible masse moléculaire sont éluées plus tardivement puisqu'elles passent par les pores du gel.

Cette technique a été utilisée lors de l'analyse structurale de la protéine de capside du FIV par diffusion des rayons X aux petits angles *(*SAXS*, Small Angle X-ray Scattering)*. La sortie de colonne (« Superdex S200 5/150 GL » (GE Healthcare)) des protéines est suivie par mesure de l'absorbance à 280 nm et la fraction d'élution contenant la protéine d'intérêt est soumise aux rayons X.

3.4. Diffusion dynamique de la lumière

La diffusion dynamique de la lumière (DLS, *Dynamic Light Scattering*) est une technique permettant de mesurer le rayon hydrodynamique R_h de particules en solution à partir de la mesure de leur coefficient de diffusion. Elle est basée sur le principe de diffusion de la lumière par des molécules soumises à des mouvements thermiques aléatoires, dits « browniens », dus à leurs collisions avec les molécules de solvant (Brown *et al.*, 1866). Plus une particule sera grosse, plus son mouvement brownien sera lent tandis que pour une particule de plus petite taille, ce mouvement sera plus rapide.

Pour déterminer ces coefficients de diffusion, la lumière diffusée est mesurée à différents temps ce qui permet de calculer la fonction d'auto-corrélation de l'intensité diffusée et de déterminer le coefficient de diffusion transversale D_T des particules en solution. Le R_h des particules est ensuite calculé à partir de ce coefficient grâce à l'équation de Stokes-Einstein :

$$R_h = {(K_B \cdot T)}\Big/{(6\pi \cdot \eta \cdot D_T)}$$

$$Avec\ R_h : rayon\ hydrodynamique, K_B : constante\ de\ Boltzmann, T : température,$$

$$\eta : viscosité\ et\ D_T : coefficient\ de\ diffusion\ transversale\ des\ particules.$$

Une droite d'étalonnage représentant les diamètres hydrodynamiques obtenus pour des standards protéiques globulaires de masse moléculaire connue permet d'estimer la masse moléculaire des protéines à partir de leur R_h. L'état d'oligomérisation des protéines en solution est ensuite déduit de leur masse moléculaire apparente.

La DLS permet également d'estimer le coefficient de polydispersité (PdI) d'une solution : une solution protéique étant considérée comme monodisperse si elle présente un PdI inférieur à 20%.

Les expériences de DLS, pour évaluer l'état d'oligomérisation des protéines purifiées, ont été réalisées à 20°C avec un système de mesure Zetasizer Nano S (Malvern Instruments) dans lequel est placé une cuve (Brand) contenant la solution protéique d'intérêt pure et filtrée sur une membrane de 0,02 µm. Chaque mesure a été réalisée en triplicat et résulte de la moyenne de 12 mesures indépendantes.

4. Suivi de l'assemblage de la CA du FIV *in vitro* et identification d'inhibiteurs

La protéine de capside du FIV étant capable de s'oligomériser afin de former la capside virale mature (Serriere *et al.*, 2013), des tests de suivi de l'assemblage de la CA *in vitro* ont été mis au point au laboratoire.

4.1. Suivi de l'assemblage de la CA *in vitro* par DLS

La DLS permettant de mesurer le rayon hydrodynamique R_h de particules en solution, elle a été utilisée afin de suivre l'assemblage de la CA au cours du temps en présence de sels.

Le suivi de l'assemblage de la CA à 7 $mg.mL^{-1}$ en tampon HEPES (50 mM HEPES, 100 mM NaCl, pH 6,5) a été réalisé, afin de vérifier qu'elle soit bien fonctionnelle dans ce tampon au pH acide, utilisé lors de l'étape de cristallogenèse. La stabilité de la CA en solution avant l'ajout de sels est vérifiée puis 1 M final de NaCl est ajouté dans la cuvette afin d'initier l'assemblage capsidique. Les expériences sont réalisées à 37°C dans l'appareil de mesure Zetasizer Nano S et les mesures sont effectuées toutes les 2 minutes pendant 80 minutes.

4.2. Identification d'inhibiteurs de l'assemblage de la CA *in vitro* par spectrophotométrie UV

De nombreux composés, issus de la chimiothèque du Dr. G. Alvarez (*Universidad de la Republica*, Paysandú, Uruguay), ont été testés *in vitro* dans le but d'évaluer leur effet sur l'oligomérisation de la CA et d'identifier un inhibiteur potentiel de l'assemblage capsidique. Ces composés sont synthétisés chimiquement à partir de produits naturels tels que les plantes et présentent un faible coût de production.

La technique de DLS utilisée pour le suivi de l'assemblage de la CA *in vitro* ne permettant le criblage que d'une molécule à la fois, avec un temps d'expérimentation assez long et demandant une quantité de protéine importante, un nouveau protocole de suivi de l'assemblage par spectrophotométrie, sans ajout de sels, a été mis au point afin de cribler un grand nombre de molécules simultanément.

La spectrophotométrie permet, par mesure de l'absorbance de la solution protéique à 340 nm, de suivre l'assemblage de la CA au cours du temps. En effet, lors de l'assemblage de la CA la turbidité de la solution augmente au fil du temps ce qui permet une mesure d'absorption de la

solution par spectrophotométrie UV. Ce procédé pouvant s'effectuer en plaque 384 puits, un grand nombre de molécules peuvent être testées simultanément en quelques heures, facilitant ainsi le criblage d'inhibiteurs de l'assemblage capsidique.

Les mesures sont effectuées sur un appareil Tecan Infinite M1000, disponible sur la plateforme *Protein Science Facility* (PSF, Lyon) et sont réalisées en plaque 384 puits avec une CA purifiée et concentrée à 5 $mg.mL^{-1}$ en tampon Phosphate. Les mesures d'absorbance à 340 nm sont réalisées toutes les 30 secondes pendant 30 minutes à 39°C.

En pratique, 25 µL de solution protéique est déposée dans chaque puits, puis le composé est ajouté à une concentration finale de 25 µM. La plaque est ensuite agitée pendant 5 secondes avant le début des mesures. Les différents composés testés étant solubilisés dans du DMSO (diméthylsulfoxide) ou du glycérol, des expériences témoins sont réalisées dans des puits contenant uniquement la solution protéique seule, du DMSO/glycérol seul (0,8% (v/v)) et la solution protéique en présence de DMSO/glycérol (sans composés).

5. Étude structurale des protéines d'intérêt

5.1. Étude par diffusion des rayons X aux petits angles

5.1.1. Principe de la diffusion des rayons X aux petits angles

La diffusion des rayons X aux petits angles (SAXS) est une méthode d'analyse structurale permettant d'obtenir des informations à basse résolution sur la conformation et l'assemblage de macromolécules en solution (Putnam *et al.*, 2007).

En principe, un faisceau monochromatique de rayons X irradie une solution de macromolécules qui va alors diffuser ces rayons dans l'espace selon un angle de diffusion faible (**Figure 23**).

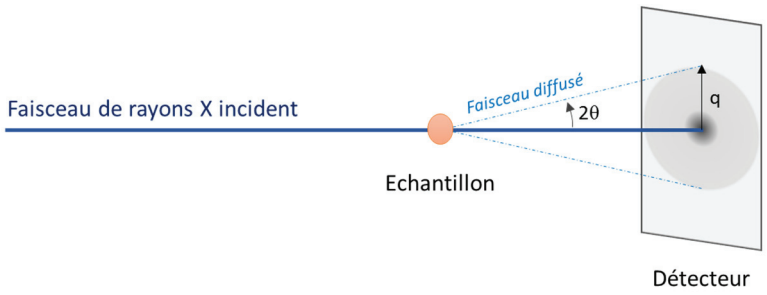

Figure 23 | Schématisation de la diffusion des rayons X aux petits angles (SAXS). D'après (Putnam *et al.*, 2007). L'échantillon (orange) est placé dans le collimateur du faisceau de rayons X incident (bleu) et diffuse ce rayon selon un angle de diffusion 2θ. Les intensités de diffusion sont mesurées sur le détecteur placé à la sortie du rayon incident.

Les intensités de diffusion des rayons X mesurées sur le détecteur permettent ainsi de calculer différents paramètres caractérisant la macromolécule étudiée :

- Le rayon de giration (R_g), reflétant la taille de la macromolécule, est calculé grâce à la courbe de diffusion (**Figure 24A**) obtenue à partir des valeurs d'intensités mesurées :

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

$$Avec\ I(0) : valeur\ de\ l'intensité\ diffusée\ à\ angle\ nul,$$

$$I(q) : intensités\ diffusées\ mesurées\ à\ l'angle\ de\ diffusion\ 2\theta$$

$$et\ q : vecteur\ de\ diffusion\ (q = 4 \cdot \pi \cdot sin\theta / \lambda)$$

– La masse moléculaire observée, permettant de déterminer l'état d'oligomérisation de la macromolécule, est calculée à partir de la valeur d'intensité diffusée à angle nul I(0) (**Figure 24A**) :

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

$$Avec\ V_{sol} : volume\ de\ la\ solution\ protéique, c : concentration\ protéique,$$

$$M_w : masse\ moléculaire, N : nombre\ d'Avogadro,$$

$$m_p : nombre\ d'électrons\ dans\ la\ molécule,$$

$$\rho_s : densité\ du\ solvant$$

$$et\ V_p : volume\ spécifique\ partiel\ de\ la\ molécule$$

– La fonction de distribution des distances P(r) (**Figure 24B**), décrivant l'ensemble des distances interatomiques de la macromolécule, dont la distance interatomique maximale D_{max}, est calculée par transformée de Fourier inverse des intensités mesurées.

Ces différentes données permettent ensuite de calculer *ab initio* le modèle d'enveloppe SAXS (**Figure 24C**) de la macromolécule en solution.

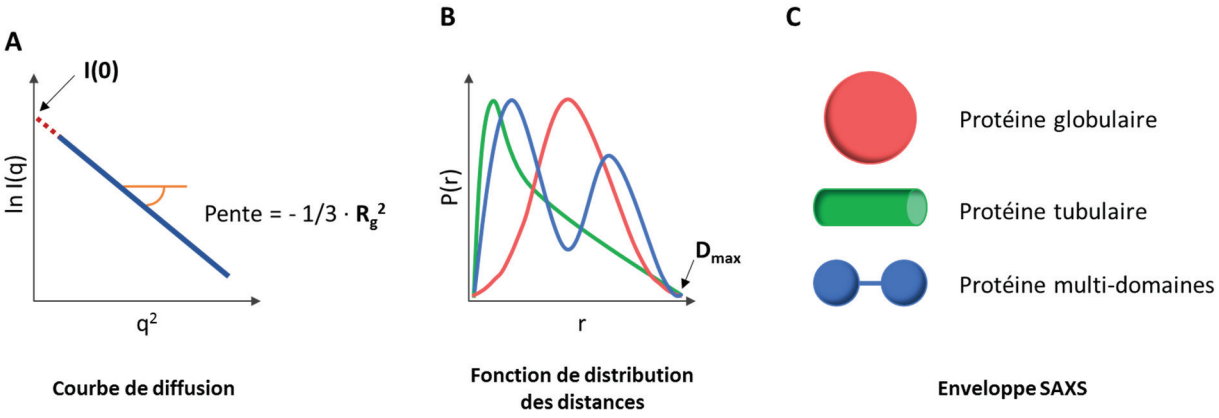

Figure 24 | (A) Courbe de diffusion et (B) fonction de distribution des distances permettant de calculer (C) le modèle *ab initio* d'une macromolécule, avec des contraintes de R_g et de D_{max}. D'après (Koch et Svergun, 2003).

5.1.2. Acquisition et analyse des données

Les données SAXS ont été enregistrées sur la ligne de lumière BM29 de l'*European Sychrotron Radiation Facility* (ESRF) de Grenoble, à température ambiante avec un faisceau de rayons X incidents de longueur d'onde λ = 0,991 Å. Cette technique nécessitant un échantillon protéique monodisperse, l'acquisition des données a été précédée d'une étape de chromatographie d'exclusion stérique (**Matériels et méthodes 3.3**) permettant de séparer les différents états d'oligomérisation de la protéine d'intérêt.

Dans le cas de la CA du FIV, les tampons HEPES (50 mM HEPES, 100 mM NaCl, pH 6,5) et Phosphate (50 mM NaH_2PO_4, 100 mM NaCl, pH 7,4) ont été testés et les échantillons injectés ont été concentrés entre 1 et 28 $mg.mL^{-1}$.

Les données SAXS obtenues ont été analysées en utilisant la suite ATSAS (Petoukhov *et al.*, 2012) afin d'obtenir les paramètres à utiliser en tant que contraintes pour le calcul de l'enveloppe SAXS : le R_g de la protéine en solution étant calculé par le logiciel PRIMUS (Petoukhov *et al.*, 2012) tandis que la P(r) et le D_{max} sont calculés par GNOM (Petoukhov *et al.*, 2012). Les 50 modèles basses résolutions générées *ab initio* par le programme DAMMIF (Franke et Svergun, 2009), à partir des contraintes fournies, ont ensuite été moyennés et filtrés par DAMAVER (Volkov et Svergun, 2003) pour obtenir l'enveloppe finale SAXS. La visualisation de l'enveloppe SAXS et l'ajustement de la structure cristallographique de la CA du FIV (PDB ID : 5NA2) dans celle-ci ont été effectués par le logiciel Chimera (Pettersen *et al.*, 2004). La qualité de cet ajustement a ensuite été estimée à l'aide du programme Crysol (Svergun *et al.*, 1995).

Les différentes étapes de la technique de SAXS, de l'enregistrement des données à la résolution de l'enveloppe SAXS, sont récapitulées en **Figure 25**.

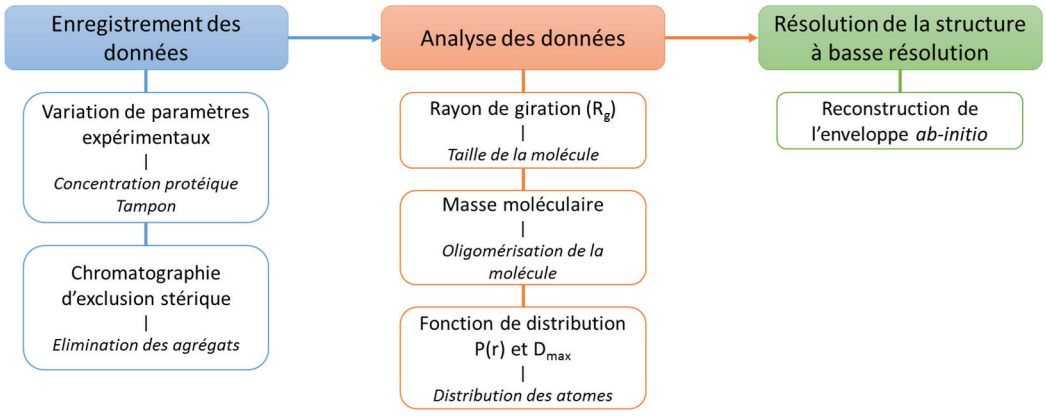

Figure 25 | Schéma récapitulatif de la collecte et l'analyse des données SAXS pour la résolution d'une structure à basse résolution.

5.2. Étude structurale par cristallographie aux rayons X

5.2.1. Cristallogenèse

5.2.1.1. Principe de la cristallogenèse

Sous certaines conditions, les protéines en solution peuvent s'assembler et s'empiler afin de former un édifice solide périodique, tridimensionnel et ordonné, appelé cristal. Cet édifice cristallin résulte principalement en l'augmentation progressive de la concentration en protéines et en agent précipitant d'une solution jusqu'à atteindre la zone de nucléation où, en raison de la saturation en protéines, des cristaux commencent à se former (**Figure 26**).

La croissance cristalline fait intervenir des sels, des tampons ainsi qu'un agent précipitant dont la nature et la concentration influent sur la solubilité de la protéine, elle-même spécifique de chaque protéine.

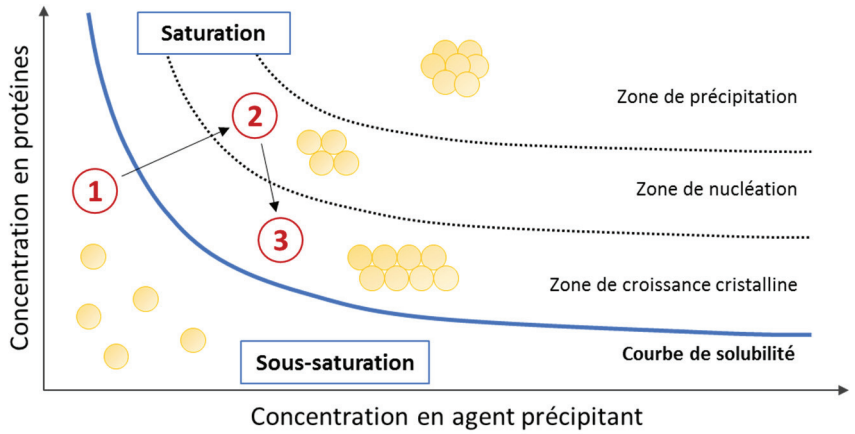

Figure 26 | Représentation schématique d'un diagramme de phase. D'après (Chayen et Saridakis, 2008).
La courbe de solubilité (bleu) sépare la zone de sous-saturation des trois zones de saturation (croissance cristalline, nucléation et précipitation). La formation de cristaux se fait lorsque les protéines (jaune) en solution (1) sont suffisamment concentrées pour former des germes cristallins (2) avant d'atteindre un équilibre de concentrations permettant la croissance de ces germes (3).

La première étape de la cristallogenèse, consistant en la recherche d'une condition de cristallisation spécifique d'une protéine donnée, est souvent limitante et nécessite le criblage d'un très grand nombre de conditions différentes. La variation de certains paramètres tels que la concentration en protéines, la nature et la concentration de l'agent précipitant, la température de croissance, le pH de la solution ou encore la présence de sels permet d'obtenir

la combinaison optimale pour la croissance de cristaux de taille suffisante en vue d'une analyse par diffraction aux rayons X.

Il existe plusieurs méthodes de cristallisation, dont l'objectif est de porter une solution de la zone de sous-saturation à la zone de croissance cristalline (**Figure 26**). Parmi celles-ci, la méthode de cristallisation par diffusion de vapeur est la mieux adaptée pour la cristallisation des protéines. En effet, elle permet une augmentation douce et progressive de la concentration en protéine. Elle consiste en la diffusion d'espèces volatiles (eau, solvants organiques), dans un système clos, de la goutte composée d'un agent précipitant et de la protéine d'intérêt vers le réservoir contenant ce même agent précipitant. Cette diffusion, due à la tension de vapeur entre les deux milieux, a lieu jusqu'à atteindre un équilibre de concentration en agent précipitant entre le réservoir et la goutte. Elle entraîne avec elle une modification du volume de la goutte et par conséquent une variation de la concentration de ses constituants, dont l'augmentation lente de la concentration en protéines par rapport à l'état initial.

La méthode de cristallisation par diffusion de vapeur peut être mise en œuvre selon deux procédés différents : les techniques de la goutte assise et de la goutte suspendue (**Figure 27**).

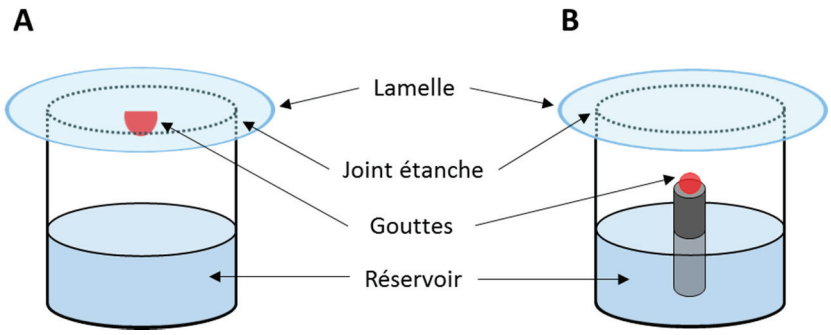

Figure 27 | Représentations schématiques des deux techniques mises en œuvre pour la méthode de diffusion de vapeur : (A) technique de la goutte suspendue et (B) technique de la goutte assise. La goutte contient un mélange de solution protéique et d'agent précipitant tandis que le réservoir ne contient que l'agent précipitant. L'ensemble est fermé hermétiquement par un joint étanche.

La technique de la goutte assise est généralement utilisée pour le criblage de conditions de cristallisation en plaques de 96 puits tandis que celle de la goutte suspendue est plus couramment utilisée pour l'optimisation de conditions de cristallisation en plaques de 24 puits.

5.2.1.2. Criblage des conditions de cristallisation

Les essais de criblage de conditions de cristallisation des protéines d'intérêt sont réalisés dans des plaques de 96 puits (MRC 2 Well Crystallization Plate, Hampton Research). Ces plaques sont constituées de réservoirs contenant 80 µL d'agent précipitant et de deux plateformes par réservoir, sur lesquelles sont déposées des gouttes de 400 nL résultant du mélange de 200 nL de solution protéique et du même volume d'agent précipitant.

Le dépôt de ces nanogouttes est réalisé à l'aide du robot nanopipetteur Mosquito (TTP Labtech) et les plaques sont ensuite conservées dans la chambre à cristaux thermostatée à 19°C du robot de visualisation automatique RockImager 182 (Formulatrix). Les deux robots utilisés sont disponibles au sein de la plateforme PSF de Lyon.

Les paramètres permettant la formation de cristaux pour les protéines d'intérêt étant encore inconnus, de nombreuses conditions de cristallisation sont testées en faisant varier :

- Les kits commerciaux de criblage de conditions de cristallisation utilisés, permettent notamment de tester différents agents précipitants (sels ou polyéthylène glycol (PEG)) à différentes concentrations et différents pH (4 à 9,5) ;
- Le tampon de stockage et la concentration de la protéine d'intérêt ;

Une fois la condition de cristallisation trouvée, elle peut être affinée et optimisée en faisant varier légèrement les conditions de cristallisation afin d'obtenir des cristaux de meilleure qualité pour une analyse par diffraction aux rayons X. Pour cette optimisation, la technique de la goutte suspendue est utilisée. Le dépôt de microgouttes (2 µL) contenant différents ratios solution protéique:agent précipitant (1:1, 1:2, 2:1) se fait manuellement sur des lamelles siliconées qui seront par la suite déposées sur des plaques de 24 puits constituées de grands réservoirs (contenant 500 µL d'agent précipitant) et placées à 19°C.

5.2.1.3. Technique de co-cristallisation

La technique de co-cristallisation est utilisée pour obtenir des cristaux de protéines en présence d'un partenaire biologique ou d'inhibiteurs potentiels (dans le cas de la CA du FIV). Pour cela, le complexe protéine:partenaire est formé en ajoutant directement le partenaire seul lors du dépôt de la goutte contenant la protéine d'intérêt et l'agent précipitant. Les conditions de cristallisation du complexe sont alors criblées afin d'obtenir un cristal contenant les deux partenaires en interaction.

Dans le cas de la CA du FIV, la condition de cristallisation identifiée pour la protéine seule a été utilisée comme guide pour la co-cristallisation en présence de composés potentiellement inhibiteurs, bien que les conditions de cristallisation des complexes puissent varier de celles des protéines seules. En pratique, le composé est dilué dans la solution de cristallisation à une concentration finale de 2,5 mM en présence de 10% DMSO. Le composé est ensuite ajouté à la solution protéique (7 $mg.mL^{-1}$) à raison d'un ratio 1/1 (v/v) pour former une goutte de 2 µL, selon la technique de la goutte suspendue.

Avant d'être illluminés par un faisceau de rayons X, les cristaux obtenus sont pêchés et cryo-protégés par trempage dans une solution d'éthylène glycol (10% à 20% (v/v)). Cette étape de cryo-protection permet d'empêcher la formation de microcristaux de glace autour des cristaux de protéines lors de la congélation en azote liquide. La congélation des cristaux, elle, permet d'éviter les dégâts de radiation dus aux rayons X et, par conséquent, leur dégradation.

Les données de diffraction des cristaux ont ensuite été enregistrées sur la ligne ID30-B à l'ESRF de Grenoble, grâce au rayonnement synchrotron d'une longueur d'onde de l'ordre de l'angström (Å), dans le but d'obtenir des jeux de données de diffraction à haute résolution qui permettraient de résoudre la structure de la protéine d'intérêt.

5.2.2. Principe de la diffraction aux rayons X

La diffraction aux rayons X est basée sur le phénomène de diffraction de ces rayonnements par les atomes d'un cristal dans des directions privilégiées de l'espace. Cette diffraction résulte de l'interaction entre les couches électroniques de chaque atome du cristal et le faisceau incident. En raison de la périodicité du cristal, des phénomènes d'interférences constructives et destructives apparaissent lors de la diffraction des rayons X, déterminant ainsi les différentes directions privilégiées de réémission des ondes dans l'espace.

Les ondes réémises par un cristal dans des directions bien précises, appelées réflexions, sont repérées par les indices (h,k,l) dans l'espace réciproque. Elles sont caractérisées par une phase (φ) et une amplitude (|F|), calculée à partir des intensités mesurées (I(*hkl*)) de leurs taches de diffraction (**Figure 28**).

Expérimentalement, seules les intensités I(*hkl*) des ondes diffractées peuvent être mesurées par le détecteur et permettent d'avoir accès à leurs amplitudes :

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

$$Avec\ |F(hkl)| : amplitude\ de\ l'onde$$

Les informations de phase φ_{hkl}, elles, ne sont pas mesurables lors de l'expérimentation. Des méthodes de phasage, comme celle du remplacement moléculaire, sont nécessaires pour estimer les phases des ondes diffractées ; afin d'en déduire les cartes de densité électronique.

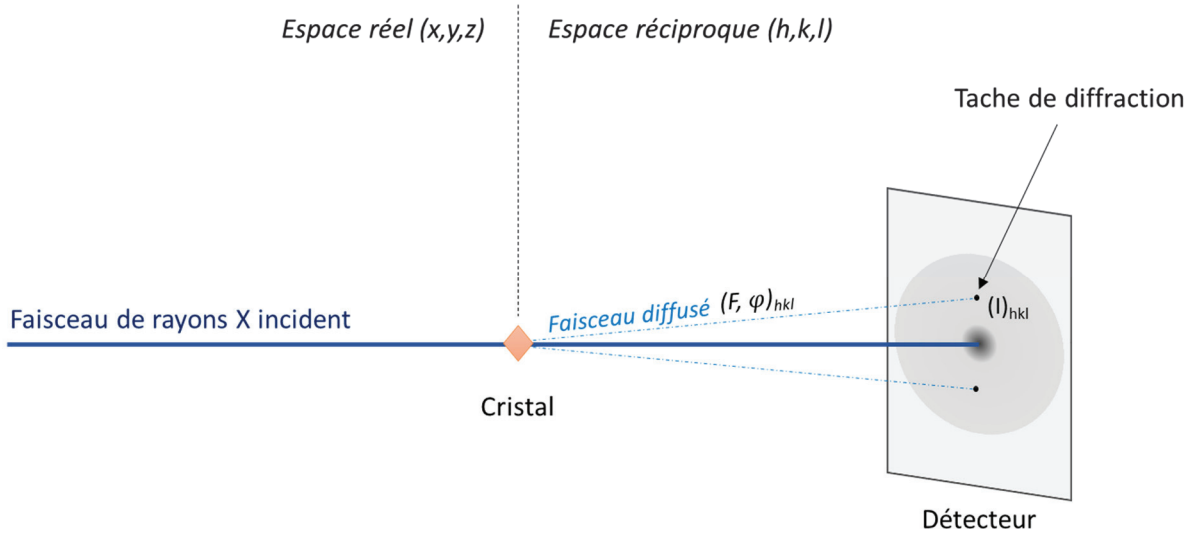

Figure 28 | Représentation schématique de la diffraction d'un cristal aux rayons X. Au contact du cristal, le faisceau incident est réémis dans l'espace réciproque (h,k,l) et est à l'origine d'une tache de diffraction d'intensité (I) mesurée par le détecteur. Le faisceau diffusé, lui, est caractérisé par une information d'amplitude (|F|) et une information de phase (φ).

Les informations d'amplitude et de phase obtenues permettent ensuite de générer les cartes de densité électronique nécessaires à la construction de la structure cristallographique. En effet, l'onde diffractée est définie par un facteur de structure F*(hkl)*, combinant sa phase et son amplitude, qui permet, par la suite, le calcul de la fonction de densité électronique du cristal.

Ce facteur de structure F*(hkl)* est égal à la somme des facteurs de diffusion atomique des atomes de la maille cristalline (de coordonnées x_j, y_j, et z_j), en considérant les décalages de phase des ondes diffusées par les différents atomes :

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

$$Avec\ hkl : indices\ du\ réseau\ réciproque,\ \varphi_{hkl} : phase\ pour\ la\ réflexion\ hkl,$$
$$f_j : facteur\ de\ diffusion\ de\ l'atome\ j\ et\ n : nombre\ d'atomes\ dans\ la\ maille$$

A partir de ce facteur de structure F*(hkl)*, il est possible de calculer la fonction de densité électronique ρ(x,y,z) pour toutes les positions (x,y,z) du cristal. Cette densité électronique correspond à la répartition des électrons du cristal par unité de volume et est égale à la transformée de Fourier inverse de F*(hkl)* :

$$\rho(hkl) = \frac{1}{V} \cdot \sum_{h=\infty}^{+\infty} \sum_{k=\infty}^{+\infty} \sum_{l=\infty}^{+\infty} F(hkl) \cdot e^{-2\pi \cdot i(hx_j + ky_j + lz_j)}$$

$$Avec\ V : volume\ du\ cristal\ de\ protéine$$

Ainsi, lorsque l'amplitude et la phase de toutes les ondes diffractées sont connues, il est possible de remonter, par le calcul, à la structure de la protéine au sein du cristal.

5.2.3. Analyse des données de diffraction

Les données de diffraction sont enregistrées sous forme d'un ensemble d'images, appelées clichés de diffraction, comportant des taches de diffraction dont l'intensité est mesurable.

5.2.3.1. Traitement des données de diffraction

La première étape du traitement des clichés de diffraction consiste en leur indexation, qui permet d'attribuer à chaque tache de diffraction des coordonnées (h,k,l) dans l'espace réciproque. Au cours de cette étape, différents paramètres de maille et groupes d'espace sont proposés par le logiciel utilisé, pour le cristal étudié.

Après l'indexation et choix du groupe d'espace adéquat, les intensités des taches de diffraction sont intégrées. La valeur du bruit de fond aux alentours de chaque tache est alors soustraite à sa valeur d'intensité.

Enfin, les intensités mesurées pour une même réflexion pouvant varier d'une image à l'autre, leurs valeurs sont mises à l'échelle, puis moyennées de façon à obtenir des réflexions uniques.

La qualité des jeux de données est estimée par de nombreux paramètres dont :

- Le facteur d'accord R_{sym}, évaluant l'écart entre les intensités de réflexions équivalentes et leur moyenne ;
- Le rapport signal sur bruit des intensités ($I/\sigma(I)$) ;
- La complétude, correspondant au rapport entre le nombre de réflexions uniques théoriques et mesurées ;
- La redondance, caractérisant le nombre de fois où l'intensité d'une réflexion unique est mesurée.

Dans le cas de la CA du FIV, les intensités diffractées sont intégrées puis mises à l'échelle avec les programmes XDS et XSCALE (Kabsch, 2010), respectivement.

Les informations de phase, non mesurables, étant manquantes lors de l'acquisition des données, elles peuvent être calculées grâce à trois méthodes différentes : le remplacement moléculaire, le remplacement isomorphe multiple et la méthode du signal anomal. La méthode de remplacement moléculaire, seule utilisée au cours de ce projet, sera développée dans la partie suivante.

5.2.3.2. Solution du problème de phase par remplacement moléculaire

Le remplacement moléculaire (Rossmann, 1990) utilise les informations de phase d'un modèle structural proche de la protéine étudiée (dit modèle-guide) et les intensités mesurées expérimentalement pour calculer la fonction de densité électronique de la protéine d'intérêt.

Le choix du (ou des) modèle(s)-guide(s) se base sur la similarité de séquence entre les protéines. En général, une identité supérieure à 20% peut permettre une similarité structurale suffisante avec la protéine d'intérêt. En outre, le (ou les) modèle(s)-guide(s) peu(ven)t modéliser tout ou partie de la structure d'intérêt.

En pratique, la structure de la protéine homologue est placée dans la maille de la structure à résoudre selon différentes orientations et positions, jusqu'à obtenir une orientation et une position optimale de la protéine, pour lesquelles les intensités calculées d'après la position du modèle-guide sont en accord avec les intensités observées expérimentalement (protéine d'intérêt). Ceci permet d'estimer des phases pour les données expérimentales, à partir des

facteurs de structure calculés (F_{calc}) à partir du (des) modèle(s)-guide(s). Une carte de densité électronique peut ainsi être générée pour la protéine d'intérêt.

Au cours de ce projet, cette technique a été utilisée pour résoudre la structure tridimensionnelle de la protéine de capside du FIV à l'aide du programme MrBUMP (Keegan et Winn, 2008) de la suite CCP4 (Winn *et al.*, 2011). Les structures du CA_{NTD} de Relik (*Rabbit Endogenus Lentivirus type K*) (PDB ID : 2XGU) (Goldstone *et al.*, 2010) et du CA_{CTD} de la CA du FIV (PDB ID : 5DCK) (Khwaja *et al.*, 2016) ont été utilisés comme modèles-guides.

5.2.3.3. Affinement de la structure

Une fois la carte de densité électronique calculée et le modèle construit pour la protéine d'intérêt, il est nécessaire d'affiner ses différents paramètres structuraux (position spatiale des atomes, facteur d'agitation thermique, d'occupation atomique, etc.). Cette étape permet d'améliorer l'adéquation entre le modèle construit (F_{calc}) et les données expérimentales (F_{obs}) mais aussi entre le modèle et les contraintes physico-chimiques intrinsèques au repliement d'une protéine.

Plusieurs cycles d'affinement sont nécessaires afin d'obtenir une conformation stable et de moindre énergie pour la protéine.

Au cours de ces cycles, l'évolution de la qualité du modèle est suivie grâce aux facteurs d'accord résiduels R_{work} et R_{free}, qui calculent l'adéquation entre les facteurs de structure observés (F_{obs}) et les facteurs de structure calculés pour le modèle construit (F_{calc}). Le facteur $R_{free,}$ qui utilise une proportion de réflexions choisies aléatoirement n'étant pas incluses dans le processus d'affinement (typiquement 5 % des réflexions totales), constitue un indicateur non-biaisé de l'avancement de l'affinement.

La structure de la CA du FIV a été affinée avec le programme Refine de la suite PHENIX (Adams *et al.*, 2011). La reconstruction du modèle atomique dans les cartes de densité électroniques a été réalisée avec le programme WinCOOT (Emsley *et al.*, 2010).

RESULTATS ET DISCUSSIONS

1. Étude de la protéine Tax-2 du virus HTLV-2

La première partie de ce projet doctoral porte sur l'étude de la protéine transactivatrice de la transcription Tax-2 du virus HTLV-2. Cette protéine étant un élément crucial pour la réplication et la pathogénicité de ce virus, une meilleure connaissance de sa structure, jusqu'alors inconnue, pourrait conduire à la découverte de nouvelles stratégies thérapeutiques contre ce virus.

Le protocole d'expression de la protéine Tax précédemment mis au point au laboratoire ne permettant pas l'expression de la protéine Tax-2 native sous forme soluble, plusieurs constructions plus ou moins grandes de celle-ci seule ont été réalisées, tout en conservant un maximum de domaines fonctionnels. De plus, une construction plasmidique permettant la co-expression de Tax-2 avec son partenaire cellulaire CREB a été réalisée. En effet, le recrutement du facteur de transcription CREB par Tax étant crucial pour sa fonction de transactivation de la transcription, l'étude du complexe Tax-2:CREB permettrait de mieux comprendre les mécanismes impliqués dans cette fonction.

1.1. Constructions plasmidiques

Après amplification par PCR, les gènes *tax-2* et *creb* d'intérêt ont été clonés dans les vecteurs d'expression pET-30a et/ou pETM-30 pour les protéines Tax-2 seules et dans le vecteur pET-DUET1 pour la co-expression de Tax-2 entière (non optimisée) avec CREB.

Les différentes protéines Tax-2 et CREB attendues, en fusion avec leur étiquette d'affinité et/ou de solubilité, sont présentées en **Figure 29**.

Les masses moléculaires théoriques attendues (**Tableau 12**) pour les différentes protéines produites ont été estimées par le logiciel ProtParam (http://web.expasy.org/protparam/), à partir de leurs séquences en acides aminés, elles-mêmes déduites de leurs séquences en acides nucléiques.

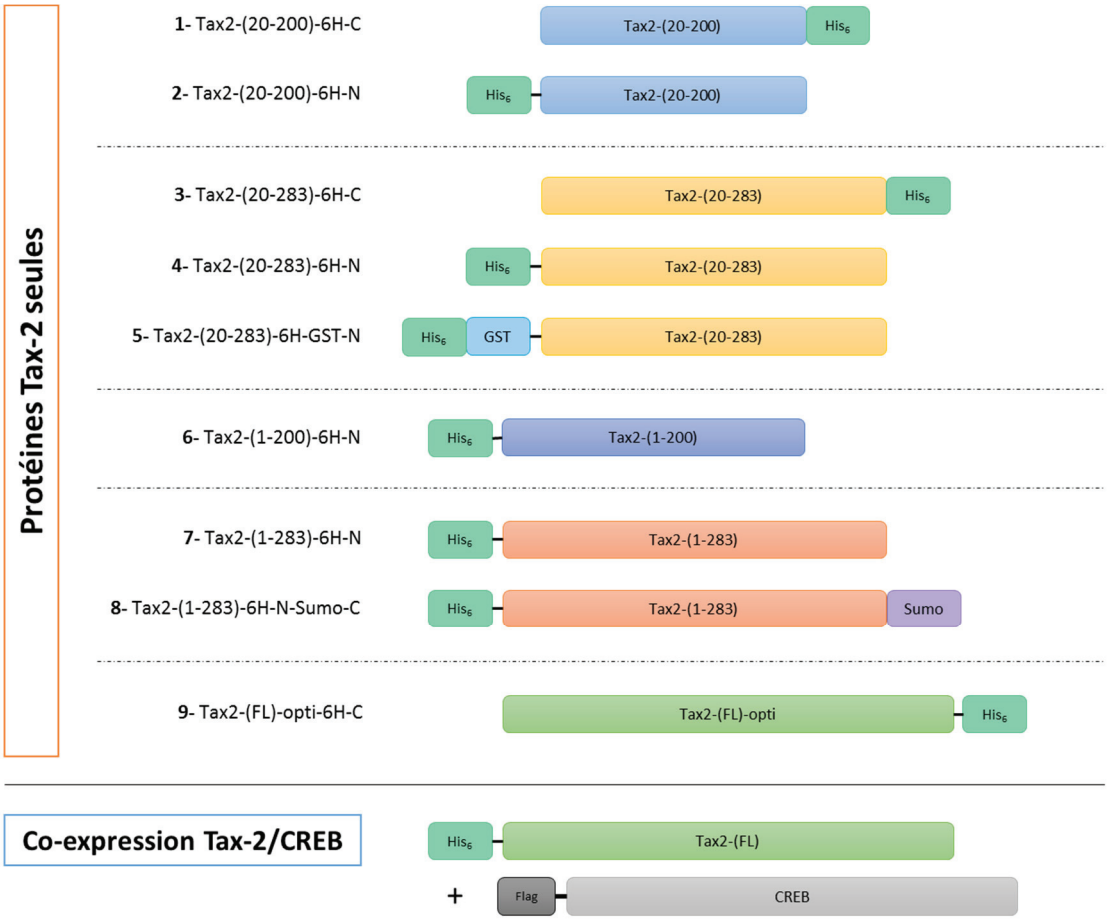

Figure 29 | Schématisation des protéines de fusion Tax-2 seules et de Tax-2/CREB attendues.

		Masses moléculaires théoriques (kDa)
Expression Tax-2 seules		
1	Tax2-(20-200)-6H-C	22,1
2	Tax2-(20-200)-6H-N	22,3
3	Tax2-(20-283)-6H-C	30,9
4	Tax2-(20-283)-6H-N	33,3
5	Tax2-(20-283)-6H-GST-N	56,3
6	Tax2-(1-200)-6H-N	24,3
7	Tax2-(1-283)-6H-N	33,4
8	Tax2-(1-283)-6H-N-Sumo-C	44,4
9	Tax2-(FL)-opti-6H-C	39,3
Co-expression Tax-2 et CREB		
10	Tax2-(FL)-CREB	Tax-2 : 38,9 CREB : 37,9

Tableau 12 | Masses moléculaires théoriques attendues pour les protéines Tax-2 et CREB.

1.2. Expression et purification des protéines Tax-2 seules

Après transformation de bactéries compétentes *E. coli* par les plasmides recombinants contenant les gènes d'intérêt, différents tests d'expression ont été réalisés. Pour cela, plusieurs conditions de surproduction, faisant varier bactéries d'expression, milieux de culture, températures d'incubation, temps d'induction et concentrations en IPTG, ont été testées pour chaque construction. Les stratégies d'expression, de solubilisation et de purification des protéines Tax-2 sont présentées en **Figure 30**.

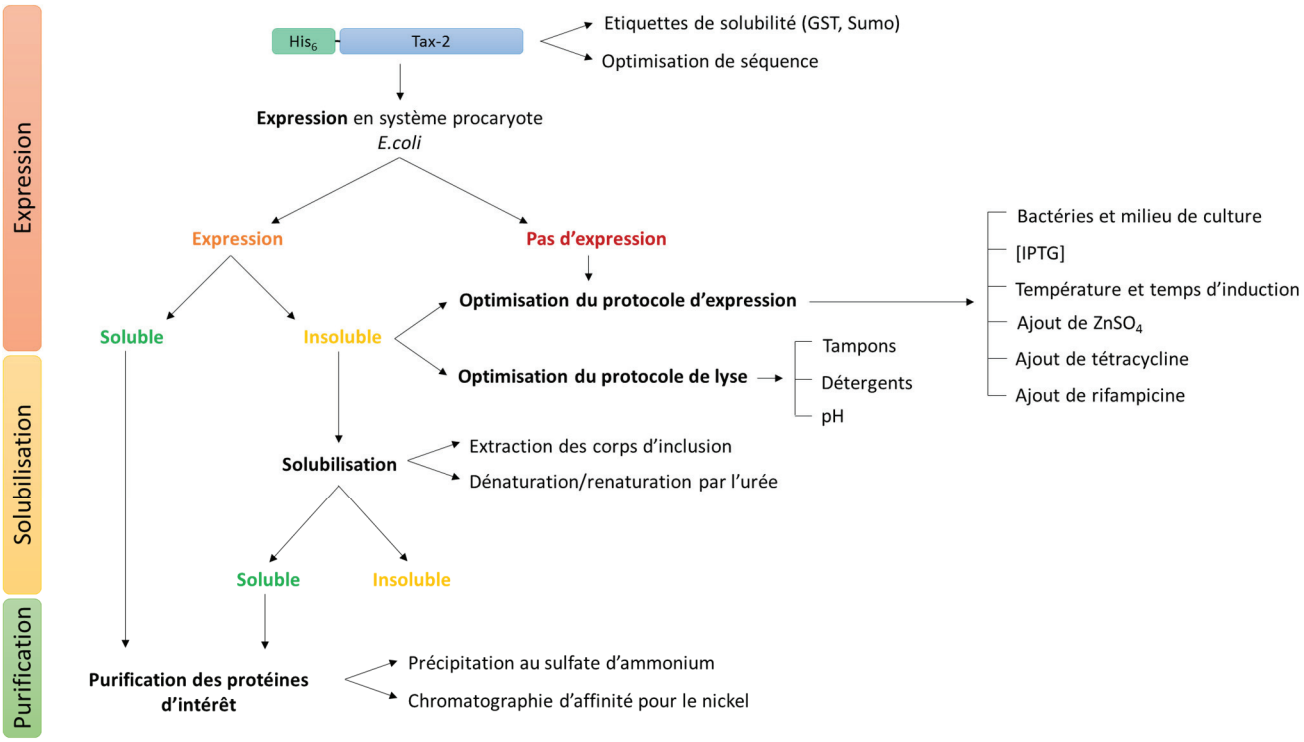

Figure 30 | Stratégies d'expression, de solubilisation et de purification des protéines Tax-2.

1.2.1. Tests d'expression

L'expression des protéines sous forme soluble a ensuite été analysée par SDS-PAGE et par *western blot*. Les tableaux suivants (**Tableau 13** et **Tableau 14**) résument les différentes conditions d'expression testées pour les constructions Tax-2 et leurs résultats.

De manière générale, l'utilisation de différentes souches bactériennes pour l'expression de Tax-2, d'étiquettes de solubilité ou encore l'utilisation de différents tampons pour la lyse bactérienne n'ont pas permis d'améliorer sa production et/ou sa solubilité. Les résultats plus détaillés pour l'expression, la solubilisation et la purification de chaque construction sont présentés dans les chapitres suivants.

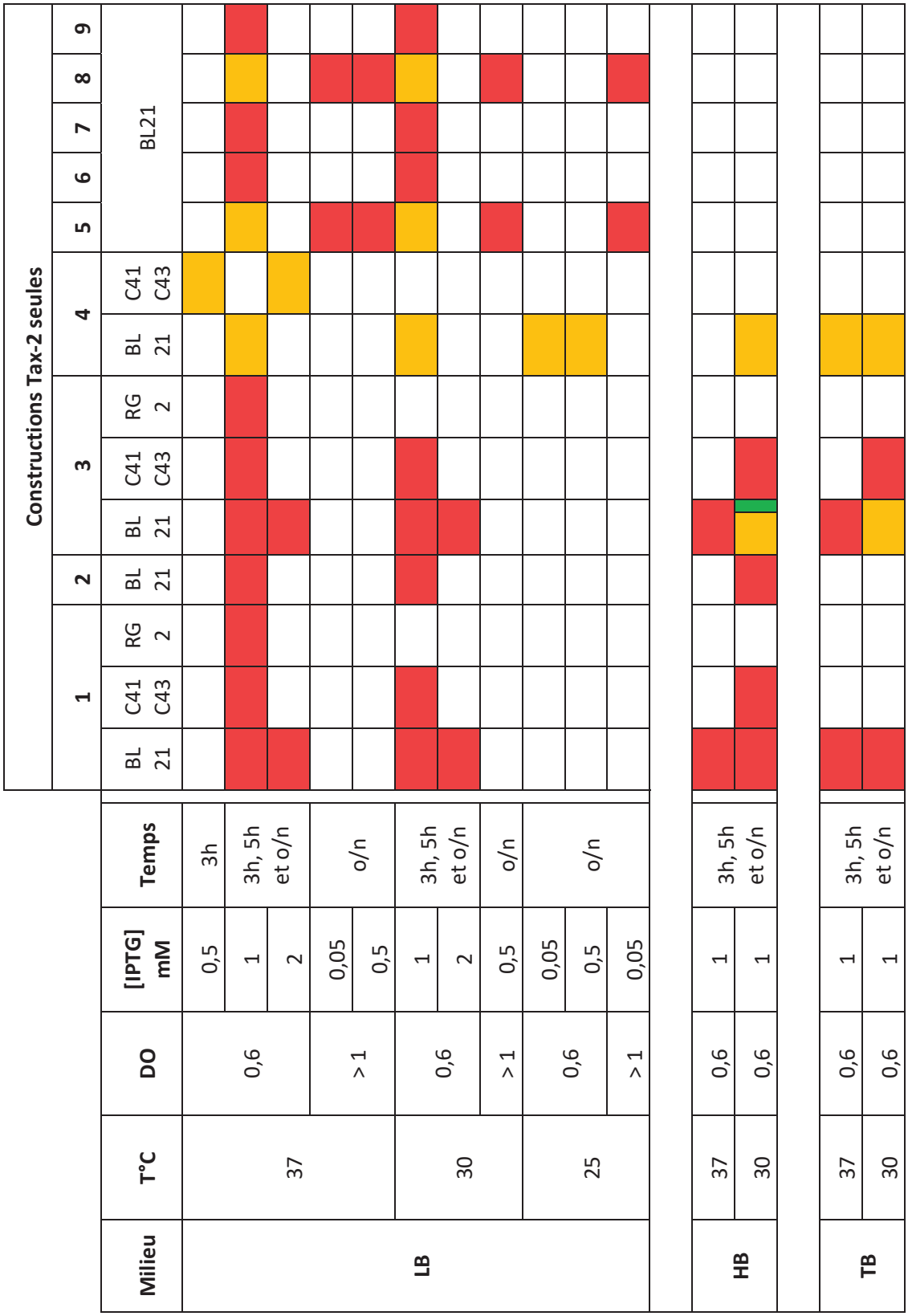

Tableau 13 | Tableau des tests d'expression « classique » des constructions Tax-2 en milieux LB, HB et TB en bactéries *E. coli* BL21 (DE3) pLysS (BL21), C41 et C43 (C41/C43) ou Rosetta-Gami2 (RG2). En rouge : pas d'expression, en jaune : expression insoluble et en vert : expression soluble. o/n : pendant la nuit. Les cases bicolores jaune/vert correspondent à la proportion de protéines insolubles/solubles après expression.

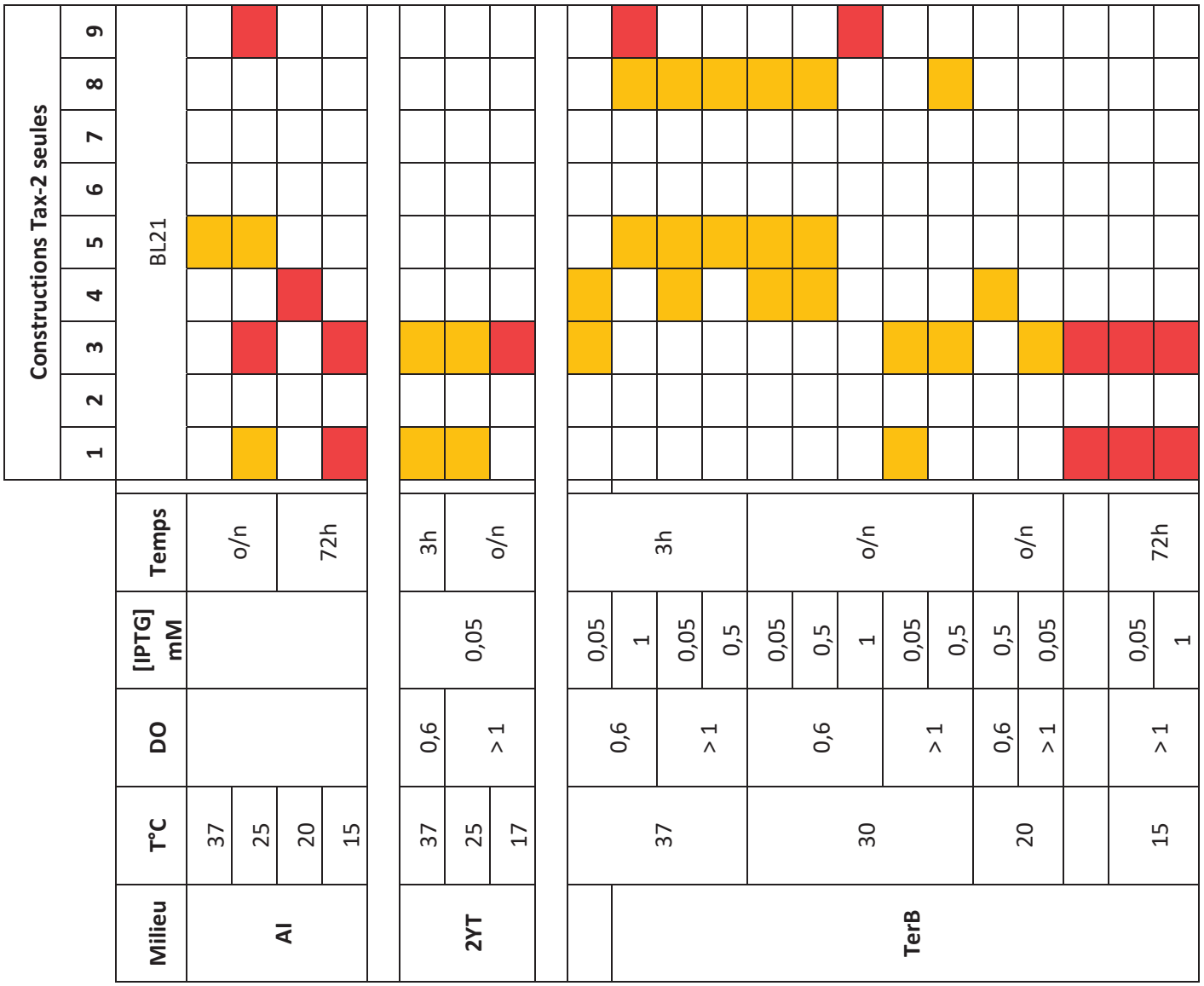

Tableau 14 | Tableau des tests d'expression « classique » des constructions Tax-2 en milieux AI, 2YT et TerB en *E. coli* BL21 (DE3) pLysS. En rouge : pas d'expression et en jaune : expression insoluble. o/n : pendant la nuit.

1.2.1.1. Tax-2 en fusion avec une étiquette His_6

- **Tax2-(20-200)**

Pour les deux fragments Tax2-(20-200) (constructions 1 et 2, **Tableau 13**), aucune expression n'a été détectée quelle que soit la position (N- ou C-terminale) de l'étiquette His_6, dans les milieux « classiques » LB, HB et TB. L'expression de Tax2-(20-200)-6H-C en milieu AI, 2YT et TerB (construction 1, **Tableau 14**) a abouti à la production de la protéine d'intérêt, cependant celle-ci est retrouvée dans la fraction insoluble lors de la lyse des bactéries.

- **Tax2-(20-283)**

Aucune expression n'a été détectée pour le fragment Tax2-(20-283)-6H-C en milieu LB et TB (construction 3, **Tableau 13**), quelles que soient les conditions testées. Son expression en milieu HB à 30°C avec 1 mM d'IPTG pendant la nuit, elle, s'est faite majoritairement sous forme insoluble (**Tableau 13**). Comme précédemment, l'expression en milieu AI, 2YT et TerB (construction 3, **Tableau 14**) n'a abouti qu'à la production d'une protéine sous forme insoluble.

Le changement de la position de l'étiquette His_6 en amino-terminal (construction 4, **Tableau 13** et **Tableau 14**) a permis d'exprimer la protéine Tax2-(20-283)-6H-N, mais celle-ci reste également dans la fraction insoluble, quel que soit le milieu utilisé.

- **Tax2-(1-200) et Tax2-(1-283)**

Comme pour les fragments plus courts associés, l'expression des fragments Tax2-(1-200) et Tax2-(1-283) en milieu LB (constructions 6 et 7, **Tableau 13**) n'a pas abouti à la production de protéines, quelle que soit la condition testée.

- **Tax2-(FL)-opti**

L'optimisation de la séquence de Tax-2 (construction 9) n'a pas permis l'expression de la protéine d'intérêt sous forme entière quelles que soient les conditions testées (**Tableau 13**).

1.2.1.2. Tax-2 en fusion avec une étiquette de solubilité

- **En fusion avec la GST**

L'ajout d'une étiquette de solubilité GST en amino-terminal de la construction Tax2-(20-283), en plus de son étiquette His_6, (construction 5) n'a pas permis son expression sous forme soluble quelles que soient les conditions testées (**Tableau 13** et **Tableau 14**).

- **En fusion avec SUMO**

Comme dans le cas de la GST, la construction Tax2-(1-283) fusionnée à une étiquette de solubilité SUMO en C-terminal (construction 8, **Tableau 13** et **Tableau 14**) est exprimée sous forme insoluble pour toutes les conditions testées.

1.2.2. Optimisation des tests d'expression

Les protocoles d'expression présentés précédemment ne permettant pas l'expression des protéines Tax-2 sous forme soluble, d'autres protocoles d'expression faisant intervenir des additifs ou des antibiotiques ont été testés. Le zinc, par exemple, peut être ajouté au milieu de culture dans le but d'améliorer la stabilité des protéines possédant un domaine à doigt de zinc. D'autre part, des antibiotiques tels que la tétracycline et la rifampicine peuvent être utilisés afin de contrôler les étapes de transcription et de traduction des protéines d'intérêt, respectivement. Les protéines dont l'expression avec les protocoles précédents a abouti à une forme insoluble, ont été choisies pour ces différents tests.

Le tableau suivant (**Tableau 15**) résume les différentes conditions d'expression optimisées testées pour les protéines Tax-2 seules et leurs résultats.

					Constructions								
					1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Test	**Milieu**	**DO**	**[IPTG]**	**T°C**	BL21								
Ajout de zinc	LB	0,6	0,5	30									
	TerB	0,6	0,5	37									
				30									
Ajout de tétracycline	TB	0,6	0,5	30									
	AI			25									
Ajout de rifampicine	LB + 0,4 % glucose	0,6	1	37									

Tableau 15 | Tableau des tests d'expression « adaptée » des constructions Tax-2 en *E. coli* BL21 (DE3) pLysS.
En rouge : pas d'expression, en jaune : expression insoluble.

De manière générale, pour l'expression des protéines Tax-2 seules, l'utilisation de zinc, de tétracycline ou de rifampicine dans les milieux de culture n'a pas permis d'améliorer leur solubilité. Des résultats plus détaillés pour l'expression de certaines constructions sont présentés dans les chapitres suivants.

1.2.2.1. Expression en présence de $ZnSO_4$

La protéine Tax possédant un domaine en doigt de zinc à son extrémité amino-terminale, un test de production en présence de $ZnSO_4$ a été réalisé en vue de stabiliser ce domaine.

Les expériences présentées ci-dessous ont été réalisées dans le milieu TerB à 30°C avec une induction à 0,5 mM d'IPTG et l'ajout de 50 µM de zinc lors de l'induction pour les constructions Tax2-(20-283)-6H-GST-N (5) et Tax2-(1-283)-6H-N-Sumo-C (8) (**Figure 31**).

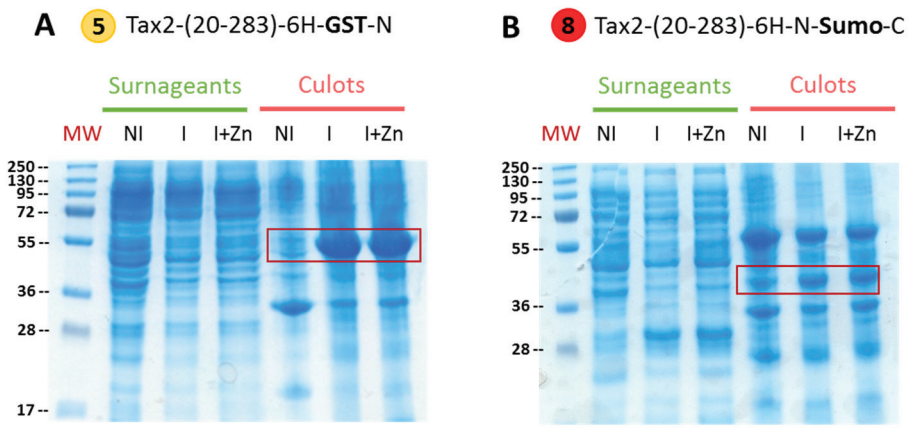

Figure 31 | Analyse par SDS-PAGE de la solubilité des fragments protéiques A) Tax2-(20-283)-6H-GST-N et B) Tax2-(1-283)-6H-N-Sumo-C lors d'une expression en milieu TerB à 30°C en présence 0,5 mM IPTG et ajout de 50 µM de $ZnSO_4$. Pistes MW : marqueurs de masses moléculaires, NI : fraction non induite, I : fraction induite sans $ZnSO_4$, I+Zn : fraction induite en présence de $ZnSO_4$.

Cependant, l'ajout de zinc dans le milieu de culture n'a pas permis d'exprimer les protéines d'intérêt sous forme soluble dans le cas des constructions Tax2-(20-283)-6H-GST-N et Tax2-(1-283)-6H-N-Sumo-C, quelles que soient la condition et la construction testées (**Figure 31, pistes culots**).

1.2.2.2. Expression avec ajout de tétracycline : arrêt de la traduction

La tétracycline est un antibiotique permettant de bloquer l'étape de synthèse protéique en empêchant la fixation de l'aminoacyl-ARNt dans le site A du ribosome. En ajoutant cet antibiotique dans le milieu de culture, après 1h d'induction, la synthèse de la protéine d'intérêt est bloquée, permettant ainsi de favoriser un repliement correct des protéines produites sous forme soluble par l'action des protéines chaperonnes.

L'expérience a été réalisée dans le milieu TB à 30°C avec une induction à 0,5 mM d'IPTG, puis l'ajout de 200 µg.mL^{-1} de tétracycline 1h après l'induction pour les constructions Tax2-(20-283)-6H-C (3) et Tax2-(20-283)-6H-N (**Figure 32**), exprimées sous forme insoluble dans cette condition sans tétracycline.

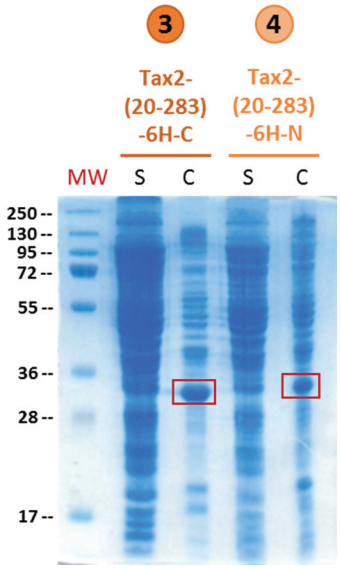

Figure 32 | Analyse par SDS-PAGE de la solubilité des protéines Tax2-(20-283)-6H-C et Tax2-(20-283)-6H-N lors d'une expression en milieu TB à 30°C en présence de 0,5 mM IPTG et ajout de 200 µg.mL^{-1} de tétracycline après 1h d'induction. Pistes MW : marqueurs de masses moléculaires, S : surnageant et C : culot.

L'ajout de tétracycline après l'induction n'a pas permis d'exprimer les protéines d'intérêt sous forme soluble : en effet, celles-ci restent dans la fraction insoluble (**Figure 32, pistes culots**) comme lorsqu'elles sont exprimées dans la même condition sans tétracycline.

1.2.2.3. Expression avec ajout de rifampicine : arrêt de la transcription

La rifampicine est un antibiotique permettant de stopper l'étape de transcription des ARN messagers en se liant spécifiquement à l'ARN polymérase bactérienne. En ajoutant cet antibiotique dans le milieu de culture 1h après l'induction, la transcription est bloquée et seule l'étape de traduction des ARNm déjà produits peut continuer. Ceci permettrait à la machinerie procaryote de n'effectuer qu'une seule tâche à la fois, à savoir la production des protéines d'intérêt sous forme soluble.

Les expériences ont été réalisées dans le milieu LB, additionné de 0,4% de glucose, à 37°C avec une induction à 1 mM d'IPTG, puis l'ajout de 100 µg.mL^{-1} de rifampicine 1h après l'induction pour les constructions 1, 2, 3, 4, 5, 7, 8 et 9. (**Figure 33**).

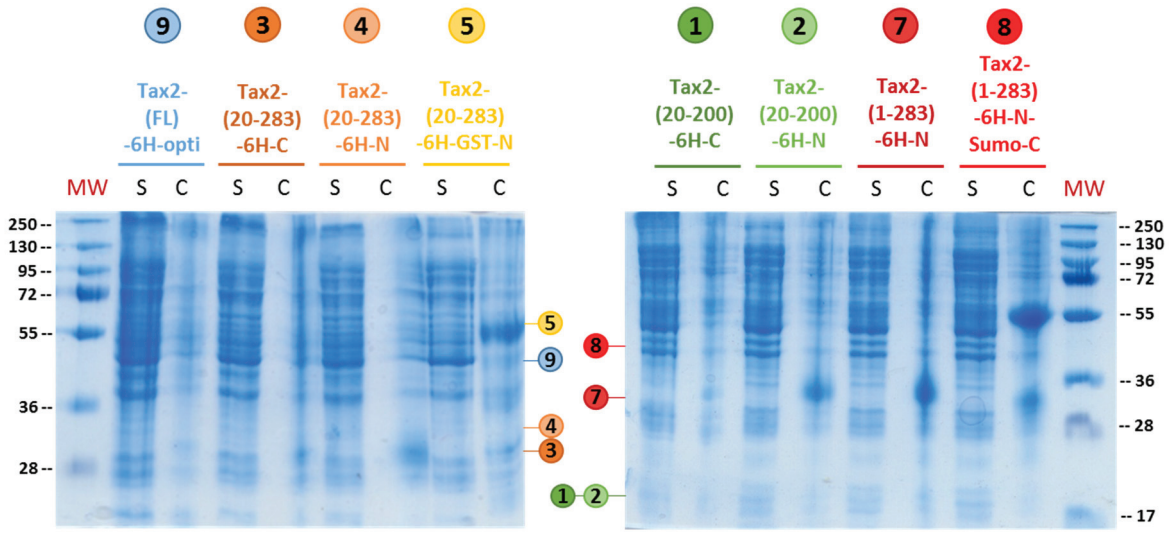

Figure 33 | Analyse par SDS-PAGE de la solubilité de différentes constructions Tax-2 lors d'une expression en milieu LB à 37°C en présence de 1 mM IPTG et ajout de 100 µg.mL^{-1} de rifampicine après 1h d'induction. Pistes MW : marqueurs de masses moléculaires, S : surnageant et C : culot. La taille attendue pour chaque construction est représentée par des billes numérotées et colorées pour chaque gel.

L'utilisation d'un milieu LB supplémenté en glucose avec ajout de rifampicine après induction n'a pas permis d'exprimer les différentes constructions Tax-2, que ce soit sous forme soluble ou insoluble. En effet, aucune bande correspondant à la taille attendue pour ces constructions n'est observée lors de l'analyse SDS-PAGE (**Figure 33**).

1.2.3. Solubilisation des protéines Tax-2

L'expression de gènes recombinants *tax-2* en système procaryote conduisant à la production de protéines insolubles, deux méthodes de solubilisation ont été envisagées.

1.2.3.1. Dénaturation par le chlorure de guanidinium

L'expérience d'extraction des corps d'inclusion de la construction Tax2-(20-283)-6H-C (3) a été réalisée après son expression en milieu HB à 30°C avec une induction à 1 mM d'IPTG pendant la nuit. Après lyse des bactéries et plusieurs étapes de lavage, les corps d'inclusion ont été traités par 6 M de chlorure de guanidinium (**Figure 34**).

La dénaturation par le chlorure de guanidinium n'a pas permis d'extraire la protéine Tax-2-(20-283)-6H-C des corps d'inclusion. En effet, d'après l'analyse SDS-PAGE, celle-ci est toujours retrouvée dans la fraction insoluble (**Figure 34, piste C_{final}**) après traitement par le chlorure de guanidinium.



Figure 34 | Analyse par SDS-PAGE de l'extraction de Tax2-(20-283)-6H-C des corps d'inclusion par dénaturation avec 6 M de chlorure de guanidinium. Pistes MW : marqueurs de masses moléculaires, SA, S_{H2O}, SP, SD : surnageants après tampon A, choc thermique, tampon P, tampon B et tampon D contenant 6 M de chlorure de guanidinium, respectivement et C_{final} : culot final.

1.2.3.2. Dénaturation par l'urée

L'expérience d'extraction des corps d'inclusion de la construction Tax2-(20-283)-6H-N (4) a été réalisée après son expression en milieu TerB à 30°C avec une induction à 0,5 mM d'IPTG pendant la nuit. Après lyse des bactéries, les corps d'inclusion ont été traités par différentes concentrations d'urée (2 à 6 M) à pH 9 (**Figure 35**).

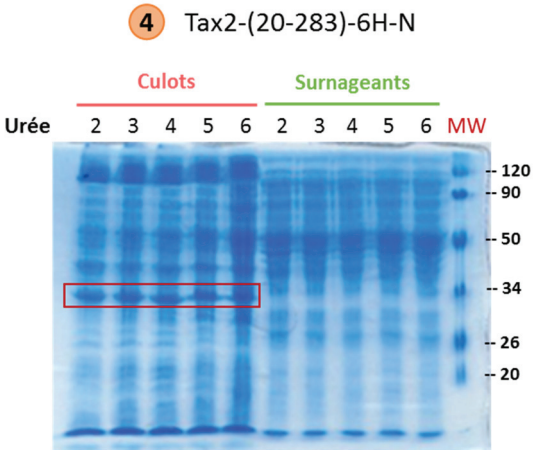

Figure 35 | Analyse par SDS-PAGE de la solubilisation de Tax2-(20-283)-6H-N par dénaturation par l'urée à pH 9, après une expression en milieu TerB à 30°C en présence de 0,5 mM IPTG. Piste MW : marqueurs de masses moléculaires.

Comme le test de solubilisation précédent, la dénaturation par l'urée n'a pas permis d'extraire la protéine Tax-2-(20-283)-6H-N des corps d'inclusion, quelles que soient les concentrations en urée (**Figure 35, pistes culots**) ou les pH (8, 9 ou 10, données non présentées) testés.

1.2.4. Précipitation au sulfate d'ammonium et purification par Ni-NTA de Tax-2

Une minorité de la protéine Tax2-(20-283)-6H-C (3) produite lors d'une expression en milieu HB à 30°C avec une induction à 1 mM d'IPTG durant la nuit semblant être soluble (**Figure 36, piste S**), un test de précipitation au sulfate d'ammonium a été effectué afin de la purifier.

Le pourcentage optimal de sulfate d'ammonium permettant de précipiter la protéine d'intérêt à 4°C a été estimé à 5% (**Figure 36, piste C5**).

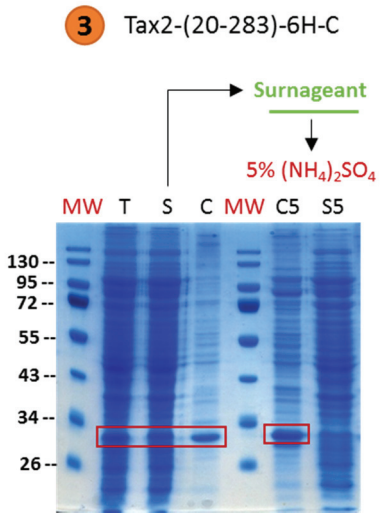

Figure 36 | Analyse par SDS-PAGE de la précipitation au sulfate d'ammonium (5%) de la fraction soluble de Tax2-(20-283)-6H-C. Pistes MW : marqueurs de masses moléculaires, T : fraction totale, S : surnageant et C : culot, C5 et S5 : culot et surnageant après la précipitation au sulfate d'ammonium.

La précipitation de la protéine Tax2-(20-283)-6H-C (3) au sulfate d'ammonium a ensuite été suivie d'une étape de purification par chromatographie d'affinité pour le nickel.

L'expérience de purification a été réalisée par centrifugation en tube Eppendorf avec les tampons de lavage et d'élution W1 et E1 (**Tableau 10, Matériels et méthodes 2.2.2**) contenant 50 mM et 300 mM d'imidazole, respectivement.

D'après les analyses SDS-PAGE et *western blot* effectuées (**Figure 37**), aucune protéine portant une étiquette His_6 n'est éluée à la taille attendue pour la protéine d'intérêt. La protéine Tax2-(20-283)-6H-C n'a donc pas pu être purifiée par chromatographie d'affinité pour le nickel.

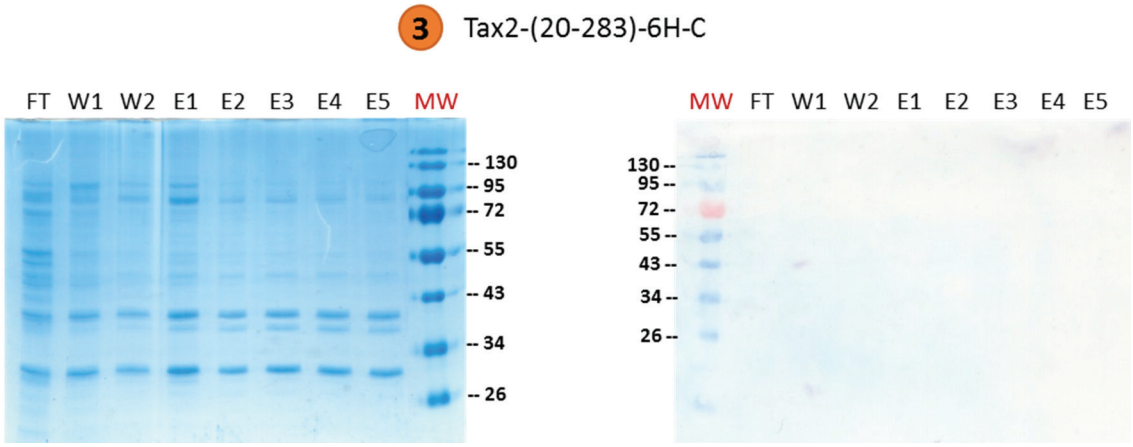

Figure 37 | Analyse par SDS-PAGE et western-blot de la purification par Ni-NTA de la protéine Tax2-(20-283)-6H-C précipitée à 5% d'ammonium sulfate. Pistes MW : marqueurs de masses moléculaires, FT : F*low-Through*, W : fractions de lavage et E : fractions d'élution.

1.3. Co-expression de la protéine Tax-2 entière avec CREB

De même que pour les protéines Tax-2 seules, aucune condition d'expression en milieu LB ou AI n'a permis l'expression de la protéine Tax-2 entière sous forme soluble en présence de son partenaire CREB, tout comme l'expression avec ajout de $ZnSO_4$ (**Tableau 16**).

La co-expression de ces deux protéines en présence de rifampicine, elle, a permis l'obtention de la protéine Tax-2 entière sous forme soluble (**Tableau 16**). Cette condition d'expression et la purification de la protéine Tax-2 entière sont détaillées dans les paragraphes suivants.

					Co-expression Tax-2 et CREB
	Milieu	**DO**	**[IPTG]**	**T°C**	BL21
	LB	0,6	1	37 / 30	
	AI			37 / 25	
Ajout de $ZnSO_4$	TerB	0,6	0,5	37 / 30	
Ajout de rifampicine	LB + 0,4 % glucose	0,6	1	37	

Tableau 16 | Tableau des tests de co-expression des protéines Tax-2 et CREB en *E. coli* BL21 (DE3) pLysS. En rouge : pas d'expression, en jaune : expression insoluble et en vert : expression soluble de la protéine Tax-2 entière.

1.3.1. Expression avec ajout de rifampicine

Un test de co-expression de Tax-2 et CREB avec ajout de rifampicine a été realisé dans le milieu LB additionné de 0,4% de glucose, à 37°C avec une induction à 1 mM d'IPTG et l'ajout de 100 µg.mL^{-1} de rifampicine 1h après l'induction (**Figure 38**).

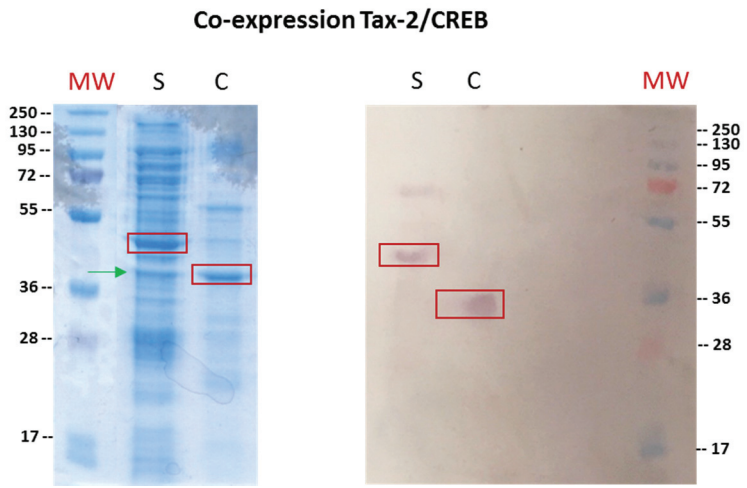

Figure 38 | Analyse par SDS-PAGE et *western blot* de la co-expression des protéines Tax-2 entière et CREB en milieu LB à 37°C en présence de 1 mM IPTG avec ajout de 100 µg.mL^{-1} de rifampicine 1h après l'induction. Pistes MW : marqueurs de masses moléculaires, S : surnageant et C : culot.

D'après les analyses SDS-PAGE et *western blot* anti-histidines réalisées, il semblerait que l'utilisation d'un milieu LB supplémenté en glucose avec ajout de rifampicine après induction ait permis l'expression de la protéine Tax-2 entière avec son étiquette His_6 partiellement sous forme soluble. En effet, une bande à environ 40 kDa est révélée (**Figure 38, pistes S, carré rouge**) sachant que la taille attendue pour Tax-2 est de 39,3 kDa.

Cependant, une bande à ~36 kDa est également révélée dans la fraction insoluble sur le *western blot* anti-histidines (**Figure 38, piste S**), suggérant qu'une forme incomplète de la protéine Tax-2 s'accumule sous forme de corps d'inclusion lors de son expression. Cette forme incomplète pourrait être une conséquence de la présence de codons rares dans la protéine Tax-2 qui troublerait la traduction protéique.

Concernant la protéine CREB, une bande à environ 38 kDa apparaît sur l'analyse SDS-PAGE (**Figure 38, SDS-PAGE piste S, flèche verte**) sachant que sa masse moléculaire théorique est de 37,9 kDa : un *western blot* anti-Flag devra être effectué afin de confirmer qu'il s'agit bien de la protéine CREB.

1.3.2. Purification par Ni-NTA

La protéine Tax-2 entière soluble issue de la co-expression avec son partenaire cellulaire CREB a ensuite été purifiée par chromatographie d'affinité pour le nickel.

L'expérience de purification a été réalisée par centrifugation avec les tampons de lavage et d'élution W2 et E2 (**Tableau 10, Matériels et méthodes 2.2.2**).

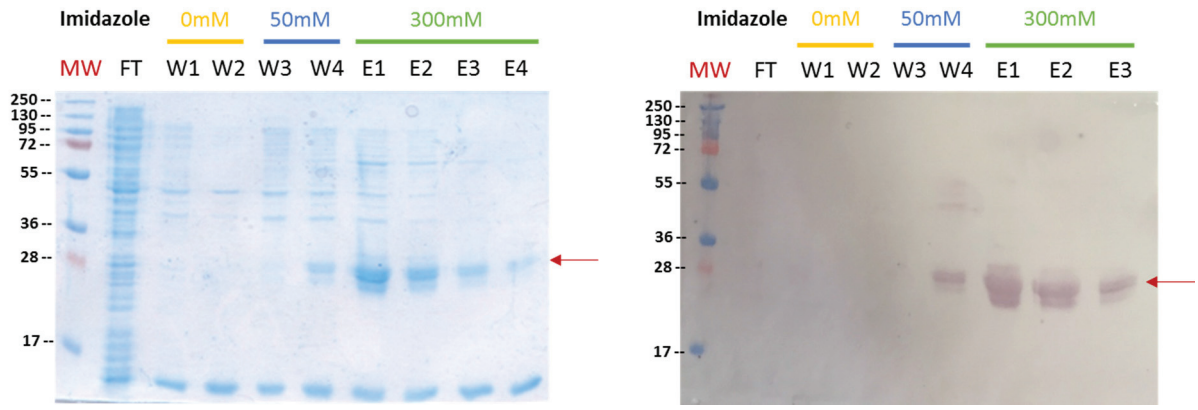

Figure 39 | Analyse par SDS-PAGE et western-blot de la purification par Ni-NTA de Tax2-(FL)-CREB, après une expression en milieu LB à 37°C en présence de 1 mM IPTG avec ajout de 100 µg.mL^{-1} de rifampicine 1h après l'induction. Pistes MW : marqueurs de masses moléculaires, W : fractions de lavage et E : fractions d'élution.

D'après les analyses SDS-PAGE et *western blot* effectuées, une protéine portant une étiquette His_6 apparait à ~ 28 kDa (**Figure 39, pistes E1, E2 et E3**). Il semblerait donc que la protéine Tax-2 soit dégradée lors de l'étape de purification. De plus, la protéine d'intérêt est éluée dès le passage du tampon de lavage contenant 50 mM d'imidazole (**Figure 39, piste W4**), suggérant que cette concentration en imidazole est suffisante pour son élution. Cependant, de nombreux contaminants sont également élués en même temps que la protéine Tax-2 (**Figure 39, pistes W4, E1 et E2**).

L'expression de Tax2-(FL)-CREB étant concluante en petit volume (40 mL), une expression dans un plus grand volume de LB a été réalisée pour cette construction (1 L) afin de purifier une quantité notable de protéines Tax-2 (en complexe avec CREB). Un cocktail d'anti-protéases différent a également été ajouté au tampon de lyse dans le but de s'affranchir du phénomène de dégradation de la protéine Tax-2 entière observé. Cependant, aucune protéine d'intérêt n'a été produite lors de son expression en grand volume (données non présentées).

1.4. Discussion

L'expression en système bactérien (*E. coli*) est la stratégie la plus couramment utilisée pour l'expression de protéines hétérologues, en raison de sa facilité de mise en œuvre et de ses rendements de production élevés. Cependant, malgré ces avantages, l'expression dans ce système des protéines recombinantes Tax-2 d'HTLV-2 seules n'a pas permis d'aboutir à une production efficace des protéines seules. En effet, quelles que soient les constructions plasmidiques ou les conditions d'expression et de lyse expérimentées, les protéines Tax-2 d'intérêt ne sont soit pas exprimées, soit exprimées sous forme insoluble dans des corps d'inclusion (**Tableau 13, Tableau 14 et Tableau 15**). Afin de résoudre ce problème de solubilité des protéines exprimées, des méthodes de solubilisation faisant intervenir des agents dénaturants, comme le chlorure de guanidinium et l'urée, ont été testées. Cependant, celles-ci n'ont pas permis d'obtenir les protéines d'intérêt sous forme soluble.

En parallèle, la séquence de la protéine Tax-2 a été optimisée pour l'expression en système bactérien en remplaçant les codons rares de la séquence par des codons utilisés plus fréquemment chez *E. coli*. Cependant, cette optimisation n'a pas permis d'améliorer la production et la solubilité de la protéine Tax-2 entière. Il semblerait donc que les problèmes de production et de solubilisation de la protéine Tax-2 ne soient pas dus aux codons rares.

Dans le cadre des protéines Tax-2 seules, il pourrait être intéressant de tester d'autres systèmes d'expression tels que la levure ou le système acellulaire, qui possèdent une machinerie d'expression eucaryote. Ceci permettrait potentiellement d'améliorer l'expression sous forme soluble des protéines d'intérêt. Cependant, l'expression en système eucaryote permet la prise en charge des nombreuses modifications post-traductionnelles que subissent Tax et dont les machineries sont absentes des systèmes bactériens (sumoylation, etc.). Or, ces longues chaînes supplémentaires peuvent constituer un écueil pour la cristallographie aux rayons X en raison de leur flexibilité. De plus, les protéines Tax peuvent subir différentes modifications de façon simultanée ou non et celles-ci ne se font pas forcément sur les mêmes résidus d'acides aminés (Turci *et al.*, 2012). Ces différences de modifications post-traductionnelles de la protéine Tax au sein d'un même échantillon entraînerait un problème de polydispersité des échantillons pour les étapes de caractérisations biochimiques et biophysiques ainsi que pour la cristallogenèse. C'est pourquoi, ce type d'expression n'a pas été favorisé/expérimenté lors de ce travail doctoral.

Néanmoins, la construction Tax2-(FL)-CREB, permettant l'expression de la protéine Tax-2 sous forme entière avec son partenaire cellulaire CREB, reste prometteuse. En effet, son expression, en système procaryote, sous forme soluble a été détectée dans le milieu LB supplémenté en glucose lors de l'ajout de rifampicine dans le milieu de culture après induction (**Figure 38**). Sa purification par chromatographie d'affinité pour le nickel permet de récupérer la protéine d'intérêt, bien que des contaminants soient également co-élués (**Figure 39**). Des étapes de purification supplémentaires, par chromatographie d'exclusion stérique notamment, pourraient permettre de récupérer la protéine Tax-2 pure. Ces contaminants pourraient également être éliminés lors d'une seconde étape de purification par chromatographie d'affinité, dirigée contre l'étiquette Flag de CREB, qui permettrait de récupérer un complexe Tax-2:CREB pur (dans le cas où celui-ci se formerait).

Cependant, la protéine Tax-2 produite lors de la co-expression semble être dégradée lors de l'étape de purification où une bande à environ 28 kDa apparaît (**Figure 39**) alors que la taille attendue pour notre protéine d'intérêt est de 39,3 kDa. Afin d'éviter cette dégradation protéique, il serait utile de tester de nouveaux cocktails d'anti-protéases plus complets que celui actuellement utilisé. Celui envisagé pour la suite est le « cOmplete ULTRA EDTA-free » (Roche) qui, en plus de la formulation classique, contient des anti-protéases aspartiques. L'utilisation de ce dernier au sein du laboratoire a déjà permis de s'affranchir de ce problème de dégradation.

Enfin, il semblerait que la protéine Tax-2 ait du mal à s'exprimer dans un grand volume de milieu de culture de LB ; et ce dans les mêmes conditions que lors des tests (seule la préculture utilisée différait). Or, la production de protéine Tax-2 en grande quantité est essentielle à la mise en place des techniques de caractérisations biophysiques et biochimiques de la protéine ainsi que pour la cristallographie aux rayons X. En effet, cette dernière demande une concentration en protéine souvent élevée et un volume notable de protéines en vue du criblage de nombreuses conditions de cristallisation. De nouveaux tests d'expression en petit et grand volumes (à partir d'une même préculture) sont actuellement en cours afin de vérifier la co-expression de la protéine Tax-2 avec son partenaire CREB dans différents volumes de LB (50 mL, 500 mL et 1 L). Dans le cas où celle-ci ne s'exprimerait toujours pas en grand volume, il sera envisagé d'effectuer de multiples expressions en petit volume jusqu'à obtenir la quantité de protéine nécessaire pour la réalisation des différentes expériences mentionnées ci-dessus.

2. Études fonctionnelles et structurales de la protéine de capside du virus FIV

La seconde partie de ce projet doctoral concerne la caractérisation fonctionnelle et structurale de la protéine de capside du FIV, essentielle aux étapes d'assemblage et de maturation de ce virus. La résolution de sa structure permettrait de mieux comprendre les mécanismes moléculaires mis en jeu lors de l'assemblage du FIV et ses spécificités par rapport aux autres rétrovirus.

Les études structurales antérieures effectuées au laboratoire sur la protéine de capside sauvage n'ayant pas été concluantes, malgré le test de plusieurs milliers de conditions de cristallisation, deux modifications ont été apportées à celle-ci dans le but d'améliorer sa cristallisation. Tout d'abord, le résidu N-terminal (Pro1) a été muté en thréonine (Thr1) afin de permettre le clivage de l'étiquette His_6 par le Facteur X, laissant ainsi l'extrémité N-terminale de la protéine intacte (tous les résidus du site de clivage étant éliminés). Ensuite, les 11 résidus d'acides aminés de l'extrémité C-terminale, prédits comme flexibles par le serveur Ronn (https://www.strubi.ox.ac.uk/RONN), ont été supprimés afin de favoriser la croissance des cristaux. Préalablement aux études structurales, des études fonctionnelles ont été conduites afin de confirmer que cette construction de la CA du FIV, CAΔC-PT, était fonctionnelle pour l'assemblage.

2.1. Surproduction, purification et clivage de l'étiquette d'affinité

2.1.1. Purification de la CA

Après une expression en milieu LB à 25°C pendant la nuit en présence de 1 mM d'IPTG, les bactéries sont lysées dans le tampon LEW (**Tableau 11**, **Matériels et méthodes 2.3.2.2**) et la protéine de capside est purifiée par chromatographie d'affinité pour le nickel (**Figure 40**).

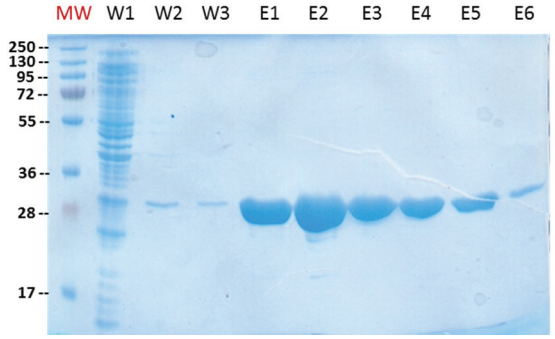

Figure 40 | Analyse par SDS-PAGE de la purification sur résine Ni-TED de la CAΔC-PT du FIV.
Pistes MW : marqueurs de masses moléculaires, W : fractions de lavage et E : fractions d'élution.

D'après l'analyse SDS-PAGE, les fractions d'élutions ne contiennent qu'une seule bande à la taille attendue pour la protéine d'intérêt (~ 28 kDa) (**Figure 40, pistes E1 à E6**) : une seule étape de purification suffit donc à obtenir la CAΔC-PT pure, avec un rendement d'environ 15 mg de protéines par litre de culture.

2.1.2. Clivage de l'étiquette d'affinité

Après purification, l'étiquette His_6 de la CA est clivée par la protéase Facteur Xa pendant la nuit à 19°C (**Figure 41**). Sur l'analyse SDS-PAGE, une diminution de la taille de la protéine d'intérêt est observée après clivage (**Figure 41, piste - His_6**), signifiant que l'étiquette His_6 de la CA a bien été enlevée.

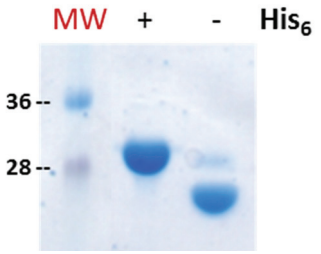

Figure 41 | Analyse par SDS-PAGE du clivage de l'étiquette His_6 de la CAΔC-PT du FIV par le Facteur X, après purification par chromatographie d'affinité. Pistes MW : marqueurs de masses moléculaires, + et - : avant et après clivage, respectivement.

Après clivage de son étiquette His_6, la CAΔC-PT est de nouveau purifiée par chromatographie d'affinité afin d'éliminer l'étiquette His_6 ainsi que la fraction résiduelle de protéines n'ayant pas été clivées. La CAΔC-PT pure, dépourvue de son étiquette, est ensuite concentrée et stockée dans le tampon de stockage adéquat (**Tableau 11, Matériels et méthodes 2.3.2.2**), ceci permettant d'effectuer les différentes études fonctionnelles et structurales décrites ci-après.

2.2. Etudes fonctionnelles de la CA du FIV

La CA étudiée étant mutée sur son premier résidu (P1T) et clivée de 11 résidus à son extrémité C-terminale, ses capacités à s'oligomériser en solution et à s'assembler correctement *in vitro* ont été évaluées par DLS. En effet, les expériences effectuées au laboratoire sur la CA native du FIV ont démontré que celle-ci était capable de s'oligomériser *in vitro* afin de former une capside de forme sphérique (Serriere *et al.*, 2013).

2.2.1. Etat d'oligomérisation de la CA en solution

Les expériences d'études structurales nécessitant des échantillons monodisperses, l'état d'oligomérisation de la CAΔC-PT du FIV (sans étiquette His_6) en solution dans le tampon HEPES a été évalué par DLS (**Figure 42**).

À 7 $mg.mL^{-1}$, un seul pic est observé pour la CAΔC-PT, confirmant que la protéine est homogène en solution avec un indice de polydispersité de 0,3. Celle-ci n'a donc pas subi de phénomène d'agrégation lors de l'étape de concentration. Le diamètre hydrodynamique moyen observé, à partir de trois mesures successives, est de 6,8 nm avec un écart-type de ± 0,4 nm.

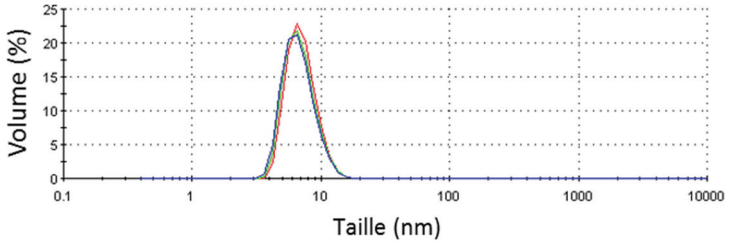

Figure 42 | Analyse DLS de la monodispersité de la CAΔC-PT à 7 $mg.mL^{-1}$ en tampon HEPES (50 mM HEPES, 100 NaCl, pH 6,5). Les courbes rouge, bleue et verte représentent les trois mesures successives effectuées.

Par comparaison avec les tailles des protéines globulaires de référence, ce diamètre correspond à une protéine d'environ 60 kDa, soit à une forme dimérique de la CAΔC-PT. En effet, la masse moléculaire de celle-ci est d'environ 28 kDa pour la forme monomérique.

De plus, ce résultat est en accord avec ceux obtenus pour la CA native du FIV en solution pour laquelle un diamètre de 6,8 nm (± 0,3 nm) a été observé à cette même concentration.

2.2.2. Suivi de l'assemblage *in vitro* par DLS

La capacité de la CA étudiée à s'assembler *in vitro* a également été évaluée par DLS. Son assemblage a été initié par une augmentation de la force ionique de la solution protéique et a été suivi pendant 80 minutes à 37°C (**Figure 43**).

Avant l'ajout de sels (**Figure 43, T=0**), le diamètre moyen des particules en solution correspond à celui observé précédemment pour la particule sous forme dimérique (**Résultats et discussions 2.2.1**).

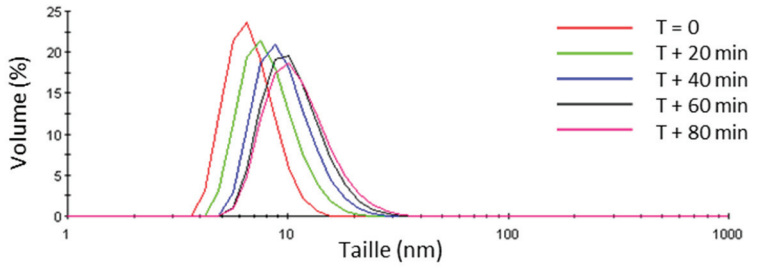

Figure 43 | Suivi de l'assemblage capsidique *in vitro* par DLS de la CA concentrée à 7 mg.mL^{-1} en tampon HEPES lors de l'ajout de 1 M final de NaCl.

Après l'ajout de 1 M final de NaCl à la solution protéique, le diamètre moyen des particules augmente jusqu'à atteindre environ 22,9 nm (**Figure 43, T=20, T=40, T=60 et T=80**). Cette augmentation du diamètre hydrodynamique traduit un changement d'état oligomérique de la protéine en solution, passant ainsi d'une forme dimérique à une forme d'oligomérisation supérieure.

Par conséquent, les deux modifications apportées à la CA du FIV n'ont pas altéré sa capacité à s'oligomériser en solution : la CAΔC-PT représente donc une forme fonctionnelle de la CA du FIV.

2.3. Etudes structurales de la CA du FIV

2.3.1. Etude structurale par cristallographie aux rayons X

2.3.1.1. Cristallogenèse

Plus de 3 000 conditions de cristallisation distinctes, utilisant de nombreux kits commerciaux et différentes concentrations de protéine, ont été testées à 19°C (**Tableau 17**) pour la CAΔC-PT du FIV en présence ou en absence de son étiquette His_6. Ces tests ont été menés sur la CAΔC-PT stockée dans le tampon HEPES (50 mM HEPES, 100 mM NaCl, pH 6,3).

Au bout de deux à trois semaines, des cristaux ont commencé à apparaitre dans certaines conditions de cristallisation (kit PEG II, PEG Ion et Crystal Screen) uniquement pour la protéine clivée de son étiquette His_6 (**Tableau 17**). Le pouvoir de diffraction de ces cristaux a donc été testé aux rayons X à la source synchrotron de l'ESRF de Grenoble. Cependant, la plupart d'entre eux se sont avérés être des cristaux de sels.

				Kits							
				PEG I	PEG II	PEG Ion	Crystal Screen	$AmSO_4$	JSCG+	5k	Morpheus
CA du FIV	+ His_6	Concentration ($mg.mL^{-1}$)	3								
			7								
			8								
			12								
			23								
	- His_6		3								
			6								
			7								
			9								
			14								

Tableau 17 | Tableaux des différentes conditions de cristallogenèse testées à 19°C pour la CA du FIV en présence ou en absence de son étiquette His_6. Les différents kits de criblage testés sont colorées en rouge: pas d'apparition de cristaux, en jaune : cristaux de sels et en vert : cristaux de protéine. Les fournisseurs des différents kits utilisés sont les suivants : PEG I, PEG II, $AmSO_4$ et JSCG+ : Qiagen, PEG Ion et Crystal Screen : Hampton Research, Morpheus : Molecular Dimensions et 5k : plateforme PSF.

Néanmoins, une condition de cristallisation, correspondant à la condition E11 du kit de cristallisation PEG II (contenant 0,2 M de magnésium sulfate, 20% (m/v) de PEG 4000 et 10 % (v/v) de glycérol), a permis d'obtenir un cristal d'une longueur d'environ 100 µm (**Figure 44**). Ce cristal a diffracté les rayons X jusqu'à une résolution de 10 Å avec un profil de diffraction correspondant à celui d'une protéine. L'échantillon utilisé étant pur (**Figure 40**), il s'agit donc bien d'un cristal de CAΔC-PT. Cependant, la résolution à laquelle le cristal a diffracté étant trop faible, il n'a pas été possible de résoudre la structure de la protéine.

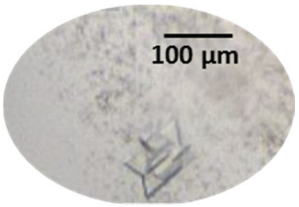

Figure 44 | Premiers cristaux de CA obtenus sans étiquette His_6 à 7 $mg.mL^{-1}$ en tampon HEPES. Cristaux obtenus dans la condition E11 du kit PEG II contenant 0,2 M de magnésium sulfate, 20% (m/v) de PEG 4000 et 10 % (v/v) de glycérol.

Une optimisation de la condition de cristallisation identifiée pour la CAΔC-PT a ensuite été réalisée, en plaques 24 puits, en faisant varier les proportions protéine:agent précipitant au sein de la goutte (1:1, 1:2 ou 2:1 (v:v)). Cependant, aucune de ces conditions n'a permis de reproduire le cristal précédemment obtenu.

Enfin, des essais de co-cristallisation de la CAΔC-PT du FIV à 7 $mg.mL^{-1}$ ont été réalisées avec plusieurs composés potentiellement inhibiteurs (fournis par le Dr. G. Alvarez, Paysandú), solubilisés dans du DMSO, dans la condition de cristallisation précédemment identifiée (E11 PEG II). Cette co-cristallisation a permis d'obtenir de nombreux cristaux (**Figure 45**) dont le pouvoir de diffraction a été testé aux rayons X sur la ligne de lumière ID30-B de l'ESRF. Tous les cristaux obtenus ont diffractés les rayons X selon un profil caractéristique d'un cristal de macromolécule, confirmant qu'il s'agit bien de cristaux de la protéine de capside.

Après indexation des différents jeux obtenus, tous les cristaux présentent le même groupe d'espace monoclinique C121 avec les paramètres de maille suivants : a = 122 Å, b = 74 Å, c = 77 Å et $\alpha = \gamma = 90°$, $\beta = 128°$.

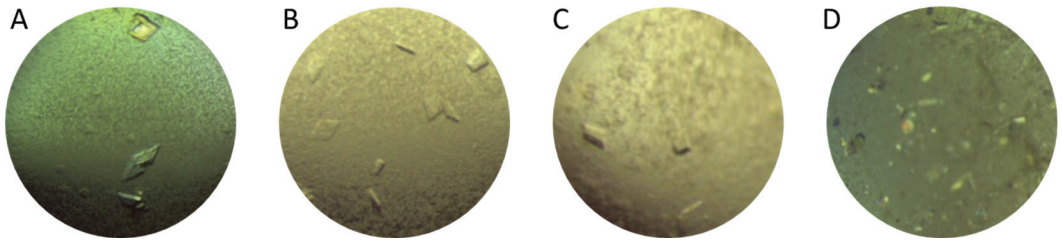

Figure 45 | Cristaux de la CA du FIV à 7 $mg.mL^{-1}$ en tampon HEPES obtenus lors de la co-cristallisation avec les composés (A) 675, (B) 121, (c) pg146 et (D) 685. Les composés ont été solubilisés dans du DMSO.

Les cristaux obtenus lors de la co-cristallisation avec le composé 675 (**Figure 45A**) ayant le meilleur pouvoir de diffraction (1,9 Å), les jeux de données de diffraction de ce cristal ont été utilisés pour résoudre la structure de la CAΔC-PT. Les statistiques d'enregistrement des données de diffraction et d'affinement de la structure de la CA sont résumées dans le **Tableau 18**.

Enregistrement de données	**FIV CA**
Groupe d'espace	C121
Paramètres de maille *a, b, c* (Å)	122, 74, 77
α, β, γ (°)	90, 128, 90
Echelle de résolution (Å)	59,7 – 1,90 (2,01 – 1,90)
R_{sym} (%)	3,54 (25,8)
I/σ(I)	16,42 (3,33)
Complétude (%)	91,2 (93,0)
Redondance	2,38
Affinement	
Echelle de résolution (Å)	59,7 – 1,9
Nombre de réflexions uniques	39142
R_{work} (%)/R_{free} (%)	19/22,5
Nombre d'atomes dans la protéine	3330
Nombre d'atomes (eau/glycérol)	697/6
Facteur B moyen (Å^2)	22,9
Déviation de coordonnées	
RMSD (*Root-Mean-Square Deviation*) des longueurs de liaisons (Å)	0,007
RMSD des angles (°)	0,857
PDB ID	***5NA2***

Tableau 18 | Statistiques d'enregistrement des données de diffraction et d'affinement pour la CA du FIV. Les valeurs entre parenthèses représentent celles attribuées à la coquille de résolution la plus haute.

2.3.1.2. Résolution de la première structure de la CA du FIV

La structure de la CAΔC-PT du FIV a été résolue à 1,9 Å par la méthode du remplacement moléculaire en utilisant les structures du CA_{NTD} de Relik (PDB ID : 2XGU) (Goldstone *et al.*, 2010) et du CA_{CTD} de la CA du FIV (PDB ID : 5DCK) (Khwaja *et al.*, 2016) comme modèles-guides. La structure de la CAΔC-PT a été résolue dans le groupe d'espace monoclinique C121, avec deux monomères de la protéine dans l'unité asymétrique (UA). Les deux chaînes (A et B) présentent une structure équivalente avec un RMSD de superposition de 0.63 Å pour les paires de Cα.

- Structure cristalline de la CA du FIV

La protéine de capside du FIV est repliée en deux domaines composés d'hélices α et séparés par une boucle flexible (**Figure 46**). Son domaine N-terminal (CA_{NTD}) est composé d'un motif *β-hairpin* de 12 résidus d'acides aminés, suivi de 7 hélices α (numérotées α1 à α7) et contient une boucle de liaison à la cyclophiline A (CypA-BL) située dans la boucle L5 (entre les hélices α4 et α5).

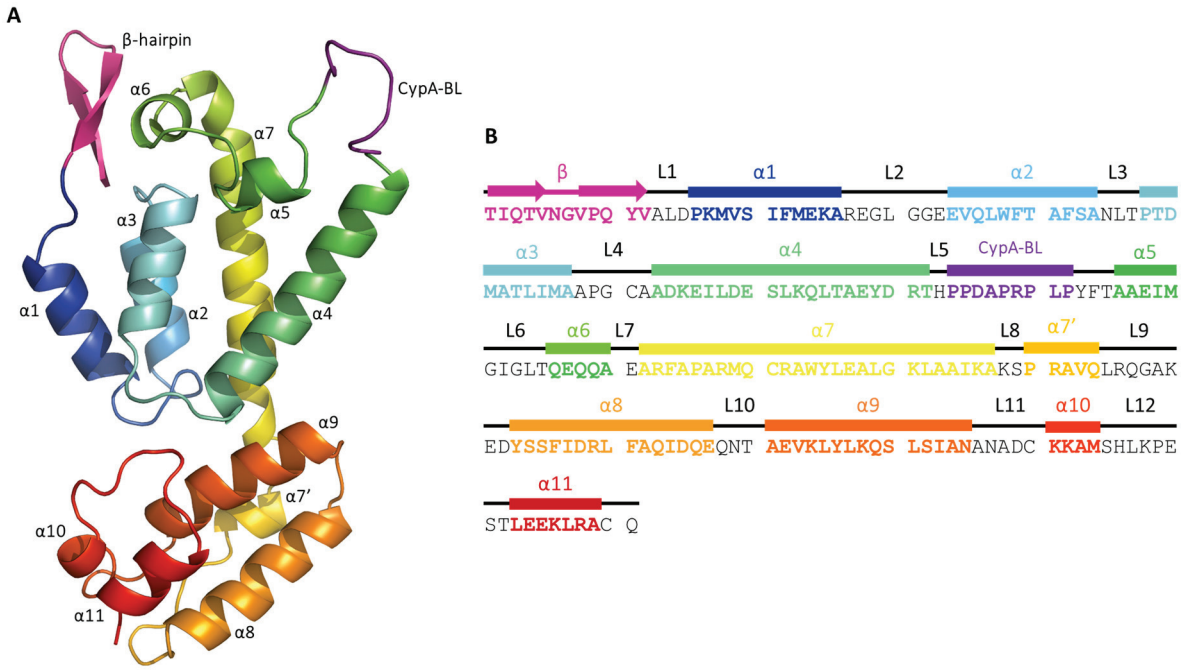

Figure 46 | Structures de la protéine de capside du FIV. (A) Structure cristallographique d'un monomère de CA du FIV. (B) Séquence de la CA du FIV et structures secondaires associées. Les hélices et boucles sont colorées du bleu au rouge de l'extrémité N-terminale à l'extrémité C-terminale et sont numérotées de α1 à α11 et de L1 à L12, respectivement.

Le domaine CA_{CTD}, lui, contient 4 hélices α (numérotées α8 à α11) ainsi que la zone d'homologie majeure (MHR, *Major Homology Region*) présente dans toutes les CA rétrovirales.

- Comparaison avec les structures connues de CA de lentivirus

Malgré de faibles similarités de séquence, le repliement global de la CA du FIV est similaire à celui des CA d'autres rétrovirus dont le VIH et EIAV (29% et 39%, respectivement). La comparaison de la structure de la CA du FIV avec celles d'autres rétrovirus (**Figure 47**) permet tout de même de distinguer quelques différences structurales.

Premièrement, l'orientation du CA_{CTD} de la CA du FIV par rapport à son CA_{NTD} est différente de celle des deux autres CA (**Figure 47A** et **Figure 47B**). En effet, ces deux domaines sont plus proches l'un de l'autre, ce qui rend la structure de la CA du FIV plus compacte que celles observées pour le VIH et EIAV.

Cependant, les CA_{NTD} et CA_{CTD} de la CA du FIV peuvent être superposés séparément avec ceux du VIH, d'EIAV (données non présentées) et de Relik (**Figure 47C**) avec un RMSD compris entre 0,5 et 1 Å, ce qui démontre la conservation globale du repliement de chaque domaine pour la CA du FIV.

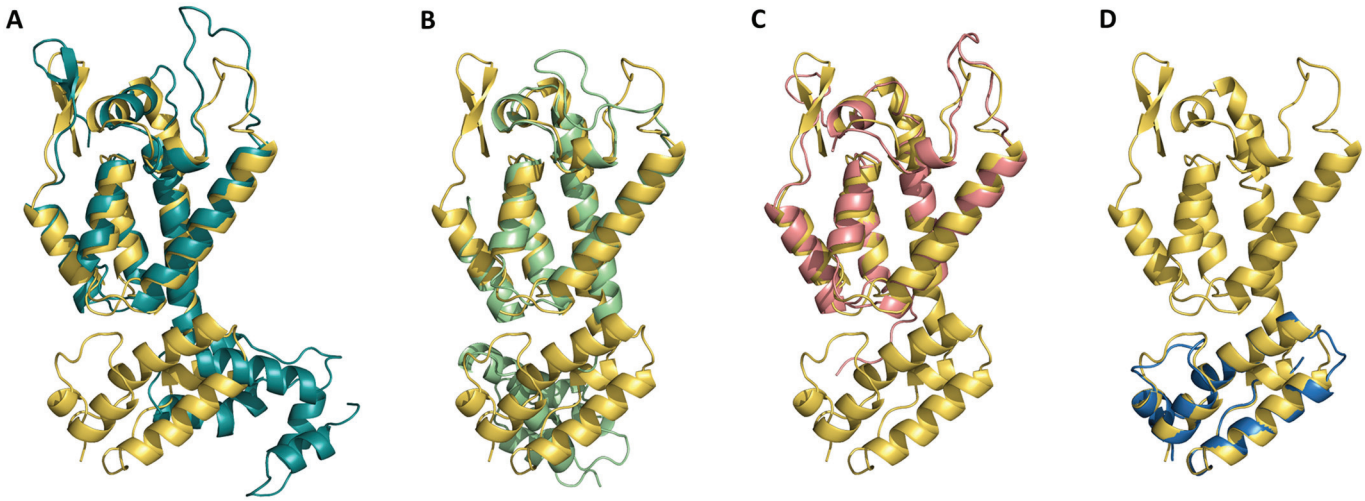

Figure 47 | Comparaison structurale des protéines de capside rétrovirales. Superposition d'un monomère de la CA du FIV (jaune) avec les CA entières de (A) VIH (teal, PDB ID : 3NTE (Du *et al.*, 2011)), (B) EIAV (vert, PDB ID : 2EIA (Jin *et al.*, 1999)) et les domaines (C) CA_{NTD} de Relik (rose, PDB ID : 2XGU (Goldstone *et al.*, 2010)) et (D) CA_{CTD} du FIV (bleu, PDB ID : 5DCK (Khwaja *et al.*, 2016)).

De façon plus détaillée, les 12 premiers résidus d'acides aminés de l'extrémité N-terminale de la CA du FIV se replient pour former un motif *β-hairpin* qui apparaît plus grand que celui du VIH (**Figure 47A**). Cette région n'est pas observée dans la structure de la CA d'EIAV (**Figure 47B**) et ne contient pas de structures secondaires dans celle de Relik (**Figure 47C**).

Une autre différence entre la CA du FIV et celles des autres rétrovirus concerne la boucle de liaison à la cyclophiline A, présente dans le CA_{NTD}, qui apparait plus petite que celle des autres lentivirus (**Figure 47**). La CypA-BL de la CA du FIV contient notamment une liaison peptidique *cis* entre les résidus d'acides aminés Arg89 et Pro90.

Cette liaison se trouve à 100% dans une conformation *cis* à la fois dans les chaînes A et B du dimère présent dans l'unité asymétrique (**Figure 48**). Cette conformation ne possède aucun équivalent publié dans les structures connues de CA rétrovirales.

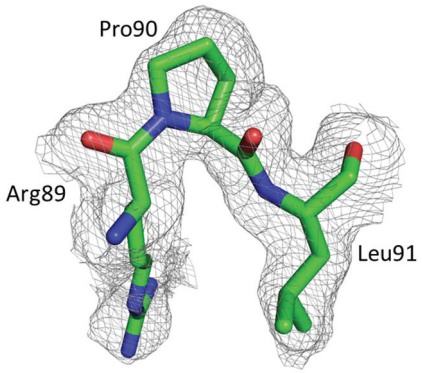

Figure 48 | Carte de type « *omit map* », contourée à 1σ, de la liaison peptidique *cis* Arg89-Pro90 présente dans la CypA-BL de la CA du FIV Les atomes d'oxygène et d'azote sont colorés en rouge et en bleu, respectivement.

Concernant le domaine carboxy-terminal, l'hélice 3_{10} observée dans la CA du VIH (entre les hélices α7 et α8) est remplacée par une hélice α dans le cas de la CA du FIV (numérotée α7') dont nous avons résolu la structure. Bien que cette région soit présente dans la structure du domaine CA_{CTD} seul de la CA du FIV (Khwaja *et al.*, 2016), celle-ci ne contient aucune structure secondaire (**Figure 47D**).

Enfin, les résidus d'acides aminés Cys190 et Cys210 de la CA du FIV présentent une conformation alternée pour laquelle un pont disulfure intramoléculaire se forme avec un pourcentage d'occupation d'environ 75% (données non présentées). Cette conformation alternée et le pont disulfure en résultant sont absents de la structure du CA_{CTD} de la CA du FIV récemment publiée (Khwaja *et al.*, 2016).

- <u>Interfaces oligomériques présentes dans le cristal de CA du FIV</u>

Au sein du cristal, la CA du FIV est organisée sous forme d'un dimère de dimères (**Figure 49A**), dont l'assemblage a été prédit comme stable en solution par le logiciel PISA (Krissinel et Henrick, 2007) permettant de détecter des oligomères biologiques. Le dimère A-B, formé par les chaînes A et B, présent dans l'unité asymétrique est lié à son symétrique A'-B' *via* un axe cristallographique d'ordre 2, afin de former le tétramère A-B-A'-B' (**Figure 49A**).

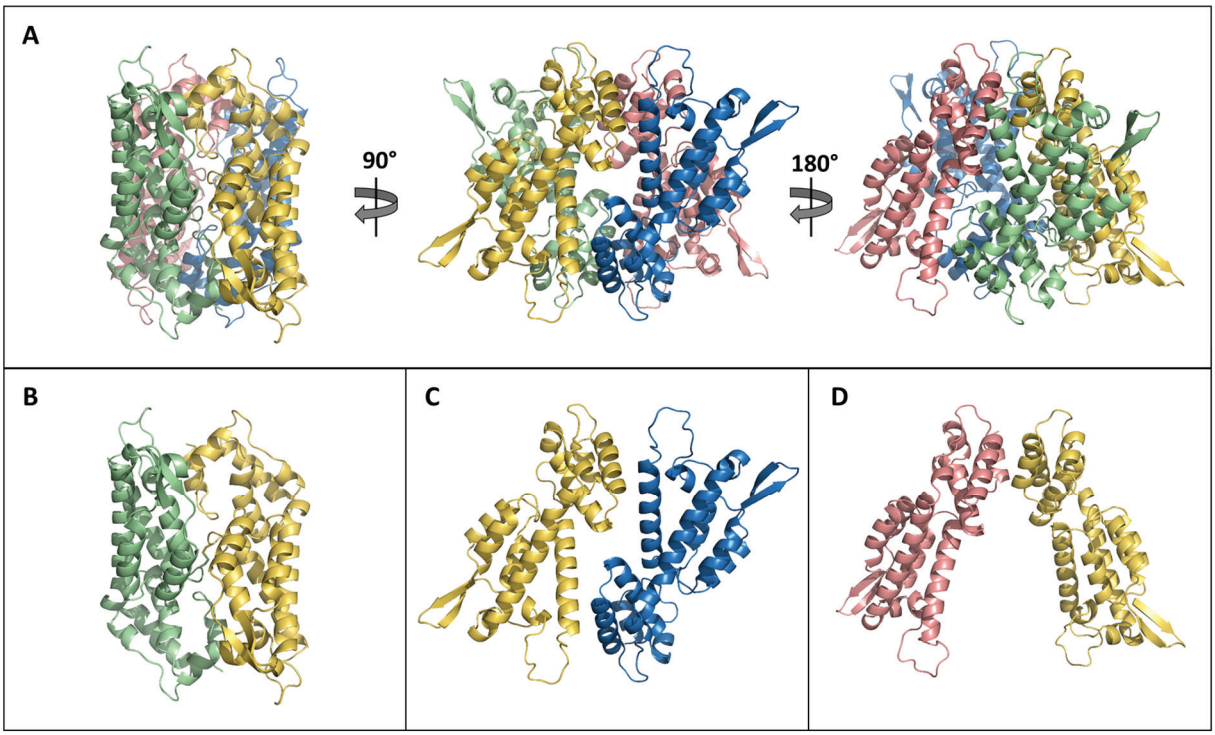

Figure 49 | Assemblage et interfaces de la CA du FIV présentes dans le cristal. (A) Assemblage de la CA du FIV sous forme tétramérique. Les chaînes A, B, A' et B' sont colorées en jaune, vert, bleu et rose, respectivement. Les trois orientations présentées correspondent au trois dimères présentés en dessous. (B) Interface 1 correspondant au dimère A-B de l'unité asymétrique. (C) Interface 2 correspondant au dimère A-A' impliquant des interactions entre la chaîne A de l'UA et son symétrique A'. (D) Interface 3 correspondant au dimère A-B' impliquant des interactions entre la chaîne A d'un dimère de l'UA et la chaîne symétrique B'.

La première interface stable (interface 1) (**Figure 49B**) correspond au dimère A-B de l'unité asymétrique, dans lequel les deux monomères de CA sont orientés de façon antiparallèle. La surface enfouie moyenne pour ce dimère est d'environ 1 100 $Å^2$ par monomère avec un ΔG de -21,1 kcal/mol. Il est à noter que les deux monomères A et B de ce dimère sont liés de façon covalente par un pont disulfure entre les résidus d'acides aminés Cys61 de chaque chaîne.

Cependant, d'après les analyses SDS-PAGE réalisées en conditions non réductrices (**Figure 50**) et de spectrométrie de masse (données non présentées), ce pont disulfure n'est pas présent dans l'échantillon protéique utilisé pour la cristallogenèse. Ces résultats suggèrent que le pont disulfure observé est un artéfact de cristallisation ayant permis de stabiliser le dimère au sein du cristal. Cette interface, qui est plus compacte que celle décrite pour d'autres lentivirus, ne semble pas fonctionnelle en raison des deux extrémités N-terminales de chaque chaîne qui sont orientées de façon antiparallèle. En effet, pour une interaction optimale avec la protéine de matrice du FIV, celles-ci devraient être orientées dans la même direction.

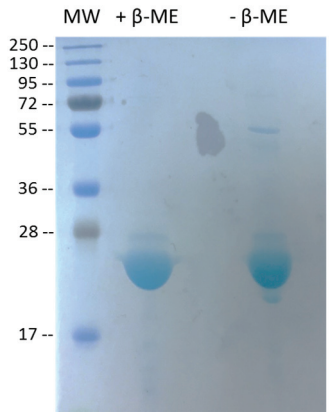

Figure 50 | Analyse SDS-PAGE de la solution protéique ayant permis l'obtention de cristaux de CA : Absence de pont disulfure entre les monomères de CA en solution. Les pistes +β-ME et –β-ME correspondent au dépôt de la protéine de CA en présence et en absence d'agent réducteur (β-mercaptoéthanol), respectivement.

A l'intérieur du tétramère, deux autres interfaces équivalentes (interface 2) correspondent aux dimères A-A' et B-B' (**Figure 49C**). Ces deux dimères sont formés par l'interaction d'une chaîne du dimère de l'unité asymétrique (A ou B) avec sa chaîne symétrique (A' ou B'). La surface enfouie moyenne pour ces dimères est d'environ 500 $Å^2$ par monomère avec un ΔG de 0 kcal/mol. De plus, ces interfaces, qui n'impliquent aucun lien covalent ni de pont salin, sont composées de deux monomères de CA orientées de façon antiparallèle. L'orientation des monomères, couplée à un ΔG nul, laisse suggérer que ces interfaces ne représentent pas non plus des dimères fonctionnels de la CA.

Une troisième interface prédite par PISA (interface 3) (**Figure 49D**) se forme entre la chaîne A de l'unité asymétrique et le symétrique de la chaîne B dans le tétramère (dimère A-B'). Cette interface, prédite comme stable, présente une surface enfouie moyenne de 500 $Å^2$ par monomère avec un ΔG de -2,4 kcal/mol. Les deux monomères de CA impliqués dans ce dimère sont orientés dans la même direction. Le domaine CA_{CTD} de la chaîne A (hélices α8 et α11) interagit avec le CA_{CTD} de la chaîne B' (hélices α8 et α11) pour former une interface proche de celle observée dans la structure du CA_{CTD} de FIV seul (Khwaja *et al.*, 2016).

Par ailleurs, une dernière interface (interface 4) (**Figure 51**), non impliquée dans le tétramère, est prédite comme un possible assemblage biologique par PISA. Cette interface, impliquée dans les contacts cristallins, comprend des interactions entre la chaîne B de l'unité asymétrique et la chaîne A d'un tétramère symétrique (dimère B-A'').

Au sein de ce dimère, le CA$_{NTD}$ de la chaîne B (hélices α2, α3 et α4) interagit avec le CA$_{NTD}$ de la chaîne A'' (hélices α3, α5 et α7). La surface enfouie moyenne pour ce dimère est d'environ 840 Å^2 par monomère avec un ΔG de -13 kcal/mol. Cette interface implique une symétrie cristallographique d'axe 2, absente des interfaces décrites pour d'autres CA lentivirales. La combinaison de l'important ΔG et de la surface enfouie avec l'orientation parallèle des deux monomères peut refléter une vraisemblance biologique pour ce dimère, bien qu'aucune interface de ce type n'ait été observée pour d'autres capsides rétrovirales.

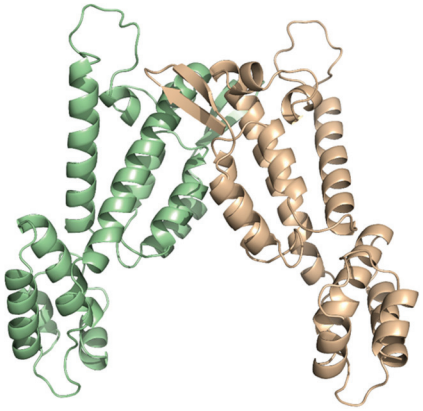

Figure 51 | Interface 4 (dimère B-A''). Cette interface, prédite par PISA, implique des interactions entre la chaîne B (verte) de l'unité asymétrique et la chaîne A'' (beige) d'un tétramère symétrique.

2.3.2. Etude structurale par SAXS

L'identification d'un dimère biologique de CA du FIV à partir des données cristallographiques n'étant pas concluante, des analyses complémentaires par SAXS ont été réalisées afin d'obtenir des informations structurales sur la protéine en solution.

D'après l'analyse de ces données (**Figure 52**), la masse molaire estimée pour la CA en tampon HEPES à 6 mg.mL^{-1}, *via* l'extrapolation de la valeur de I(0), est d'environ 27 kDa (**Figure 52A**). Cette masse moléculaire est proche de la masse moléculaire attendue pour un monomère de la protéine d'intérêt (28 kDa). Ce résultat est en adéquation avec ceux obtenus précédemment par (Serriere *et al.*, 2013) pour lesquels une concentration supérieure à 7 mg.mL^{-1} était nécessaire pour la formation d'un dimère de CA.

Le rayon de giration (R_g) de la protéine, estimé à partir de la représentation de Guinier (**Figure 52A**), est d'environ 27 Å pour un diamètre maximal (D_{max}) d'environ 75 Å, estimé à partir de la courbe de distribution des distances P(r) (**Figure 52B**). Cette courbe comporte notamment un épaulement (**Figure 52B**) caractéristique des protéines multi-domaines (Koch et Svergun, 2003). Ce résultat confirme donc que les données cristallographiques montrant que la CA du FIV est composée de plusieurs domaines, tout comme les protéines de capside d'autres rétrovirus dont la structure a été résolue (VIH, EIAV, Relik, etc…).

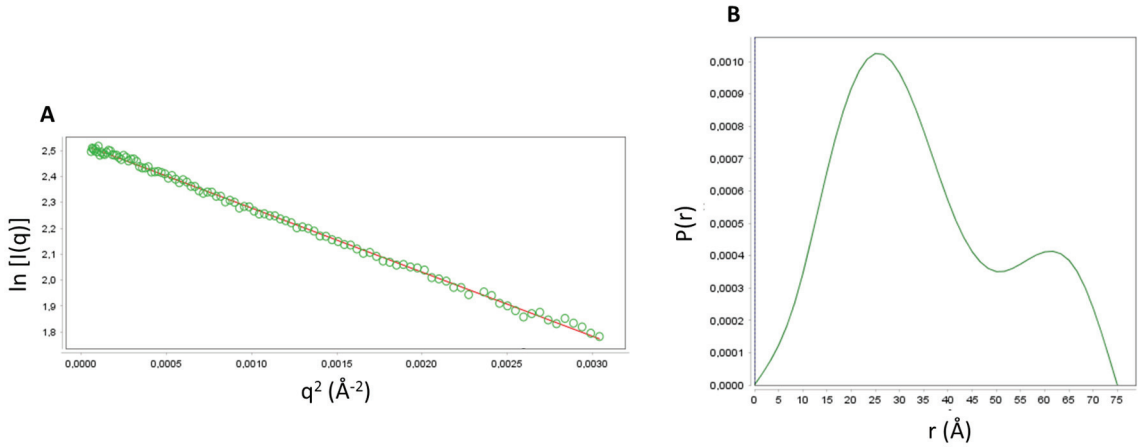

Figure 52 | Analyse des données SAXS de la protéine de capside à 6 mg.mL^{-1} en tampon HEPES : (A) représentation de Guinier et (B) courbe de distribution des distances P(r).

D'après l'enveloppe SAXS calculée *ab initio* (**Figure 53**), la protéine de capside du FIV à 6 mg/mL^{-1} adopte une forme monomérique en solution. Une bonne corrélation (χ^2 = 1,4) est observée lors de l'ajustement de la structure de notre monomère de CA dans l'enveloppe SAXS obtenue.

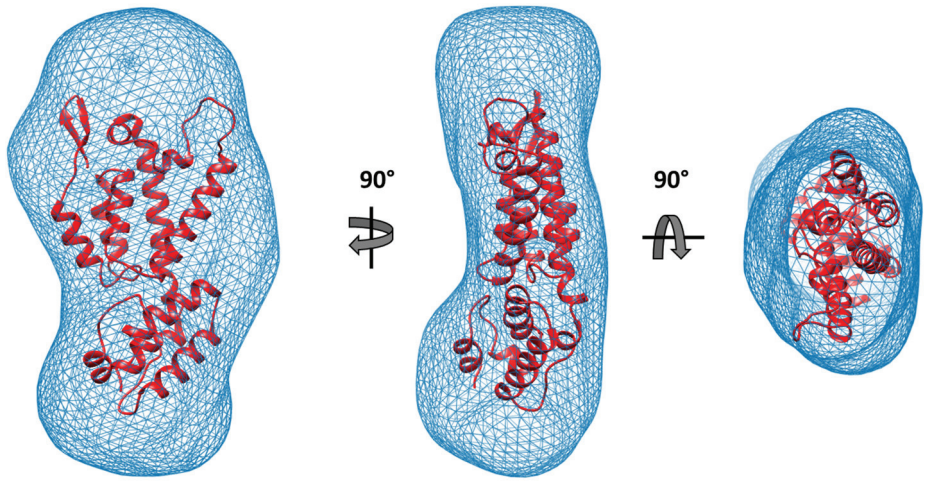

Figure 53 | Ajustement de la structure cristallographique de la CA du FIV (rouge) dans l'enveloppe SAXS obtenue pour la CA en solution (bleu).

2.4. Identification d'inhibiteurs de l'assemblage de la CA du FIV

La protéine de capside du FIV étant une protéine essentielle à la formation de virions infectieux, elle constitue une cible thérapeutique de choix dans le développement d'une stratégie anti-FIV.

Une méthode de criblage *in vitro* de composés, basée sur la spectrophotométrie à 340 nm, a été mise en place en collaboration avec le Dr. G. Alvarez (*Universidad de la Republica*, Uruguay). Cette méthode permet de suivre la turbidité de la solution protéique au cours du temps, une augmentation de celle-ci traduisant un assemblage de la protéine de capside. Ainsi, une cinquantaine de composés issus de la chimiothèque du Dr. G. Alvarez ont été testés dans le but d'évaluer leur effet sur l'oligomérisation de la CA du FIV et d'identifier un inhibiteur potentiel de l'assemblage capsidique.

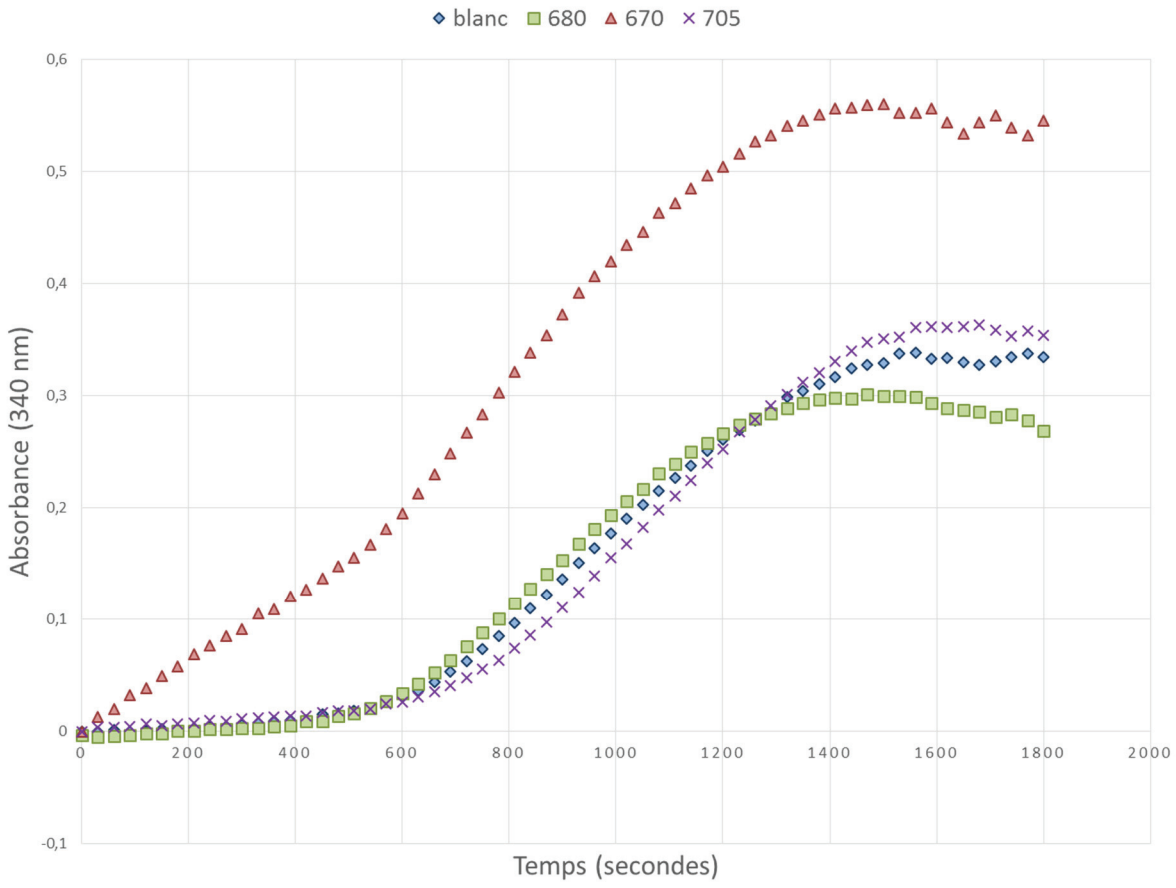

Figure 54 | Suivi de l'assemblage capsidique de la CA du FIV *in vitro* en absence (bleu) et en présence de composés.

Les tests réalisés sur le suivi de l'assemblage capside en présence de différents composés ont permis d'identifier trois catégories de molécules selon leurs effet sur l'oligomérisation de la CA *in vitro* :

- Les composés sans effet dont la courbe d'assemblage est similaire à celle de la protéine seule (**Figure 54, composé 705**) ;
- Les composés à effet « *enhancer* » qui augmentent la vitesse d'assemblage de la CA et/ou la taille finale des particules obtenues (**Figure 54, composé 670**);
- Les composés à effet inhibiteur qui diminuent la vitesse d'assemblage de la CA et/ou la taille finale des particules obtenues (**Figure 54, composé 680**).

Les composés ayant eu un effet inhibiteur sur l'assemblage de la CA du FIV sont ceux qui ont été utilisés pour la co-cristallisation avec la protéine de capside (**Résultats et discussion 2.3.1.1**). Les courbes correspondant à l'assemblage de la CA en présence des quatre composés ayant permis l'obtention de cristaux sont présentées en **Figure 55**.

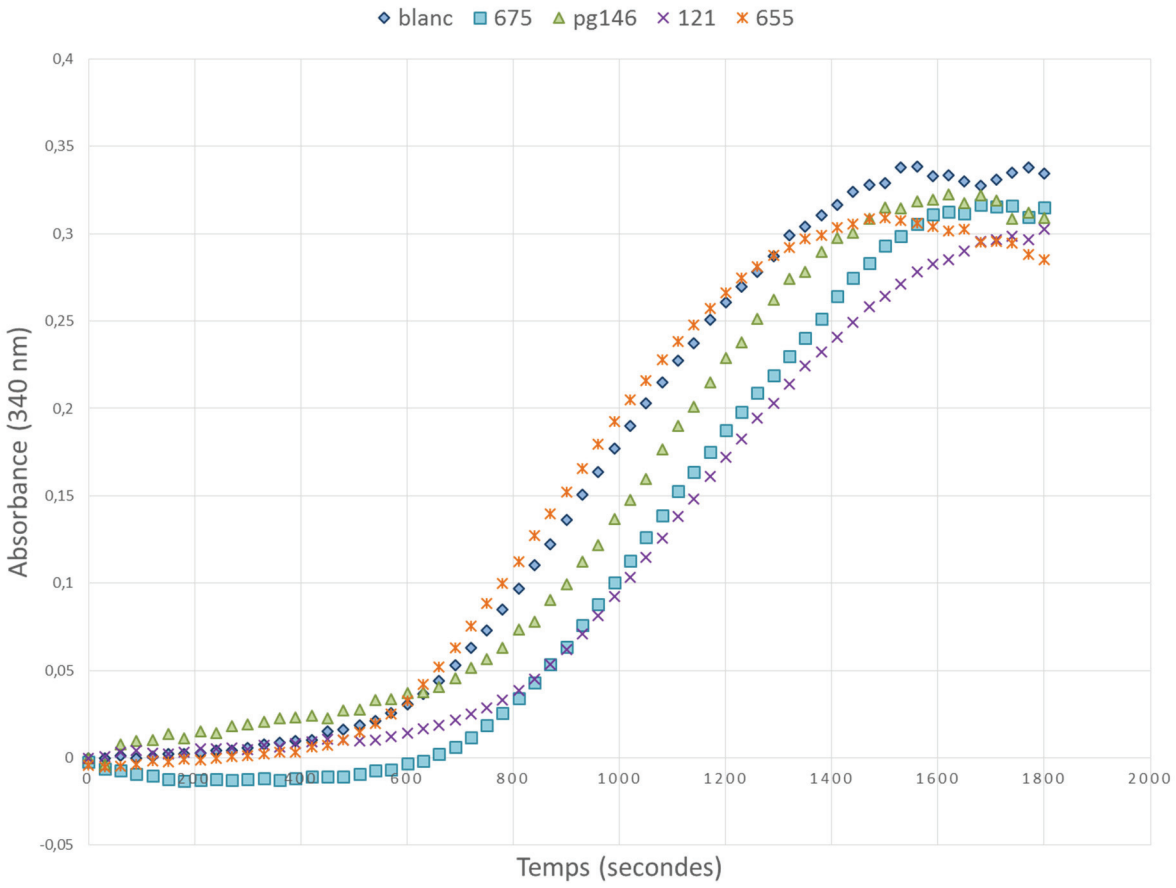

Figure 55 | Suivi de l'assemblage capsidique de la CA du FIV *in vitro* en absence (bleu) et en présence de composés inhibiteurs.

Pour le composé 655, la vitesse d'assemblage de la CA reste inchangée mais la taille finale des particules obtenues diminue. En effet, le maximum d'absorbance atteint au plateau en présence de la molécule 655 (**Figure 55, composé 655**) est inférieur à celui de la protéine seule (**Figure 55, blanc**).

Concernant les composés 675, pg146 et 121, une diminution de la vitesse d'assemblage est observée, couplée à une légère diminution de la taille finale des particules. En effet, les courbes d'assemblage pour ces trois molécules sont décalées vers la droite (**Figure 55, composés 675, pg146 et 121**) et le maximum d'absorbance atteint au plateau est inférieur à celui de la protéine seule (**Figure 55, blanc**).

2.5. Discussion

La CA native du FIV n'ayant pas permis d'obtenir des cristaux diffractant les rayons X, malgré le criblage de milliers de conditions, deux modifications de séquence lui ont été apportées. Premièrement, son premier résidu (Pro1) a été muté en Thr afin de permettre le clivage de son étiquette His_6 par le Facteur X, laissant ainsi son extrémité N-terminale intacte (tous les résidus du site de clivage étant éliminés). La seconde modification consiste en la troncation des 11 derniers résidus d'acides aminés de son extrémité C-terminale, prédits comme flexibles, afin de favoriser la croissance des cristaux. Ces deux modifications apportées à la protéine étudiée n'ont pas altéré sa capacité à s'oligomériser en solution. En effet, un dimère est observé pour la CAΔC-PT étudiée à une concentration de 7 $mg.mL^{-1}$ (**Figure 42**) comme dans le cas de la protéine CA native (Serriere *et al.*, 2013). Ces modifications n'ont également pas altéré la capacité de la CAΔC-PT à s'assembler *in vitro* en présence de sels, puisqu'une augmentation de son état d'oligomérisation est observée (**Figure 43**). La protéine CAΔC-PT étudiée représente donc bien une forme fonctionnelle de la CA du FIV.

Afin de déterminer les spécificités des mécanismes moléculaires sous-jacents à l'assemblage du FIV, nous avons étudié la structure de la sa protéine de capside entière en la comparant notamment à d'autres structure de CA rétrovirales. Comme les protéines de capside d'autres rétrovirus (VIH, EIAV ou encore Relik), la CA du FIV est majoritairement composée d'hélices α et est organisée en deux domaines : le CA_{NTD} et le CA_{CTD}. Ce résultat confirme que le repliement général des protéines de capside est très bien conservé parmi les rétrovirus.

En effet, la structure de la CA du FIV présente de nombreuses caractéristiques communes avec celles des autres lentivirus. Par exemple, elle contient à son extrémité amino-terminale un motif *β-hairpin*. Celui-ci a été identifié comme indispensable à la formation de la capside virale dans le cas du VIH puisqu'il participe directement aux interactions intermoléculaires CA-CA (Momany *et al.*, 1996) (Tang *et al.*, 2002). Au sein de la structure de la CA du FIV, ce *β-hairpin* semble adopter une conformation de type « ouverte » (**Figure 56**) comme celle décrite pour la CA du VIH. Cette conformation pourrait être importante pour l'import des dNTP à l'intérieur du virus lors de l'étape de transcription inverse (Jacques *et al.*, 2016), renforçant ainsi la pertinence biologique de notre structure.

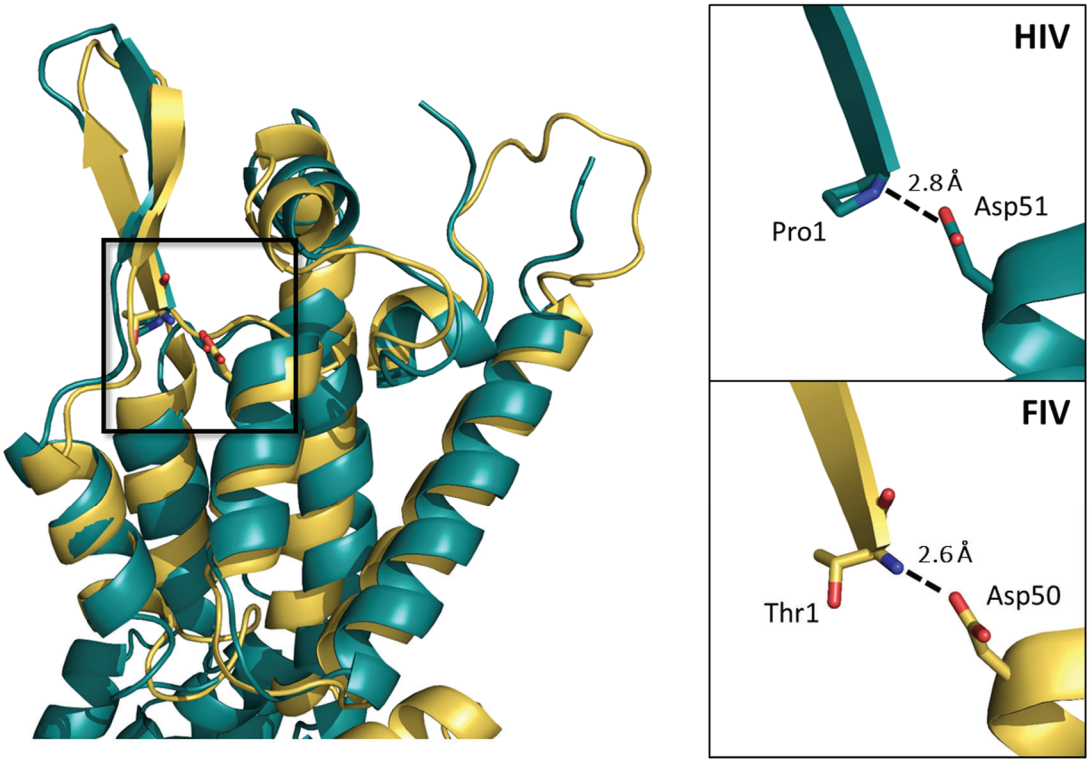

Figure 56 | Comparaison des motifs *β-hairpin* des CA du FIV et du VIH. Les CA_{NTD} du FIV (jaune) et du VIH dans sa conformation « ouverte » (bleu, PDB ID : 5HGL) ont été superposés. Deux vues détaillées, à droite, montrent les ponts salins formés entre la Pro1 et l'Asp51 du VIH (au-dessus) et entre la Thr1 et l'Asp50 du FIV (au-dessous).

Dans le cas du VIH, la Pro1 a été décrite comme essentielle pour la formation du motif N-terminal *β-hairpin* (Cortines *et al.*, 2011). Il est à noter que cette proline est conservée dans la plupart des CA lentivirales telles que celles du VIH, du FIV, du SIV ou encore d'EIAV, reflétant probablement une fonction clé dans la réplication virale. Toutefois, la mutation de la Pro1 en Thr1 introduite dans notre CA du FIV n'a pas empêché la formation du motif *β-hairpin* dans notre structure (**Figure 46**), ni déstabilisé son assemblage *in vitro* (**Figure 43**).

De plus, le pont salin entre le groupe NH^{2+} N-terminal de la Pro1 et le groupe carboxyl de la chaîne latérale de l'Asp51, servant à stabiliser le motif *β-hairpin* de la CA du VIH, possède un équivalent dans la structure de la CAΔC-PT du FIV, où un pont salin est observé entre le groupe NH^{2+} N-terminal de la Thr1 et le groupe carboxyl de la chaîne latérale de l'Asp50 (**Figure 56**). Ces deux ponts salins sont également équivalents en termes de longueur avec respectivement 2,6 Å pour le VIH et 2,8 Å pour le FIV. Une étude fonctionnelle de ce mutant P1T de la CA permettrait de mieux comprendre les fonctions de cette proline. Celle-ci pourrait, par exemple, jouer un rôle important dans la conformation adoptée par le précurseur Gag à la jonction MA-CA, afin de permettre son clivage efficace par la protéase virale pendant l'étape de maturation.

La boucle de liaison à la cyclophiline A, présente dans toutes CA rétrovirales, est également observée dans le domaine CA_{NTD} de la CA du FIV. De façon intéressante, une liaison peptidique *cis* est observée entre l'Arg89 et la Pro90 de la CypA-BL de la CA du FIV (**Figure *48***). La cyclophiline A étant une *cis-trans* peptidylprolyl isomérase possédant une forte spécificité pour les *cis*Pro (Konno *et al.*, 1996), la présence de cette liaison peptidique *cis* pourrait expliquer pourquoi le résidu Pro90 a été démontré comme essentiel pour la liaison à la cyclophiline A dans le cas du FIV (Lin et Emerman, 2006). De manière intéressante, aucune *cis*Pro n'a été décrite à ce jour dans les structures de capside rétrovirale connues.

Enfin, les deux derniers résidus du domaine CA_{CTD} de la CA du FIV ne sont pas définis en densité électronique. Cela confirme la grande flexibilité de l'extrémité carboxy-terminale de la CA, que nous avions déjà tronquée de 11 résidus d'acides aminés pour éviter les problèmes de cristallogenèse. Cette flexibilité de l'extrémité C-terminale pourrait permettre à la jonction CA-NC d'adopter une conformation optimale pour le clivage du précurseur Gag comme cela a déjà été supposé pour l'extrémité C-terminale de la protéine MA du FIV au niveau de la jonction MA-CA (Serriere *et al.*, 2013) ou encore au niveau de la jonction CA-sp1 de Gag du VIH (Wright *et al.*, 2007).

Ainsi, les différents domaines impliqués dans les fonctions des CA rétrovirales et ceux ayant été observés comme importants pour la réplication du FIV sont présents dans notre structure de la CA du FIV.

Durant l'étape d'assemblage des virions, les CA rétrovirales s'organisent en pentamères et en hexamères (Zhao *et al.*, 2013) (Gres *et al.*, 2015) afin de former une capside conique. Cependant, aucun assemblage de type pentamérique ou hexamérique n'a été observé dans le cas de la CA du FIV que nous avons cristallisé. En effet, seules des interfaces dimériques, dont la pertinence fonctionnelle reste encore à déterminer, ont été identifiées à l'intérieur du cristal. Cela reflète très probablement la plasticité intermoléculaire de la CA nécessaire aux différentes étapes d'assemblage de la capside virale.

La structure cristalline de la CAΔC-PT du FIV contient un dimère de protéine de capside (chaînes A et B) dans l'unité asymétrique. Ce dimère est lié de façon covalente par un pont disulfure entre les Cys61 de chaque monomère. Nous avons démontré que ce pont disulfure résultait d'un artéfact de cristallisation puisqu'il est n'est pas présent dans la solution protéique précédant l'étape de cristallogenèse (**Figure 50**). Cette absence de pont disulfure est cohérente avec les résultats de (Nath et Peterson, 2001) selon lesquels les Cys61 ne sont pas impliquées dans la formation de ponts disulfures dans la CA du FIV. Cet artéfact de cristallisation est très certainement dû à la présence de DMSO, introduit dans la condition de cristallisation lors de l'ajout du composé potentiellement inhibiteur. En effet, il a été démontré que le DMSO promeut l'oxydation des thiols en disulfures à pH faible et à température ambiante (Liu *et al.*, 2009). La présence de ce pont disulfure artéfactuel pourrait avoir stabilisé le dimère de CA lors du processus de cristallisation. Cela expliquerait la formation des cristaux en présence de DMSO alors que ceux-ci n'ont pas été obtenus dans les mêmes conditions en absence de DSMO. Cependant, la présence de ce pont pourrait également avoir stabilisé les dimères de CA dans des interfaces non fonctionnelles. Des expériences de cristallogenèse en absence de DMSO sont en cours afin d'identifier des interfaces oligomériques fonctionnelles pour la CA du FIV. En parallèle, des mutants de la CA du FIV Cys61Ser sont actuellement en construction dans le but d'obtenir la structure de la CA du FIV en absence du pont disulfure observé entre les Cys61.

Au sein de la CA du FIV les Cys190 et Cys210, qui sont conservées dans toutes les CA rétrovirales, présentent une conformation alternée. L'une de ces conformations, présente dans 75% des protéines du cristal, implique un pont disulfure intrachaîne qui est absent de la structure du CA_{CTD} de FIV récemment publiée (Khwaja *et al.*, 2016). Bien qu'il puisse également être du à la présence de DMSO dans le cristal, la présence de ce pont disulfure est en accord avec les études biochimiques menées par (Nath et Peterson, 2001) qui ont démontré que les Cys190 et Cys210 de la CA du FIV sont impliquées dans un pont disulfure

intramoléculaire, qui est indispensable à l'assemblage de la capside du FIV et à son pouvoir infectieux (Nath et Peterson, 2001). De plus, la présence d'un pont disulfure dans le domaine CTD de la CA est extrêmement conservée dans les CA rétrovirales comme celles du VIH et d'EIAV (Jin *et al.*, 1999) (Gamble *et al.*, 1997). Ainsi, cette caractéristique structurale est probablement pertinente pour la biologie et les fonctions de la CA du FIV.

Comme attendu d'après les données biochimiques (Nath et Peterson, 2001), le dernier résidu Cys121 libre de la CA du FIV (hélice α7) n'est pas impliqué dans la formation d'un pont disulfure.

Curieusement, l'atome de soufre de cette cystéine forme, avec ceux des Met51 (hélice α3) et Met100 (hélice α5), un *cluster* de soufre dans lequel ces atomes sont alignés et séparés de 4 Å (**Figure 57**). Cette structuration est inédite et spécifique à la CA du FIV mais aucune donnée sur sa fonction, s'il y en a une, n'est disponible actuellement.

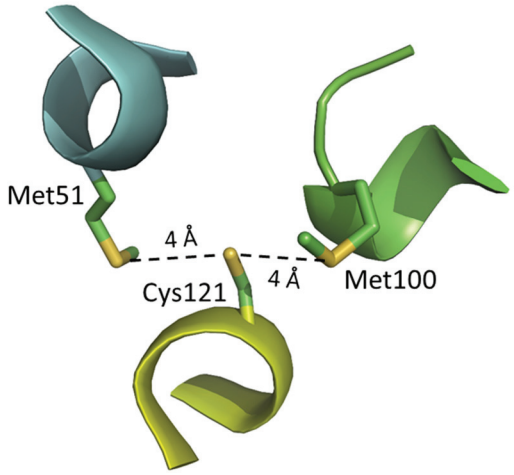

Figure 57 | *Cluster* de soufre impliquant les atomes de soufre des chaînes latérales des résidus d'acides aminés Met51 (α3), Met100 (α5) et Cys121 (α7).

Enfin, nous avons pu confirmer que la structure cristalline obtenue pour la forme monomérique de la CA du FIV est proche de celle en solution en utilisant des expériences de SAXS. En effet, une bonne corrélation est observée lors de l'ajustement de la structure du monomère de CA obtenue par cristallographie aux rayons X dans l'enveloppe SAXS expérimentale (**Figure 53**). Ce résultat couplé aux caractéristiques structurales de la CA et à sa capacité à s'assembler *in vitro* confirme que la structure cristallographique de la CAΔC-PT du FIV résolue représente, dans sa forme monomérique une protéine de capside fonctionnelle, comparable à celle du VIH et d'EIAV.

Cependant, même si la structure cristallographique de la CA FIV entière monomérique démontre des caractéristiques originales, des arrangements oligomériques supérieurs tels que les pentamères ou hexamères permettraient de mieux comprendre les particularités de l'assemblage de la capside du FIV comparativement aux autres lentivirus. Cela permettrait également de valider la pertinence des résultats préalablement obtenus par notre équipe, selon lesquels l'assemblage capsidique du FIV se fait sous forme de particules sphériques, contrairement à ce qui a été observé chez d'autres rétrovirus (Serriere *et al.*, 2013).

La protéine de capside étant essentielle à la formation de particules virales infectieuses, elle constitue une cible de choix pour le développement de nouvelles stratégies antirétrovirales.

Ainsi, de nombreuses molécules de la chimiothèque du Dr. G. Alvarez ont été testées afin d'identifier de potentiels inhibiteurs de l'assemblage capsidique de la CA du FIV. Parmi ces molécules, certaines ont eu un léger effet inhibiteur sur l'assemblage de la CA du FIV *in vitro* et constituent, par conséquent, des « *hits* » potentiels. Afin de confirmer ces résultats, de nouvelles expériences de spectrophotométrie seront effectuées avec ces composés en triplicat dans le but d'obtenir des résultats statistiquement satisfaisants. De plus, des études par thermophorèse sont également en cours afin de déterminer les constantes d'affinité (K_D) de ces molécules pour la protéine de capside.

Par ailleurs, les expériences de co-cristallisation avec les composés potentiellement inhibiteurs n'ont pas permis d'obtenir la structure de la protéine de capside en présence de ceux-ci. De nouveaux essais de co-cristallisation, en augmentant la quantité de composés dans la goutte, ou de trempage sont envisagés afin de résoudre la structure de la CA du FIV en présence de son (ses) inhibiteur(s). L'obtention de la structure tridimensionnelle du complexe CA:inhibiteur permettrait d'identifier les sites de fixation de ces composés et de comprendre leurs mécanismes d'inhibition de l'assemblage capsidique. En effet, ceux-ci pourraient par exemple se lier aux interfaces CA_{NTD}-CA_{NTD}, CA_{NTD}-CA_{CTD} ou CA_{CTD}-CA_{CTD} se formant entre les CA de la capside virale et, ainsi, déstabiliser la formation de la capside virale.

CONCLUSION ET PERSPECTIVES

Ce projet doctoral a porté sur l'étude fonctionnelle et structurale de deux protéines rétrovirales d'intérêt thérapeutique, impliquées dans deux étapes essentielles de la réplication des rétrovirus, à savoir la régulation de l'expression des protéines virales et l'assemblage des nouvelles particules virales. En effet, les rétrovirus, dont le plus connu reste le VIH, constituent à l'heure actuelle un problème de santé publique majeur tant chez l'Homme que chez l'animal. La compréhension des déterminants structuraux sous-jacents à la fonction de leurs protéines virales constitue donc une étape essentielle dans le développement de nouvelles stratégies antirétrovirales efficaces.

De manière générale, l'étude fonctionnelle et structurale de protéines d'intérêt thérapeutique est caractérisée par quatre étapes majeures : i) l'expression et la purification des protéines, ii) leur caractérisation biochimique et fonctionnelle, iii) la résolution de leurs structures et iv) l'identification d'inhibiteurs de leur activité ; chaque étape pouvant donc être limitante dans l'avancée d'un tel projet.

Concernant la première étape, l'expression des protéines en système bactérien et leur purification par chromatographie d'affinité sont les stratégies les plus couramment utilisées en raison de leur facilité à mettre en œuvre. Cependant, dans le cas de protéines difficiles telles que les protéines toxiques ou membranaires, cette étape peut vite devenir limitante car ces méthodes ne permettent pas d'aboutir à une production efficace des protéines étudiées.

C'est le cas de la protéine transactivatrice de la transcription Tax-2 du virus HTLV-2 étudiée au cours de ce projet. En effet, à ce jour, nous n'avons pas réussi à produire cette protéine seule sous forme soluble en système bactérien. Cependant, de récents essais de co-expression en présence de son partenaire cellulaire, le facteur de transcription CREB, se sont avérés prometteurs. Cette condition de co-expression doit encore être validée puis optimisée dans le but d'obtenir des complexes Tax-2:CREB purs et en quantité suffisante pour envisager des essais de cristallogenèse. Dans le cas contraire, de nouvelles stratégies d'expression utilisant des systèmes eucaryotes ou acellulaires devront être mises en place pour aboutir à la production de la protéine Tax-2.

En parallèle, la protéine Tax-1 du virus HTLV-1 a également été étudiée au laboratoire par le Dr. M. Dujardin. L'expression et la purification de Tax-1, elles, ont été réalisées avec succès, avec le protocole de co-expression présenté dans ce manuscrit, bien que la quantité de protéine pure obtenue soit assez faible.

La caractérisation biochimique des protéines étudiées constituent également une étape essentielle de ce type de projet afin d'étudier des protéines fonctionnelles et de pertinence biologique. Pour cela, de nombreuses techniques biochimiques et biophysiques peuvent être utilisées afin de déterminer les états oligomériques des protéines d'intérêt, leur composition en structures secondaires ou encore leur fonctionnalité.

Concernant la protéine Tax-2, lorsque celle-ci aura pu être produite, des caractérisations biochimiques et biophysiques permettront d'obtenir des informations sur sa composition en structures secondaires et sa fonctionnalité, respectivement.

Dans le cas de l'étude de la protéine de capside du FIV, dont les protocoles d'expression et de purification avaient déjà été mis au point avant mon arrivée, des études de suivi de son assemblage *in vitro* par DLS et de mesure de la turbidité ont permis de confirmer que celle-ci était toujours fonctionnelle, malgré les modifications apportées à sa séquence.

Afin de mieux comprendre les mécanismes moléculaires impliqués dans les fonctions biologiques de ces protéines d'intérêt thérapeutique, la résolution de leurs structures constitue une étape cruciale.

Dans le cas des protéines Tax-1 et Tax-2, l'obtention de leurs structures tridimensionnelles permettrait, par exemple, de mieux comprendre les différences de pathogénicité observées entre les virus HTLV-1 et HTLV-2. Pour la protéine Tax-1, une condition de cristallisation a pu être identifiée et a permis d'obtenir des cristaux dont le pouvoir de diffraction a été testé aux rayons X. Si un jeu de données à 4 Å a été enregistré, il ne s'agit cependant pas de la protéine Tax seule. En effet, la protéine GroEL de *E. coli* a également été identifiée dans le cristal. La structure de GroEL est actuellement en cours de résolution par remplacement moléculaire, ce qui permettra, par la suite, de résoudre celle de Tax-1 dans le cas où une densité électronique supplémentaire et correspondante serait observée. Cette condition de cristallisation de Tax-1 pourrait servir de guide pour cristalliser la protéine Tax-2 lorsque celle-ci aura été produite efficacement. Dans le cas de l'obtention de complexes Tax:CREB, des études par cryo-microscropie électronique, nécessitant des macromolécules de grande taille, sont également possibles afin de déterminer la structure de ce complexe.

Concernant la protéine de capside du FIV, l'étape de cristallogenèse a été la plus limitante, avec plus de 5000 conditions de cristallisation différentes testées pour les différentes constructions de CA produites au laboratoire. Néanmoins, une condition permettant l'obtention de cristaux de CA a pu être identifiée, ce qui nous a permis de résoudre la

structure tridimensionnelle d'un mutant fonctionnel à 1,9Å de résolution par cristallographie aux rayons X. Cette structure a permis de confirmer, pour la CA du FIV, la présence des différents domaines impliqués dans les fonctions des CA rétrovirales ainsi que ceux ayant été décrits comme important pour la réplication du FIV.

Cependant, aucun assemblage de type pentamérique ou hexamérique, nécessaire à la formation d'une capside virale fonctionnelle, n'a été observé dans la structure cristalline que nous avons obtenue pour la CA du FIV. De nouveaux essais de cristallogenèse sont actuellement en cours afin d'obtenir sa structure dans des arrangements oligomériques supérieurs. En effet, obtenir la structure de la CA du FIV sous forme pentamérique et hexamérique permettrait d'identifier les interfaces d'interactions protéiques impliquées dans l'assemblage de ce virus et, ainsi, de mieux comprendre les spécificités de son assemblage comparativement aux autres lentivirus.

Enfin, la résolution de structures de protéines représentant des cibles thérapeutiques constitue également une base solide pour l'identification et la conception rationnelle de molécules inhibant leur activité.

Pour la protéine Tax, par exemple, l'obtention de sa structure permettrait d'identifier une molécule inhibant sa dimérisation, démontrée comme essentielle pour son activité de transactivateur de la transcription. Une telle molécule constituerait la première étape d'une stratégie thérapeutique anti-HTLV.

Dans le cas de l'étude de la CA du FIV, de nombreuses molécules, issues de la chimiothèsque de nos collaborateurs uruguayens, ont été criblées *in vitro* afin d'identifier un inhibiteur ciblant l'assemblage de la CA. À ce jour, une douzaine de composés ont été identifiés comme potentiellement inhibiteurs de l'assemblage de la CA *in vitro*. La détermination des constantes d'affinité entre ces composés et la CA est en cours de réalisation. Néanmoins, l'obtention de la structure tridimensionnelle de la CA constitue une base pour l'amélioration des composés inhibiteurs grâce à des techniques de modélisation *in silico*. L'amélioration des composés identifiés pourront ainsi être réalisées afin, par exemple, d'augmenter leur affinité pour la protéine cible et, par conséquent, de potentialiser leur action. Ces molécules candidates pourront ensuite être synthétisées par l'équipe de chimistes du Dr. G. Alvarez en Uruguay, puis testées *in vitro* afin de vérifier l'amélioration de leurs effets sur l'assemblage de la CA du FIV. Ces molécules pourront également être utilisées pour co-cristalliser la CA en présence de ses inhibiteurs et obtenir la structure tridimensionnelle du complexe CA:inhibiteur. Cela

permettrait de confirmer les sites de fixation de ces composés prédits par *docking* et de comprendre leurs mécanismes d'inhibition de l'assemblage capsidique.

Enfin, à plus long terme, la molécule inhibitrice identifiée pour la CA du FIV pourrait, par exemple, servir de modèle pour la conception d'une molécule inhibant l'assemblage du VIH, si elle cible une zone structurellement conservée de la CA.

Pour conclure, les rétrovirus étant capable de développer des résistances aux thérapies antirétrovirales, de nouvelles molécules inhibitrices doivent encore être identifiées afin d'étayer les stratégies de multi-thérapie existantes. Cela permettrait de cibler conjointement plusieurs étapes essentielles de la réplication des rétrovirus et de lutter efficacement contre leur infection. Pour cela, l'étude structurale et fonctionnelle de protéines cruciales pour la réplication et la pathogénicité des rétrovirus est une étape essentielle dans le développement de thérapies antirétrovirales. En effet, en permettant une meilleure compréhension des déterminants structuraux sous-jacents à la fonction de ces protéines, ce type d'étude ouvre la voie à l'identification de nouveaux composés antirétroviraux spécifiques à chaque rétrovirus.

BIBLIOGRAPHIE

A

Abdusetir Cerfoglio, J. C., S. A. Gonzalez and J. L. Affranchino (2014). "Structural elements in the Gag polyprotein of feline immunodeficiency virus involved in Gag self-association and assembly." J Gen Virol **95**(Pt 9): 2050-2059.

Adams, P. D., P. V. Afonine, G. Bunkóczi, V. B. Chen, N. Echols, J. J. Headd, L.-W. Hung, S. Jain, G. J. Kapral, R. W. Grosse Kunstleve, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2011). "The Phenix software for automated determination of macromolecular structures." Methods **55**(1): 94-106.

Affranchino, J. L. and S. A. Gonzalez (2010). "*In vitro* assembly of the feline immunodeficiency virus Gag polyprotein." Virus Res **150**(1-2): 153-157.

Alexandre, C. and B. Verrier (1991). "Four regulatory elements in the human c-fos promoter mediate transactivation by HTLV-1 Tax protein." Oncogene **6**(4): 543-551.

Alfadhli, A., T. C. Dhenub, A. Still and E. Barklis (2005). "Analysis of human immunodeficiency virus type 1 Gag dimerization-induced assembly." J Virol **79**(23): 14498-14506.

Amarasinghe, G. K., R. N. De Guzman, R. B. Turner, K. J. Chancellor, Z. R. Wu and M. F. Summers (2000). "NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition." J Mol Biol **301**(2): 491-511.

Araujo, A. and W. W. Hall (2004). "Human T-lymphotropic virus type II and neurological disease." Ann Neurol **56**(1): 10-19.

B

Bartman, M. T., Z. Kaidarova, D. Hirschkorn, R. A. Sacher, J. Fridey, G. Garratty, J. Gibble, J. W. Smith, B. Newman, A. E. Yeo and E. L. Murphy (2008). "Long-term increases in lymphocytes and platelets in human T-lymphotropic virus type II infection." Blood **112**(10): 3995-4002.

Basbous, J., A. Bazarbachi, C. Granier, C. Devaux and J. M. Mesnard (2003). "The central region of human T-cell leukemia virus type 1 Tax protein contains distinct domains involved in subunit dimerization." J Virol **77**(24): 13028-13035.

Bazarbachi, A., Y. Plumelle, J. Carlos Ramos, P. Tortevoye, Z. Otrock, G. Taylor, A. Gessain, W. Harrington, G. Panelatti and O. Hermine (2010). "Meta-analysis on the use of zidovudine and interferon-alfa in adult T-cell leukemia/lymphoma showing improved survival in the leukemic subtypes." J Clin Oncol **28**(27): 4177-4183.

Bendinelli, M., M. Pistello, S. Lombardi, A. Poli, C. Garzelli, D. Matteucci, L. Ceccherini-Nelli, G. Malvaldi and F. Tozzini (1995). "Feline immunodeficiency virus: an interesting model for AIDS studies and an important cat pathogen." Clin Microbiol Rev **8**(1): 87-112.

Bertazzoni, U., M. Turci, F. Avesani, G. Di Gennaro, C. Bidoia and M. G. Romanelli (2011). "Intracellular localization and cellular factors interaction of HTLV-1 and HTLV-2 Tax proteins: similarities and functional differences." Viruses **3**(5): 541-560.

Bisset, L. R., H. Lutz, J. Boni, R. Hofmann-Lehmann, R. Luthy and J. Schupbach (2002). "Combined effect of zidovudine (ZDV), lamivudine (3TC) and abacavir (ABC) antiretroviral therapy in suppressing *in vitro* FIV replication." Antiviral Res **53**(1): 35-45.

Bocanegra, R., M. A. Fuertes, A. Rodriguez-Huete, J. L. Neira and M. G. Mateu (2015). "Biophysical analysis of the MHR motif in folding and domain swapping of the HIV capsid protein C-terminal domain." Biophys J **108**(2): 338-349.

Boxus, M., J. C. Twizere, S. Legros, J. F. Dewulf, R. Kettmann and L. Willems (2008). "The HTLV-1 Tax interactome." Retrovirology **5**: 76.

Briggs, J. A., T. Wilk, R. Welker, H. G. Krausslich and S. D. Fuller (2003). "Structural organization of authentic, mature HIV-1 virions and cores." Embo j **22**(7): 1707-1715.

C

Cachat, A., S. A. Chevalier, S. Alais, N. L. Ko, L. Ratner, C. Journo, H. Dutartre and R. Mahieux (2013). "Alpha interferon restricts human T-lymphotropic virus type 1 and 2 de novo infection through PKR activation." J Virol **87**(24): 13386-13396.

Carrio, M. M. and A. Villaverde (2001). "Protein aggregation as bacterial inclusion bodies is reversible." FEBS Lett **489**(1): 29-33.

Chayen, N. E. and E. Saridakis (2008). "Protein crystallization: from purified protein to diffraction-quality crystal." Nat Methods **5**(2): 147-153.

Ciminale, V., F. Rende, U. Bertazzoni and M. G. Romanelli (2014). "HTLV-1 and HTLV-2: highly similar viruses with distinct oncogenic properties." Front Microbiol **5**: 398.

Coleman, J. K., R. Pu, M. M. Martin, E. N. Noon-Song, R. Zwijnenberg and J. K. Yamamoto (2014). "Feline immunodeficiency virus (FIV) vaccine efficacy and FIV neutralizing antibodies." Vaccine **32**(6): 746-754.

Cortines, J. R., E. B. Monroe, S. Kang and P. E. Prevelige (2011). "A retroviral chimeric capsid protein reveals the role of the N-terminal β-hairpin in mature core assembly." J Mol Biol **410**(4): 641-652.

Currer, R., R. Van Duyne, E. Jaworski, I. Guendel, G. Sampey, R. Das, A. Narayanan and F. Kashanchi (2012). "HTLV tax: a fascinating multifunctional co-regulator of viral and cellular pathways." Front Microbiol **3**: 406.

D

de Marco, A., A. M. Heuser, B. Glass, H. G. Kräusslich, B. Müller and J. A. G. Briggs (2012). "Role of the SP2 Domain and Its Proteolytic Cleavage in HIV-1 Structural Maturation and Infectivity." J Virol **86**(24): 13708-13716.

Dick, R. A. and V. M. Vogt (2014). "Membrane interaction of retroviral Gag proteins." Front Microbiol **5**: 187.

Dodon, M. D., Z. Li, S. Hamaia and L. Gazzolo (2004). "Tax protein of human T-cell leukaemia virus type 1 induces interleukin 17 gene expression in T cells." J Gen Virol **85**(Pt 7): 1921-1932.

Dorfman, T., A. Bukovsky, A. Ohagen, S. Hoglund and H. G. Gottlinger (1994). "Functional domains of the capsid protein of human immunodeficiency virus type 1." J Virol **68**(12): 8180-8187.

Du, S., L. Betts, R. Yang, H. Shi, J. Concel, J. Ahn, C. Aiken, P. Zhang and J. I. Yeh (2011). "Structure of the HIV-1 Full-Length Capsid in a Conformationally-Trapped Unassembled State Induced by Small-Molecule Binding." J Mol Biol **406**(3): 371-386.

Dubendorff, J. W. and F. W. Studier (1991). "Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor." J Mol Biol **219**(1): 45-59.

Dumon-Seignovert, L., G. Cariot and L. Vuillard (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**(1): 203-206.

Duong-Ly, K. C. and S. B. Gabelli (2014). "Salting out of proteins using ammonium sulfate precipitation." Methods Enzymol **541**: 85-94.

E

Elder, J. H., Y. C. Lin, E. Fink and C. K. Grant (2010). "Feline immunodeficiency virus (FIV) as a model for study of lentivirus infections: parallels with HIV." Curr HIV Res **8**(1): 73-80.

Elder, J. H., M. Schnolzer, C. S. Hasselkus-Light, M. Henson, D. A. Lerner, T. R. Phillips, P. C. Wagaman and S. B. Kent (1993). "Identification of proteolytic processing sites within the Gag and Pol polyproteins of feline immunodeficiency virus." J Virol **67**(4): 1869-1876.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

F

Fassati, A. (2012). "Multiple roles of the capsid protein in the early steps of HIV-1 infection." Virus Res **170**(1-2): 15-24.

Ferreira, O. C., Jr., J. D. Planelles V Fau - Rosenblatt and J. D. Rosenblatt (1997). "Human T-cell leukemia viruses: epidemiology, biology, and pathogenesis." (0268-960X (Print)).

Feuer, G. and P. L. Green (2005). "Comparative biology of human T-cell lymphotropic virus type 1 (HTLV-1) and HTLV-2." Oncogene **24**(39): 5996-6004.

Fiorini, F., S. Stefanini, P. Valenti, E. Chiancone and D. De Biase (2008). "Transcription of the Listeria monocytogenes fri gene is growth-phase dependent and is repressed directly by Fur, the ferric uptake regulator." Gene **410**(1): 113-121.

Fossen, T., V. Wray, K. Bruns, J. Rachmat, P. Henklein, U. Tessmer, A. Maczurek, P. Klinger and U. Schubert (2005). "Solution structure of the human immunodeficiency virus type 1 p6 protein." J Biol Chem **280**(52): 42515-42527.

Franke, D. and D. I. Svergun (2009). "DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering." J Appl Crystallogr **42**(Pt 2): 342-346.

Frankel, A. D. and J. A. Young (1998). "HIV-1: fifteen proteins and an RNA." Annu Rev Biochem **67**: 1-25.

Freed, E. O. (1998). "HIV-1 gag proteins: diverse functions in the virus life cycle." Virology **251**(1): 1-15.

Freed, E. O. (2015). "HIV-1 assembly, release and maturation." Nat Rev Microbiol **13**(8): 484-496.

Freed, E. O. and A. J. Mouland (2006). "The cell biology of HIV-1 and other retroviruses." Retrovirology **3**: 77.

G

Gamble, T. R., S. Yoo, F. F. Vajdos, U. K. von Schwedler, D. K. Worthylake, H. Wang, J. P. McCutcheon, W. I. Sundquist and C. P. Hill (1997). "Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein." Science **278**(5339): 849-853.

Gessain, A. (2011). "[Human retrovirus HTLV-1: descriptive and molecular epidemiology, origin, evolution, diagnosis and associated diseases]." Bull Soc Pathol Exot **104**(3): 167-180.

Gessain, A., F. Barin, J. C. Vernant, O. Gout, L. Maurs, A. Calender and G. de The (1985). "Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis." Lancet **2**(8452): 407-410.

Gessain, A. and O. Cassar (2012). "Epidemiological Aspects and World Distribution of HTLV-1 Infection." Front Microbiol **3**: 388.

Gessain, A. and R. Mahieux (2012). "Tropical spastic paraparesis and HTLV-1 associated myelopathy: clinical, epidemiological, virological and therapeutic aspects." Rev Neurol (Paris) **168**(3): 257-269.

Goldstone, D. C., M. W. Yap, L. E. Robertson, L. F. Haire, W. R. Taylor, A. Katzourakis, J. P. Stoye and I. A. Taylor (2010). "Structural and functional analysis of prehistoric lentiviruses uncovers an ancient molecular interface." Cell Host Microbe **8**(3): 248-259.

Gomez, N. V., A. Fontanals, V. Castillo, M. A. Gisbert, A. Suraniti, G. Mira and P. B. Pisano (2012). "Evaluation of different antiretroviral drug protocols on naturally infected feline immunodeficiency virus (FIV) cats in the late phase of the asymptomatic stage of infection." Viruses **4**(6): 924-939.

Goncalves, D. U., F. A. Proietti, J. G. Ribas, M. G. Araujo, S. R. Pinheiro, A. C. Guedes and A. B. Carneiro-Proietti (2010). "Epidemiology, treatment, and prevention of human T-cell leukemia virus type 1-associated diseases." Clin Microbiol Rev **23**(3): 577-589.

Goren, I., O. J. Semmes, K. T. Jeang and K. Moelling (1995). "The amino terminus of Tax is required for interaction with the cyclic AMP response element binding protein." J Virol **69**(9): 5806-5811.

Gres, A. T., K. A. Kirby, V. N. KewalRamani, J. J. Tanner, O. Pornillos and S. G. Sarafianos (2015). "X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability." Science **349**(6243): 99-103.

H

Haoudi, A., R. C. Daniels, E. Wong, G. Kupfer and O. J. Semmes (2003). "Human T-cell leukemia virus-I tax oncoprotein functionally targets a subnuclear complex involved in cellular DNA damage-response." J Biol Chem **278**(39): 37736-37744.

Harrod, R., Y. Tang, C. Nicot, H. S. Lu, A. Vassilev, Y. Nakatani and C. Z. Giam (1998). "An exposed KID-like domain in human T-cell lymphotropic virus type 1 Tax is responsible for the recruitment of coactivators CBP/p300." Mol Cell Biol **18**(9): 5052-5061.

Hayward, J. J. and A. G. Rodrigo (2010). "Molecular epidemiology of feline immunodeficiency virus in the domestic cat (Felis catus)." Vet Immunol Immunopathol **134**(1-2): 68-74.

Higuchi, M. and M. Fujii (2009). "Distinct functions of HTLV-1 Tax1 from HTLV-2 Tax2 contribute key roles to viral pathogenesis." Retrovirology **6**: 117.

Higuchi, M., C. Tsubata, R. Kondo, S. Yoshida, M. Takahashi, M. Oie, Y. Tanaka, R. Mahieux, M. Matsuoka and M. Fujii (2007). "Cooperation of NF-kappaB2/p100 activation and the PDZ domain binding motif signal in human T-cell leukemia virus type 1 (HTLV-1) Tax1 but not HTLV-2 Tax2 is crucial for interleukin-2-independent growth transformation of a T-cell line." J Virol **81**(21): 11900-11907.

Hill, C. P., D. Worthylake, D. P. Bancroft, A. M. Christensen and W. I. Sundquist (1996). "Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly." Proc Natl Acad Sci U S A **93**(7): 3099-3104.

Hizi, A. and E. Herzig (2015). "dUTPase: the frequently overlooked enzyme encoded by many retroviruses." Retrovirology **12**: 70.

Hosie, M. J. and J. A. Beatty (2007). "Vaccine protection against feline immunodeficiency virus: setting the challenge." Aust Vet J **85**(1-2): 5-12; quiz 85.

I

Iwanaga, M., T. Watanabe and K. Yamaguchi (2012). "Adult T-cell leukemia: a review of epidemiological evidence." Front Microbiol **3**: 322.

J

Jacques, D. A., W. A. McEwan, L. Hilditch, A. J. Price, G. J. Towers and L. C. James (2016). "HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis." Nature **536**(7616): 349-353.

Jeong, S. J., M. Radonovich, J. N. Brady and C. A. Pise-Masison (2004). "HTLV-I Tax induces a novel interaction between p65/RelA and p53 that results in inhibition of p53 transcriptional activity." Blood **104**(5): 1490-1497.

Jin, D. Y. and K. T. Jeang (1997). "HTLV-I Tax self-association in optimal trans-activation function." Nucleic Acids Res **25**(2): 379-387.

Jin, Z., L. Jin, D. L. Peterson and C. L. Lawson (1999). "Model for lentivirus capsid core assembly based on crystal dimers of EIAV p26." J Mol Biol **286**(1): 83-93.

K

Kabsch, W. (2010). "XDS." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-132.

Kalyanaraman, V. S., M. G. Sarngadharan, M. Robert-Guroff, I. Miyoshi, D. Golde and R. C. Gallo (1982). "A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia." Science **218**(4572): 571-573.

Keegan, R. M. and M. D. Winn (2008). "MrBUMP: an automated pipeline for molecular replacement." Acta Crystallogr D Biol Crystallogr **64**(Pt 1): 119-124.

Kenyon, J. C. and A. M. Lever (2011). "The molecular biology of feline immunodeficiency virus (FIV)." Viruses **3**(11): 2192-2213.

Khwaja, A., M. Galilee, A. Marx and A. Alian (2016). "Structure of FIV capsid C-terminal domain demonstrates lentiviral evasion of genetic fragility by coevolved substitutions." Sci Rep **6**: 24957.

King, A. M. Q., M. J. Adams, E. B. Carstens and E. J. Leftkowitz (2011). "Virus Taxonomy : Ninth Report of the International Committe on Taxonomy of viruses."

Kinpara, S., M. Kijiyama, A. Takamori, A. Hasegawa, A. Sasada, T. Masuda, Y. Tanaka, A. Utsunomiya and M. Kannagi (2013). "Interferon-alpha (IFN-alpha) suppresses HTLV-1 gene expression and cell cycling, while IFN-alpha combined with zidovudine induces p53 signaling and apoptosis in HTLV-1-infected cells." Retrovirology **10**: 52.

Koch, M. H. J. and D. Svergun (2003). "Reports on Progression in Physics." **66**.

Konno, M., M. Ito, T. Hayano and N. Takahashi (1996). "The substrate-binding site in *Escherichia coli* cyclophilin A preferably recognizes a cis-proline isomer or a highly distorted form of the trans isomer." J Mol Biol **256**(5): 897-908.

Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." J Mol Biol **372**(3): 774-797.

L

Lairmore, M. D., R. Haines and R. Anupam (2012). "Mechanisms of human T-lymphotropic virus type 1 transmission and disease." Curr Opin Virol **2**(4): 474-481.

Lamed, R., R. Kenig, E. Morag, S. Yaron, Y. Shoham and E. A. Bayer (2001). "Nonproteolytic cleavage of aspartyl proline bonds in the cellulosomal scaffoldin subunit from Clostridium thermocellum." Appl Biochem Biotechnol **90**(1): 67-73.

Lee, R. and R. A. Schwartz (2011). "Human T-lymphotrophic virus type 1-associated infective dermatitis: a comprehensive review." J Am Acad Dermatol **64**(1): 152-160.

Lee, S. K., M. Potempa and R. Swanstrom (2012). "The choreography of HIV-1 proteolytic processing and virion assembly." J Biol Chem **287**(49): 40867-40874.

Lin, T. Y. and M. Emerman (2006). "Cyclophilin A interacts with diverse lentiviral capsids." Retrovirology **3**: 70.

Liu, S., L. Zhou, L. Chen, S. G. Dastidar, C. Verma, J. Li, D. Tan and R. Beuerman (2009). "Effect of structural parameters of peptides on dimer formation and highly oxidized side products in the oxidation of thiols of linear analogues of human beta-defensin 3 by DMSO." J Pept Sci **15**(2): 95-106.

Luttge, B. G. and E. O. Freed (2010). "FIV Gag: virus assembly and host-cell interactions." Vet Immunol Immunopathol **134**(1-2): 3-13.

Luttge, B. G., M. Shehu-Xhilaga, D. G. Demirov, C. S. Adamson, F. Soheilian, K. Nagashima, A. G. Stephen, R. J. Fisher and E. O. Freed (2008). "Molecular characterization of feline immunodeficiency virus budding." J Virol **82**(5): 2106-2119.

M

Mahieux, R. and A. Gessain (2011). "HTLV-3/STLV-3 and HTLV-4 viruses: discovery, epidemiology, serology and molecular aspects." Viruses **3**(7): 1074-1090.

Mailler, E., S. Bernacchi, R. Marquet, J. C. Paillart, V. Vivet-Boudou and R. P. Smyth (2016). "The Life-Cycle of the HIV-1 Gag-RNA Complex." Viruses **8**(9).

Mammano, F., A. Ohagen, S. Hoglund and H. G. Gottlinger (1994). "Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis." J Virol **68**(8): 4927-4936.

Matsuda, T., T. Kigawa, S. Koshiba, M. Inoue, M. Aoki, K. Yamasaki, M. Seki, K. Shinozaki and S. Yokoyama (2006). "Cell-free synthesis of zinc-binding proteins." J Struct Funct Genomics **7**(2): 93-100.

Matsuoka, M. and K. T. Jeang (2007). "Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation." Nat Rev Cancer **7**(4): 270-280.

Meertens, L., S. Chevalier, R. Weil, A. Gessain and R. Mahieux (2004). "A 10-amino acid domain within human T-cell leukemia virus type 1 and type 2 tax protein sequences is responsible for their divergent subcellular distribution." J Biol Chem **279**(41): 43307-43320.

Momany, C., L. C. Kovari, A. J. Prongay, W. Keller, R. K. Gitti, B. M. Lee, A. E. Gorbalenya, L. Tong, J. McClure, L. S. Ehrlich, M. F. Summers, C. Carter and M. G. Rossmann (1996). "Crystal structure of dimeric HIV-1 capsid protein." Nat Struct Biol **3**(9): 763-770.

Moraillon, A. (2000). Rétroviroses felines Encyclopédie Vétérinaire, Médecine Générale. **1500:** 1-9.

N

Nath, M. D. and D. L. Peterson (2001). "*In vitro* assembly of feline immunodeficiency virus capsid protein: biological role of conserved cysteines." Arch Biochem Biophys **392**(2): 287-294.

Nguyen, T. T., R. F. Bruinsma and W. M. Gelbart (2006). "Continuum theory of retroviral capsids." Phys Rev Lett **96**(7): 078102.

P

Pedersen, N. C., E. W. Ho, M. L. Brown and J. K. Yamamoto (1987). "Isolation of a T-lymphotropic virus from domestic cats with an immunodeficiency-like syndrome." Science **235**(4790): 790-793.

Percher, F., P. Jeannin, S. Martin-Latil, A. Gessain, P. V. Afonso, A. Vidy-Roche and P. E. Ceccaldi (2016). "Mother-to-Child Transmission of HTLV-1 Epidemiological Aspects, Mechanisms and Determinants of Mother-to-Child Transmission." Viruses **8**(2).

Perilla, J. R. and A. M. Gronenborn (2016). "Molecular Architecture of the Retroviral Capsid." Trends Biochem Sci **41**(5): 410-420.

Perkins, N. D. (2007). "Integrating cell-signalling pathways with NF-kappaB and IKK function." Nat Rev Mol Cell Biol **8**(1): 49-62.

Petoukhov, M. V., D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev and D. I. Svergun (2012). "New developments in the ATSAS program package for small-angle scattering data analysis." J Appl Crystallogr **45**(Pt 2): 342-350.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera - a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-1612.

Poiesz, B. J., F. W. Ruscetti, A. F. Gazdar, P. A. Bunn, J. D. Minna and R. C. Gallo (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**(12): 7415-7419.

Pornillos, O., B. K. Ganser-Pornillos, S. Banumathi, Y. Hua and M. Yeager (2010). "Disulfide bond stabilization of the hexameric capsomer of human immunodeficiency virus." J Mol Biol **401**(5): 985-995.

Pornillos, O., B. K. Ganser-Pornillos and M. Yeager (2011). "Atomic-level modelling of the HIV capsid." Nature **469**(7330): 424-427.

Putnam, C. D., M. Hammel, G. L. Hura and J. A. Tainer (2007). "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution." Q Rev Biophys **40**(3): 191-285.

Q

Qi, M., R. Yang and C. Aiken (2008). "Cyclophilin A-dependent restriction of human immunodeficiency virus type 1 capsid mutants for infection of nondividing cells." J Virol **82**(24): 12001-12008.

Qi, X., Y. Sun and S. Xiong (2015). "A single freeze-thawing cycle for highly efficient solubilization of inclusion body proteins and its refolding into bioactive form." Microb Cell Fact **14**: 24.

R

Rende, F., I. Cavallari, M. G. Romanelli, E. Diani, U. Bertazzoni and V. Ciminale (2012). "Comparison of the Genetic Organization, Expression Strategies and Oncogenic Potential of HTLV-1 and HTLV-2." Leuk Res Treatment **2012**: 876153.

Romanelli, M. G., E. Diani, E. Bergamo, C. Casoli, V. Ciminale, F. Bex and U. Bertazzoni (2013). "Highlights on distinctive structural and functional properties of HTLV Tax proteins." Front Microbiol **4**: 271.

Rossmann, M. G. (1990). "The molecular replacement method." Acta Crystallogr A **46 (Pt 2)**: 73-82.

Roucoux, D. F. and E. L. Murphy (2004). "The epidemiology and disease outcomes of human T-lymphotropic virus type II." AIDS Rev **6**(3): 144-154.

Roucoux, D. F., B. Wang, D. Smith, C. C. Nass, J. Smith, S. T. Hutching, B. Newman, T. H. Lee, D. M. Chafets and E. L. Murphy (2005). "A prospective study of sexual transmission of human T lymphotropic virus (HTLV)-I and HTLV-II." J Infect Dis **191**(9): 1490-1497.

S

Serriere, J., D. Fenel, G. Schoehn, P. Gouet and C. Guillon (2013). "Biophysical characterization of the feline immunodeficiency virus p24 capsid protein conformation and *in vitro* capsid assembly." PLoS One **8**(2): e56424.

Serriere, J., X. Robert, M. Perez, P. Gouet and C. Guillon (2013). "Biophysical characterization and crystal structure of the Feline Immunodeficiency Virus p15 matrix protein." Retrovirology **10**: 64.

Shirinian, M., Y. Kfoury, Z. Dassouki, H. El-Hajj and A. Bazarbachi (2013). "Tax-1 and Tax-2 similarities and differences: focus on post-translational modifications and NF-kappaB activation." Front Microbiol **4**: 231.

Spearman, P. (2016). "HIV-1 Gag as an Antiviral Target: Development of Assembly and Maturation Inhibitors." Curr Top Med Chem **16**(10): 1154-1166.

Studier, F. W. (1991). "Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system." J Mol Biol **219**(1): 37-44.

Studier, F. W. (2005). "Protein production by auto-induction in high density shaking cultures." Protein Expr Purif **41**(1): 207-234.

Sundquist, W. I. and H. G. Krausslich (2012). "HIV-1 assembly, budding, and maturation." Cold Spring Harb Perspect Med **2**(7): a006924.

Suzuki, T., J. I. Fujisawa, M. Toita and M. Yoshida (1993). "The trans-activator tax of human T-cell leukemia virus type 1 (HTLV-1) interacts with cAMP-responsive element (CRE) binding and CRE modulator proteins that bind to the 21-base-pair enhancer of HTLV-1." Proc Natl Acad Sci U S A **90**(2): 610-614.

Suzuki, T., Y. Ohsugi, M. Uchida-Toita, T. Akiyama and M. Yoshida (1999). "Tax oncoprotein of HTLV-1 binds to the human homologue of Drosophila discs large tumor suppressor protein, hDLG, and perturbs its function in cell growth control." Oncogene **18**(44): 5967-5972.

Svergun, D. I., C. Barberato and M. H. J. Koch (1995). "CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordonates." J. Appl. Cryst. **28**: 768-773.

T

Takahashi, T., H. Takase, T. Urano, S. Sugita, K. Miyata, N. Miyata and M. Mochizuki (2000). "Clinical features of human T-lymphotropic virus type 1 uveitis: a long-term follow-up." Ocul Immunol Inflamm **8**(4): 235-241.

Tang, C., Y. Ndassa and M. F. Summers (2002). "Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein." Nat Struct Biol **9**(7): 537-543.

Thenin-Houssier, S. and S. T. Valente (2016). "HIV-1 Capsid Inhibitors as Antiretroviral Agents." Curr HIV Res **14**(3): 270-282.

Tie, F., N. Adya, W. C. Greene and C. Z. Giam (1996). "Interaction of the human T-lymphotropic virus type 1 Tax dimer with CREB and the viral 21-base-pair repeat." J Virol **70**(12): 8368-8374.

Trono, D. (1995). "HIV accessory proteins: leading roles for the supporting cast." Cell **82**(2): 189-192.

Troyer, J. L., J. Pecon-Slattery, M. E. Roelke, W. Johnson, S. VandeWoude, N. Vazquez-Salat, M. Brown, L. Frank, R. Woodroffe, C. Winterbach, H. Winterbach, G. Hemson, M. Bush, K. A. Alexander, E. Revilla and S. J. O'Brien (2005). "Seroprevalence and genomic divergence of circulating strains of feline immunodeficiency virus among Felidae and Hyaenidae species." J Virol **79**(13): 8282-8294.

Turci, M., J. Lodewick, G. Di Gennaro, A. S. Rinaldi, O. Marin, E. Diani, C. Sampaio, F. Bex, U. Bertazzoni and M. G. Romanelli (2012). "Ubiquitination and sumoylation of the HTLV-2 Tax-2B protein regulate its NF-kappaB activity: a comparative study with the HTLV-1 Tax-1 protein." Retrovirology **9**: 102.

Turci, M., J. Lodewick, P. Righi, A. Polania, M. G. Romanelli, F. Bex and U. Bertazzoni (2009). "HTLV-2B Tax oncoprotein is modified by ubiquitination and sumoylation and displays intracellular localization similar to its homologue HTLV-1 Tax." Virology **386**(1): 6-11.

U

Uhl, E. W., T. G. Heaton-Jones, R. Pu and J. K. Yamamoto (2002). "FIV vaccine development and its importance to veterinary and human medicine: a review FIV vaccine 2002 update and review." Vet Immunol Immunopathol **90**(3-4): 113-132.

V

Vallejo, L. F. and U. Rinas (2004). "Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins." Microb Cell Fact **3**(1): 11.

Verdonck, K., E. Gonzalez, S. Van Dooren, A. M. Vandamme, G. Vanham and E. Gotuzzo (2007). "Human T-lymphotropic virus 1: recent knowledge about an ancient infection." Lancet Infect Dis **7**(4): 266-281.

Vogt, V. M. (1997). Retroviral Virions and Genomes. Retroviruses. J. M. Coffin, S. H. Hughes and H. E. Varmus. Cold Spring Harbor (NY), Cold Spring Harbor Laboratory Press
Cold Spring Harbor Laboratory Press.
Volkov, V. V. and D. I. Svergun (2003). "Uniqueness of *ab initio* shape determination in small-angle scattering." Journal of Applied Crystallography **36**(3 Part 1): 860-864.

W

Wiktor, S. Z., E. J. Pate, P. S. Rosenberg, M. Barnett, P. Palmer, D. Medeiros, E. M. Maloney and W. A. Blattner (1997). "Mother-to-child transmission of human T-cell lymphotropic virus type I associated with prolonged breast-feeding." J Hum Virol **1**(1): 37-44.

Wilen, C. B., J. C. Tilton and R. W. Doms (2012). "Molecular mechanisms of HIV entry." Adv Exp Med Biol **726**: 223-242.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-242.

Wright, E. R., J. B. Schooler, H. J. Ding, C. Kieffer, C. Fillmore, W. I. Sundquist and G. J. Jensen (2007). "Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells." Embo j **26**(8): 2218-2226.

Y

Yamamoto, J. K., R. Pu, E. Sato and T. Hohdatsu (2007). "Feline immunodeficiency virus pathogenesis and development of a dual-subtype feline-immunodeficiency-virus vaccine." Aids **21**(5): 547-563.

Yamamoto, J. K., E. Sparger, E. W. Ho, P. R. Andersen, T. P. O'Connor, C. P. Mandell, L. Lowenstine, R. Munn and N. C. Pedersen (1988). "Pathogenesis of experimentally induced feline immunodeficiency virus infection in cats." Am J Vet Res **49**(8): 1246-1258.

Yamano, Y. and T. Sato (2012). "Clinical pathophysiology of human T-lymphotropic virus-type 1-associated myelopathy/tropical spastic paraparesis." Front Microbiol **3**: 389.

Yoshida, M., M. Seiki, K. Yamaguchi and K. Takatsuki (1984). "Monoclonal integration of human T-cell leukemia provirus in all primary tumors of adult T-cell leukemia suggests causative role of human T-cell leukemia virus in the disease." Proc Natl Acad Sci U S A **81**(8): 2534-2537.

Z

Zhao, G., J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken and P. Zhang (2013). "Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics." Nature **497**(7451): 643-646.

PUBLICATIONS

Send Orders for Reprints to reprints@benthamscience.ae

Protein & Peptide Letters, **2015**, *22*, 0000-0000 1

A Staggered Decameric Assembly of Human C-Reactive Protein Stabilized by Zinc Ions Revealed by X-ray Crystallography

Christophe Guillon[a,*], Ulrick Mavoungou Bigouagou[b,c], Christelle Folio[a], Pascale Jeannin[b,c], Yves Delneste[b,c] and Patrice Gouet[a]

[a]*Biocrystallography and Structural Biology of TherapeuticTargets, UMR 5086 CNRS Université de Lyon, IBCP, CNRS, Université de Lyon, Lyon, France;* [b]*Université d'Angers, Angers, France; Inserm, UMR 892, Angers, France; CNRS, UMR 6299, Angers, France ;* [c]*CHU d'Angers, Laboratoire d'Immunologie et d'Allergologie, Angers, France*



Abstract: Human C-reactive protein (CRP) is an acute phase protein, which harbours both host defence and scavenging properties. In this study, we obtained two new crystal forms of CRP, where CRP forms a symmetric, staggered dimer of pentamers. In one of these structures, obtained in the presence of HIV-1 Tat protein, this dimer of pentamers is stabilized by two zinc ions trapped within a cleft of the effector face of CRP. These two decameric interfaces involve complementary surfaces of CRP pentamers and bury a large area of ~2000 Å^2 per pentamer, suggesting a biological role of this interface. These two novel decameric interfaces and the involvement of zinc might have important consequences in the understanding of CRP biological functions.

Keywords: C-reactive protein, CRP, decamer, pentraxin, structure, X-ray crystallography, zinc, acute phase, innate immunity.

1. INTRODUCTION

C-reactive protein (CRP) is a protein which naturally circulates in the plasma. CRP is a member of the pentraxin family, which also includes Serum amyloid P (SAP) and pentraxin-3 (PTX3/TSG-14). Pentraxins are characterized by a molecular signature HxCxS/TWxS, in which x represents any amino acid, and can be distinguished in two classes depending on the length of their N-terminal domain: short pentraxins (CRP, SAP) and long pentraxins (PTX3 and PTX4) [1-3]. Pentraxins in general, and the CRP protein in particular, are highly conserved in phylogenetically divergent animal species, from horse-shoe crabs (*Limulus limulus*) or xenopus (*Xenopus laevi*) to human [4].

In humans, as an acute phase protein, the concentration of CRP in serum rises dramatically by a 1000x- to 10000x-fold in response to tissue injury, bacterial infection, or inflammation [5, 6]. CRP is able to interact in a calcium-dependent manner with phosphocholine groups present in bacterial membranes, with fungi such as *Aspergillus fumigatus* or with yeast [1, 2, 7]. The binding of CRP leads to the clearance of microorganisms, either by phagocytosis or by the activation of the complement pathway. CRP can also interact with highly negatively charged molecules, such as nucleoproteins, and might be involved in the elimination of apoptotic bodies, a critical process to maintain tissue and immune system homeostasis and to avoid autoimmune response [1, 6, 8, 9]. Besides these protective effects, increased CRP levels in serum are also used as a prognosis marker for coronary diseases [10].

CRP is a 206 residues-long protein containing an intrachain disulfide bond and a calcium-binding site. The X-ray structure of CRP has been determined in several conditions: it harbours a lectin-like jellyroll fold that forms a pentameric ring with a 5-fold molecular symmetry through salt-bridge interactions within monomers (PDB ID: 1B09, 1GNH, 1LJ7 [11-13]). This pentamer presents itself as a disk of ~100 Å in diameter and 35 Å in depth, presenting two faces. The first face contains the calcium-binding site, which binds the bacterial or apoptotic cell ligand (dubbed the "B-face" for "binding face"). The second face (the effector "A-face" containing the single α-helix of CRP) is involved in the activation and regulation of the complement pathway and/or Fcɣ receptors, thus mediating ligand elimination via the complement cascade or opsonisation [12, 14-18]. Although the recognition of Fcγ receptors seems to involve pentameric pentraxins [17], some studies have suggested that the ability of CRP to activate the complement pathway involves a denatured monomeric form of the CRP protein [19, 20]. However, the monomerization of CRP from a pentamer is achievable *in vitro* through non physiological conditions including harsh chemical treatment, immobilization onto polystyrene, or removal of calcium ions [18, 19, 21] and monomers tend to aggregate rapidly [10]. By contrast, only the monomeric form of CRP has been described to interact with endothelial cells [22] and to induce prothrombic activities [23], suggesting that monomeric instead of pentameric CRP is involved in coronary syndromes [10]. Other reports have described that CRP exists in solution as an equilibrium between pentamers and decamers in physiological conditions, where the de-

*Address correpondence to this author at the "Biocrystallography and Structural Biology of Therapeutic Targets", UMR 5086 CNRS Université de Lyon, IBCP 7, passage du Vercors, 69367 Lyon cedex 7, France; Tel: (+33) 4 37 65 2904; Fax: (+33) 4 72 72 2631; E-mail: christophe.guillon@ibcp.fr

0929-8665/15 $58.00+.00 © 2015 Bentham Science Publishers

camer of CRP is supposedly formed by the dimerization of two CRP pentamers through interaction between their effector A-face [18]. Altogether, although the respective importance of monomeric, pentameric and decameric forms of CRP on its functions is still debated, changes in CRP oligomerization might be involved in the regulation of its function.

We have recently obtained evidence for a specific interaction of CRP with the HIV regulatory protein Tat (Mavoungou Bigouagou *et al.*, in preparation). Therefore, we performed crystallization of CRP in the presence of Tat to understand the structural basis of this interaction. In these conditions, a new crystal form was obtained, in which a CRP decamer was observed but Tat was absent in the electron density maps. This decamer of CRP, which has never been observed before, was formed by the interaction of two CRP pentamers via their A-face and involved two zinc ions in a cleft of CRP at the A-face involved in the interaction with the C1q complement protein. In the absence of Tat, the same decamer of CRP was observed in the same crystallization condition, but no zinc ions were present. This pentamer-pentamer interface of CRP and the possible role of zinc to stabilize this CRP decamer might have important biological functions.

2. MATERIAL-METHODS

2.1. Proteins and Crystallisation Conditions

The Tat protein (101 residues, 11.5kDa), derived from the sequence of primary isolate 133A (Genbank accession number DQ459120 [24]) was obtained by chemical synthesis using Fmoc chemistry, and has been previously characterized structurally, biochemically and functionally [25-27]. Tat protein was resuspended at 10 $mg.ml^{-1}$ in distilled water prior to use. The CRP protein is a recombinant protein expressed in *Escherichia coli* and commercially available (Calbiochem) at 1 $mg.ml^{-1}$ in 140 mM NaCl, 20 mM Tris-HCl, 2 mM $CaCl_2$, 0.05% NaN_3, pH 7.5. After removal of endotoxin using Profos Endotrap Red columns according to the manufacturer's protocol, CRP was concentrated at 8 $mg.ml^{-1}$ by ultracentrifugation for 10 min at 10000×g using a Vivaspin device (Millipore) with a 10 kDa cut-off. The concentration of the concentrated fraction was evaluated using a BCA protein assay (Pierce), according to manufacturer's protocol.

In an attempt to co-crystallize HIV Tat and CRP, 15.6 µl of CRP (8 $mg.ml^{-1}$ in 140 mM NaCl, 20 mM Tris-HCl, 2 mM $CaCl_2$, 0.05% NaN_3, pH 7.5) were mixed to 1 µl of Tat (10 $mg.ml^{-1}$ in distilled water) and 0.3 µl $CaCl_2$ 100 mM, incubated for 20 min at room temperature, before adding 5 µl of PBS 1x (Euromedex). In these conditions where the Tat/CRP complex is active (Mavoungou Bigouagou *et al*, in preparation), the initial molar ratio of CRP:Tat in solution was thus 5:1. The screening of crystallization conditions was performed using a Mosquito nanopipetter in 96-well plates, using the sitting-drop vapour diffusion method with 0.4 µl drops of a 1:1 v:v protein:well solution and commercial screens from Qiagen (PEG I and PEG II suites, Classic lite). A hit was obtained for condition 88 of the Classic lite suite (0.1 M Tris-HCl pH 8.0, 15% (w:v) PEG4000, 0.2 M $MgCl_2$) and was further optimized using additive screens (Hampton Research). At the end, crystals of CRP in the presence of Tat were obtained using the hanging-drop vapour-diffusion method, with drops of 4 µl made of a 1:1 ratio (v:v) of protein:well solution and equilibrated against the well solution (0.1 M Tris-HCl pH 8.0, 12% (w:v) PEG4000, 0.2 M $MgCl_2$, 50 mM LiCl) at 289 K. Platelet-shaped crystals of 100×50×5 µm were repeatedly obtained in these conditions within two to four weeks. These crystals belong to the triclinic space groups P1 with cell dimensions a = 79.3 Å, b = 96.4 Å, c = 160.4 Å α= 79.9° β= 77.1° γ= 69.4°. As a control, we attempted to crystallize CRP alone. CRP (8 $mg.ml^{-1}$), incubated for 20 minutes with $CaCl_2$, as detailed above, before addition of 1× PBS, was used for crystallization experiments in the same conditions. Crystals grew in approximately nine months and belong to the monoclinic space group $P2_1$ with cell dimensions a = 97.6 Å, b = 143.2 Å, c = 161.3 Å α=β= 90° γ= 89.7°. For data collection, both crystal types were plunged in reservoir solution supplemented with 10% (v:v) ethylene glycol before being snap frozen at 100 K in liquid nitrogen.

2.2. Data Collection and Processing

X-ray diffraction data were collected at the European Synchrotron Radiation Facility (Grenoble, France) at beamlines FIP-BM30A (CRP in the absence of Tat) and ID23-2 (CRP in the presence of Tat) (Table **1**). All collected data were processed using XDS [28]. Molecular replacement was performed for the two crystal forms using the program AMoRe [29] and CRP pentamer as the search model (PDB ID 1LJ7 [13]). The triclinic form was refined first to 2 Å resolution with four pentamers in the asymmetric unit. Crystallographic refinement was performed with Refmac and water molecules were added using Arp/Warp from the CCP4i suite [30]. Manual corrections were performed with the graphics software package TURBO-FRODO. The monoclinic form was determined in turn to 3 Å resolution with four pentamers in the asymmetric unit. Crystallographic refinement was performed with RefMac, using non crystallographic symmetry (NCS) restrains with tight restraints and chain A as a reference. Both crystals showed similar solvent contents of ~45%. The stereochemistry of the structures was analysed using the Rampage program of the CCP4i suite. As the triclinic form of CRP showed unassigned residual electron density, presence of ions in the crystal was suspected. X-Ray fluorescence spectra were measured for the crystal or individual solutions using the Rontec fluorescence detector and the XNemo suite of FIP-BM30A beamline at ESRF. Characteristic Zn emission lines were observed. New X-ray diffraction data were collected from a triclinic crystal near the Zn absorption K-edge (1.283 Å) and anomalous difference Fourier maps were calculated using the CNS software [31], allowing the unambiguous location of Zn ions in the structure. Refinement of these zinc ions was performed without any restraints of the bond distances with the surrounding atoms using Refmac. Centromer distances between pentamers were calculated using LSQKAB. Buried areas and stability of the decameric interface were calculated using the protein interfaces, surfaces and assemblies service PISA at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) [32]. Figures were generated using the PyMol software from Schrödinger.

Table 1. Refinement statistics.

PDB ID	3PVO	3PVN
Protein	CRP	CRP/Tat
Beam line	FIP-BM30A	ID23-2
Wavelength (Å)	0.97969	0.87260
Data-collection statistics		
Space group	$P2_1$	P1
Unit cell parameters	a=97.6 b=143.2 c=161.2 α=β=90.0 γ=89.7	a= 79.3 b= 96.4 c= 160.4 α= 79.9 β= 77.1 γ= 69.4
Data resolution	20-3Å (3.08-3Å)	20-2Å (2.05-2Å)
Completeness (%)	95.0 (92.6)	92.8 (94)
No. of unique reflections	86781 (12015)	275048 (19869)
Multiplicity	1.96	1.85
I/ΣI	6.7 (2.0)	9.7 (3.0)
R_{sym}	0.153	0.062
% solvent	45.4%	44.7%
Refinement statistics		
Work/test reflections	82032/4380	261507/13541
R_{work}/R_{free} (%)	22.6/28.5	17.6/23.5
Asymmetric unit content aminoacids water molecules Ca^{2+} ions Zn^{2+} ions	4 pentamers 4120 250 40 0	4 pentamers 4120 3012 40 5
R.m.s.d. bond length (Å)	0.022	0.024
R.m.s.d. bond angles °	1.70	1.94
Ramachandran plot		
Favoured region Allowed region Outlier region	3906 (95.7%) 154 (3.8%) 20 (0.5%)	3946 (96.7%) 122 (3.0%) 12 (0.3%)

2.3. Accession Numbers

The coordinates and the structure factors are deposited in the Protein Data Bank (http://www.rcsb.org) with accession numbers **3PVO** and **3PVN**. Statistics of the X-ray data for all crystals are provided in Table **1**.

3. RESULTS

The crystal cell of the 2.0 Å triclinic structure obtained in the presence of Tat contains 20 monomers of CRP organized in four pentamers: P1 (chains A to E), P2 (chains F to J), P3 (chains K to O) and P4 (chains P to T) (Fig. **1A** & **1B**). Tat is not visible in the electron density map. The structure of each CRP monomer, as well as the arrangement of the monomers within a pentamer, are very similar to those from the literature [11, 12] as structures can be superimposed with r.m.s.d of 0.4 Å on all Cα atoms. As previously described, the calcium-binding sites of the B-face of each monomer within the pentamer contain two calcium ions [12]. To facilitate the comparison of our results with these published structures, we conserved the numbering of the initial study [11] although it did not take into account the first 18 residues of CRP.

In the asymmetric unit/crystal cell, CRP pentamers are associated as symmetrical dimers of pentamers, resulting in the formation of two decamers (P1 and P4, P2 and P3) (Fig. **1A** & **1B**). The decamer results from an intermolecular interaction of the two pentamers by the A-face, opposite to the

calcium binding face, via a 2-fold non crystallographic symmetry axis parallel to the interface but not coincident with the 5-fold molecular axis of the pentamers (Fig. **1A**). This leads to a dimer of staggered pentamers, in which the centers of the two pentamers are shifted ~35Å one from the other, and in which one monomer per pentamer (E, F, K and R for pentamer P1, P2, P3, and P4, respectively) does not participate to the interface (Fig. **1A**). This shift results in a tight interface of rather complementary structures (Fig. **1B**), as examplified by the total amount of buried area (~2000 Å^2 per pentamer) and the calculated stability of the dimer of pentamer (ΔG=-13.3 kcal/mol), predicting a biological unit according to PISA. Intermolecular interactions between pentamers to form the CRP decamer involve primarily the region 170-190 of the monomers, around the α-helix of CRP of the effector face, with equivalent determinants.

One monomer per pentamer (A, G, L and S for pentamers P1, P2, P3 and P4, respectively) shows an extra density located in a cleft at residue His38 on the A-face of the monomer (Fig. **1C**). X-ray fluorescence spectra were recorded in turn and Zn ions were identified for the crystal and individual protein solutions but not for the crystallization buffer (Fig. **2**). Subsequent calculations of anomalous difference Fourier maps clearly identified the observed extra density as a bound Zn ion (Fig. **1D**). This Zn(II) ion shows a distorted trigonal bipyramidal coordination that involves the imidazole group of the His38 of this cleft and four water molecules, with an average bond length of 2.1Å coherent with Zn (Fig. **1E**). On these four water molecules, three are bridging the Zn(II) ion to the main chain carbonyl of Ile174, Tyr175 and Pro206 of the same monomer, and one bridges the Zn(II) ion to the oxygen of the hydroxyl group of residue Ser181 from the facing pentamer (Fig. **1E**). Thus, two Zn(II) ions are participating to each decameric interface by bridging His38A to Ser181T and His38S to Ser181B in the decamer P1-P4 (Fig. **1A** and data not shown), His38G to Ser181M and His38L to Ser181H in the decamer P2-P3 (data not shown). Anomalous difference Fourier maps allowed us identifying another zinc ion, present in a single monomer in the asymmetric unit (Fig. **1F** and **G**). This zinc ion shows a tetrahedral coordination involving the residue Asp60 of chain T, one water molecule, and the Lys69 and Asp70 residues of the D chain of a crystallographic symmetrical pentamer, thus participating to the crystal packing (Fig. **1H**).

As a control, we evaluated the structure of CRP for crystals grown with the same protein concentration and in the same crystallization conditions but in the absence of Tat. As for the structure obtained from crystals grown in the presence of Tat, the structure of the monomer and of the individual pentamers were similar to those described in the literature [11, 12] and two calcium ions were observed in the calcium binding site of each monomer. However, in contrast with our data on the crystals grown in the presence of Tat, this crystal form shows a rod-like packing with pentamer-pentamer interactions between A-faces as well as between B-faces (Supplementary Fig. **1**). The pentamer-pentamer interactions through the B-faces are crystallographic contacts, as exemplified by the small amount of buried area (~460Å^2 per pentamer) and the positive free energy of the interface predicted by PISA (ΔG=2.1 kcal/mol), thus unfavorable for a biological dimer. On the opposite, the pentamer-pentamer interaction through the A-faces is similar to the decameric structure we observed earlier, with a r.m.s.d of 1.16Å on all Cα atoms of the decamers (data not shown). However, this interface does not involve any zinc ion, resulting in a decrease of the predicted stability of the decamer (ΔG of -8.4 kcal/mol).

4. DISCUSSION

The crystallization conditions in the absence or presence of Tat both led to the formation of decamers, formed by the interaction of two CRP pentamers through their A-face. CRP decamers have already been detected in trace amounts in the commercial solution we used in this study [33]. They have also been described for CRP in physiological conditions that are very close to the buffer we have used for CRP crystallization, and the formation of these decamers is dependent on the presence of calcium ions in the concentration range we have used [18, 34]. Moreover, the interaction of CRP with complement factor H, one of its interactants during the cascade of innate immunity reactions, seems to involve the decameric form of CRP [34]. Noteworthy, SAP, another short pentraxin, has also been described to be a pentameric protein which can associate as a dimer of staggered pentamers. However, these decamers of SAP form in the absence of calcium [33, 35], and SAP aggregates in the presence of calcium *in vitro* [36]. Thus, the presence of decamers and the implication of cations in the pentamer-pentamer interaction seem to be a common feature in the pentraxin family in general, and for CRP in particular. Moreover, the dimeric interface we observed between the two CRP pentamers was estimated to be more stable when Zn(II) ions participate to the interface. Interestingly, the uptake of zinc by hepatic tissues is associated with the increase of CRP production during acute phase inflammation [37, 38]. Thus, the zinc ions that we evidenced are likely representing a biological feature of CRP that needs further investigations.

The dimer of CRP pentamers involving zinc ions, obtained in the presence of Tat, differs from some decameric structures of CRP described in the literature: two of these decameric structures (1B09, 1LJ7) show that each pentamer forming the decamer obey the same 5-fold molecular symmetry axis, and their relative orientation rotated only from 20° around this 5-fold axis [12, 13]. The A-A dimeric interfaces we have obtained are in fact close to a dimeric interface present in another published CRP structure (1GNH) [11]. However, this dimeric interface was assigned to a non-biological crystal interface, and contained no zinc ions. The fact that the decameric interface has been observed in two different experimental conditions, in three different crystal forms, and that it buries a large area (~ 2000 Å^2 per pentamer) with high free energies, suggests on the contrary that this decameric interface is, in fact, a biological feature of CRP. This is supported by another study demonstrating the presence of CRP decamers involving a pentamer-pentamer interaction through the A-faces of the CRP pentamers [18]. Although that study did not consider a pentamer-pentamer interaction with staggered pentamers in the decamer, their model suggested the importance of the region between residues 170 and 190 for the formation of the decamer [18]. Interestingly, 4 of the 5 H-bonds coordinating the zinc ion in our structure are directly or indirectly involving this region.

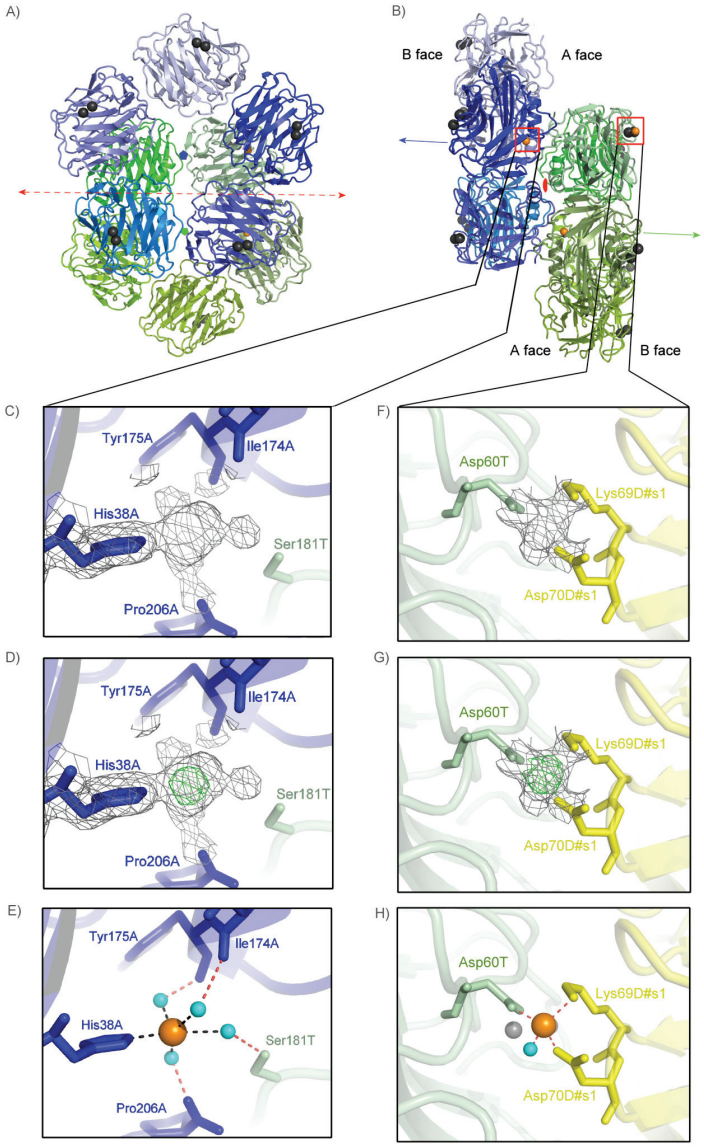

Figure 1. Dimer of CRP pentamers in the presence of Tat. **A:** Front view of the same P1-P4 dimer of pentamer from the B-face of P1, perpendicular to the 5xfold NCS of each pentamer (blue and light green pentagons). NCS 2xfold symmetrical axis is depicted as dashed red arrow.

B: side view of the P1 (blue) - P4 (green) dimer of pentamer perpendicular to the 2xfold NCS axis (red oval). NCS 5xfold symmetrical axis is depicted as blue and light green arrows for the P1 and P4 pentamers, respectively. For both figures, calcium and zinc ions are depicted by black and orange spheres, respectively.

C: Zoom on the unassigned electronic density near to His38 of chain A (P1). The 2 f_o-f_c map is displayed at 1σ.

D: The anomalous difference Fourier map at the Zn edges at 1σ (in green) is superimposed to the electronic density of figure 1C.

E: Localization and coordination of the zinc ion. Cyan spheres represent water molecules; black dashed lines display the penta-coordination of the zinc ion (orange sphere); red dashed lines represent hydrogen bonds between water molecules and residues from the two pentamers P1 (chain A) and P4 (chain T).

F: Zoom on the unassigned electronic density near to Asp60 of chain T (P4). The 2 f_o-f_c map is displayed at 1σ.

G: The anomalous difference Fourier map at the Zn edges at 1σ (in green) is superimposed to the electronic density of figure 1F.

H: Localization and coordination of the zinc ion. Orange, cyan and gray spheres represent the zinc ion, the water molecules, and the calcium ions that do not participate to the coordination, respectively; black dashed lines display the tetrahedral coordination of the zinc ion (orange sphere). The crystallographic symmetric decamer (100) is displayed in yellow.

To that matter, although the canonical sphere of coordination for zinc ions is generally considered as tetrahedral [39], our structure clearly demonstrates a distorted trigonal bipyramidal coordination for the zinc ions present at the pentamer-pentamer interface (Fig. **1E**). Penta-coordination (mostly trigonal bipyramidal) occurs in fact in between 14% to 31% of zinc-bearing proteins, with higher figures in catalytic proteins where the zinc tend to be accessible to solvent molecules [40].

Molecular modelling using X-ray and neutron scattering on SAP have also suggested the presence of dimers of pentamers in solution, involving the A-face of the pentamers, but with a D5 point group symmetry [35]. This was not confirmed by X-ray crystallography, as the decameric form of the crystal structure of SAP in the presence of a ligand showed a decameric interface through the B-faces of the pentamers, primarily due to the Ca^{2+}-associated ligand dAMP [41]. Moreover, CRP residue His38 is substituted by an arginine in SAP, which fills most of the cleft [41]. A zinc-mediated symmetric dimer of pentamers is thus unlikely to occur for SAP.

Our results show that zinc ions are involved in the formation of dimers of CRP pentamers in crystals grown in the presence, but not in the absence, of the HIV Tat protein. Zinc ions were detected by X-ray fluorescence in both the CRP and the Tat protein solutions (Fig. **2**). This could be expected as CRP has been shown to bind zinc *in vitro* [42] and Tat contains a zinc-finger domain [43]. Thus, the zinc ion present in the crystal could originate either from the CRP protein solution or from the Tat solution. One possibility is that the zinc ions are added by the Tat solution that seems to contain more zinc than CRP, and then settle in the cleft around His38 of CRP. Another possibility is that one of the regions of Tat with a strong biochemical signature [44] modifies the physicochemical properties of the crystallization conditions, resulting in the stabilization of zinc ions naturally present in the cleft of the CRP. In both cases, this cleft in CRP would act as a trap for Zn ions. Interestingly, our data do not demonstrate any significant rearrangement of the lateral chains of the residues involved in the Zn(II) binding in the structures obtained in the presence and in the absence of Tat, suggesting that zinc could participate in the stabilization of a decamer that can form anyhow. This is further suggested by the fact that the decameric interface with zinc has been calculated using PISA as more stable than in the absence of zinc, with ΔG of -13.3 kcal/mol and -8.4 kcal/mol, respectively.

Besides these structural and biochemical observations, several lines of evidence also suggest that a physiological role of this zinc-mediated interface could be considered. First, during acute phase, zinc concentration increases in hepatocytes [38], which is the cell type in which CRP is primarily produced during acute phase [45]. Thus, CRP and zinc are located in the same cells at the same moment and could therefore interact when the active production of CRP occurs *in vivo*. Second, CRP exists in an equilibrium between pentamers and decamers in physiological conditions but at low affinity (~20μM) [18]. The fact that zinc is predicted by PISA as stabilizing this decameric form suggests that Zn could play a role in this equilibrium. Third, the cleft of CRP containing the zinc ion implicated in the decameric interface in our structure has been shown to be involved in the interaction with the complement protein C1q [12, 14] as well as for interaction with the membrane associated receptors FcγRII [17]. Thus, this region of CRP is highly reactive and could be the target of a zinc-mediated regulation, although no study has yet evaluated the possible regulatory role of zinc in these interactions. Fourth, a regulatory role for a zinc-mediated dimeric interface has been shown for other proteins from the innate immune response in humans, such as IFNα2b and IFNβ [46, 47]. For these molecules, the dimeric forms have been proposed to be inactive, storage

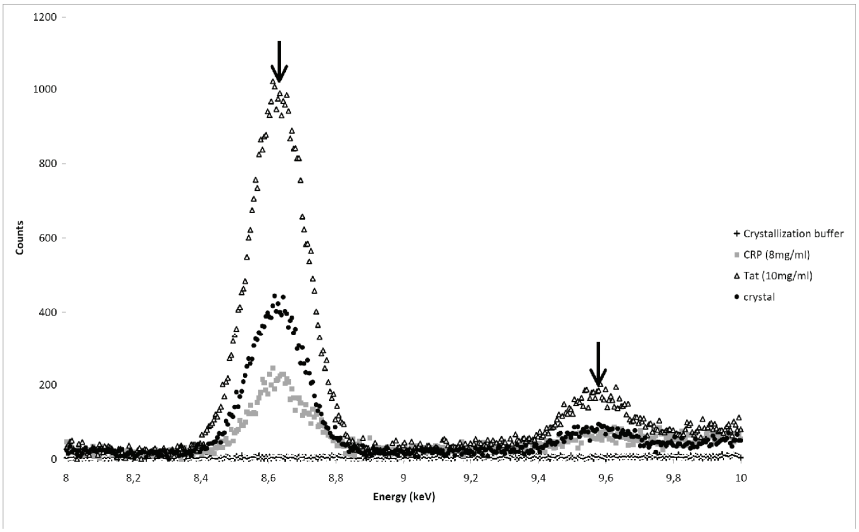

Figure 2. X-ray fluorescence emission spectra of the crystal and the various solutions used for crystallization. Arrows are located at the energy where peaks were measured with a solution of 10mM $ZnCl_2$ (data not shown).

forms of the proteins which would harbour their biological activity as a monomer [46, 47]. Here, the decameric structure of CRP masks the effector faces of the two pentamers and this decamer is predicted to be stabilized by the presence of Zn. Thus, a similar storage form of the zinc-mediated decameric structure of CRP can be envisaged. This could reflect a more general role of zinc in regulating the acute phase [48], in line with the ability of several acute-phase proteins (α2-macroglobulin, ceruloplasmin) to bind zinc *in vitro* [49, 50]. Altogether, our results suggest that zinc could have a physiological role in the regulation of CRP oligomerization and/or function *in vivo* that needs further evaluation.

5. CONCLUSION

Our results show that CRP harbours a decameric conformation in different crystal assemblies and cristallization conditions. This gives a structural support to previous studies that have described the presence of CRP as a decamer in physiological conditions. Interestingly, in the presence of Tat, this CRP decameric structure involves zinc ions located in a cleft known to interact with the complement protein C1q and with membrane associated receptors FcγR, implicated in the CRP-mediated innate immune response and the CRP-mediated opsonisation, respectively. The zinc ions present in the decameric structure of CRP are naturally present in the CRP protein solution, are involving a region important for the function of CRP and are increasing the stability of the decamer. Thus, this zinc-mediated decamers might have functional consequences on the biological properties of CRP which should be further evaluated.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

UMB, PJ and YD prepared and provided the endotoxin-free CRP protein and evidenced the Tat/CRP interaction. CF and CG performed the biochemical and crystallization studies. CG and PG collected data and analyzed the structures. All authors designed the study and wrote the paper. We thank Drs N. Bouheden and A. Galioot for technical assistance, Dr X. Robert for his help with structure analysis and bioinformatic tools, and Dr A. Beck for stimulating discussions. We also thank the staff of the ESRF BM30A beamline (FIP) for their help in the identification of the zinc ions. This work was supported by the Agence Nationale de Recherche sur le SIDA (ANRS), and UMB was a recipient from a doctoral grant from ANRS.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publishers Web site along with the published article.

REFERENCES

[1] Garlanda, C.; Bottazzi, B.; Bastone, A.; Mantovani, A., Pentraxins at the crossroads between innate immunity, inflammation, matrix deposition, and female fertility. *Annu Rev Immunol* **2005,** *23*, 337-366.

[2] Mantovani, A.; Garlanda, C.; Doni, A.; Bottazzi, B., Pentraxins in innate immunity: from C-reactive protein to the long pentraxin PTX3. *J Clin Immunol* **2008,** *28*, 1-13.

[3] Martinez de la Torre, Y.; Fabbri, M.; Jaillon, S.; Bastone, A.; Nebuloni, M.; Vecchi, A.; Mantovani, A.; Garlanda, C., Evolution of the pentraxin family: the new entry PTX4. *J Immunol* **2010,** *184*, 5055-5064.

[4] Gewurz, H.; Zhang, X. H.; Lint, T. F., Structure and function of the pentraxins. *Curr Opin Immunol* **1995,** *7*, 54-64.

[5] Du Clos, T. W., Function of C-reactive protein. *Ann Med* **2000,** *32*, 274-278.

[6] Pepys, M. B.; Hirschfield, G. M., C-reactive protein: a critical update. *J Clin Invest* **2003,** *111*, 1805-1812.

[7] Mantovani, A.; Valentino, S.; Gentile, S.; Inforzato, A.; Bottiger, B.; Garlanda, C., The long pentraxin PTX3: a paradigm for humoral pattern recognition molexules. *Ann N. Y. Acad Sci* **2013,** *1285*, 1-14.

[8] Hicks, P. S.; Saunero-Nava, L.; Du Clos, T. W.; Mold, C., Serum amyloid P component binds to histones and activates the classical complement pathway. *J Immunol* **1992,** *149*, 3689-3694.

[9] Gershov, D.; Kim, S.; Brot, N.; Elkon, K. B., C-Reactive protein binds to apoptotic cells, protects the cells from assembly of the terminal complement components, and sustains an antiinflammatory innate immune response: implications for systemic autoimmunity. *J Exp Med* **2000,** *192*, 1353-1364.

[10] Schwedler, S. B.; Filep, J. G.; Galle, J.; Wanner, C.; Potempa, L. A., C-reactive protein: a family of proteins to regulate cardiovascular function. *Am J Kidney Dis* **2006,** *47*, 212-222.

[11] Shrive, A. K.; Cheetham, G. M.; Holden, D.; Myles, D. A.; Turnell, W. G.; Volanakis, J. E.; Pepys, M. B.; Bloomer, A. C.; Greenhough, T. J., Three dimensional structure of human C-reactive protein. *Nat Struct Biol* **1996,** *3*, 346-354.

[12] Thompson, D.; Pepys, M. B.; Wood, S. P., The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure* **1999,** *7*, 169-177.

[13] Ramadan, M. A.; Shrive, A. K.; Holden, D.; Myles, D. A.; Volanakis, J. E.; DeLucas, L. J.; Greenhough, T. J., The three-dimensional structure of calcium-depleted human C-reactive protein from perfectly twinned crystals. *Acta Crystallogr D Biol Crystallogr* **2002,** *58*, 992-1001.

[14] Agrawal, A.; Shrive, A. K.; Greenhough, T. J.; Volanakis, J. E., Topology and structure of the C1q-binding site on C-reactive protein. *J Immunol* **2001,** *166*, 3998-4004.

[15] Roumenina, L. T.; Ruseva, M. M.; Zlatarova, A.; Ghai, R.; Kolev, M.; Olova, N.; Gadjeva, M.; Agrawal, A.; Bottazzi, B.; Mantovani, A.; Reid, K. B.; Kishore, U.; Kojouharova, M. S., Interaction of C1q with IgG1, C-reactive protein and pentraxin 3: mutational studies using recombinant globular head modules of human C1q A, B, and C chains. *Biochemistry* **2006,** *45*, 4096-4104.

[16] Ng, P. M.; Le Saux, A.; Lee, C. M.; Tan, N. S.; Lu, J.; Thiel, S.; Ho, B.; Ding, J. L., C-reactive protein collaborates with plasma lectins to boost immune response against bacteria. *Embo J* **2007,** *26*, 3431-3440.

[17] Lu, J.; Marnell, L. L.; Marjon, K. D.; Mold, C.; Du Clos, T. W.; Sun, P. D., Structural recognition and functional activation of FcgammaR by innate pentraxins. *Nature* **2008,** *456*, 989-992.

[18] Okemefuna, A. I.; Stach, L.; Rana, S.; Buetas, A. J. Z.; Gor, J.; Perkins, S. J., C-reactive protein exists in an NaCl concentration-dependent pentamer-decamer equilibrium in physiological buffer. *J Biol Chem* **2010,** *285*, 1041-1052.

[19] Biro, A.; Rovo, Z.; Papp, D.; Cervenak, L.; Varga, L.; Fust, G.; Thielens, N. M.; Arlaud, G. J.; Prohaszka, Z., Studies on the interactions between C-reactive protein and complement proteins. *Immunology* **2007,** *121*, 40-50.

[20] Mihlan, M.; Stippa, S.; Józsi, M.; Zipfel, P. F., Monomeric CRP contributes to complement control in fluid phase and on cellular surfaces and increases phagocytosis by recruiting factor H. *Cell Death Differ* **2009,** *16*, 1630-1640.

[21] Potempa, L. A.; Siegel, J. N.; Fiedel, B. A.; Potempa, R. T.; Gewurz, H., Expression, detection and assay of a neoantigen (Neo-CRP) associated with a free, human C-reactive protein subunit. *Mol Immunol* **1987,** *24*, 531-541.

[22] Ji, S. R.; Ma, L.; Bai, C. J.; Shi, J. M.; Li, H. Y.; Potempa, L. A.; Filep, J. G.; Zhao, J.; Wu, Y., Monomeric C-reactive protein activates endothelial cells via interaction with lipid raft microdomains. *Faseb J* **2009,** *23*, 1806-1816.

[23] Khreiss, T.; Jozsef, L.; Potempa, L. A.; Filep, J. G., Opposing effects of C-reactive protein isoforms on shear-induced neutrophil-platelet adhesion and neutrophil aggregation in whole blood. *Circulation* **2004,** *110*, 2713-2720.

[24] Guillon, C.; Stankovic, K.; Ataman-Önal, Y.; Biron, F.; Verrier, B., Evidence for CTL-mediated selection of Tat and Rev mutants after the onset of the asymptomatic period during HIV-1 infection. *AIDS Res Hum Retroviruses* **2006,** *22*, 1283-1292.

[25] Mayol, K.; Munier, S.; Beck, A.; Verrier, B.; Guillon, C., Design and characterization of an HIV-1 Tat mutant : inactivation of viral and cellular functions but not antigenicity. *Vaccine* **2007,** *25*, 6047-6060.

[26] Foucault, M.; Mayol, K.; Receveur-Bréchot, V.; Bussat, M.-C.; Klinguer-Hamour, C.; Verrier, B.; Beck, A.; Haser, R.; Gouet, P.; Guillon, C., UV and X-ray structural studies of a 101-residue long Tat protein from a HIV-1 primary isolate and of its mutated, detoxified, vaccine candidate. *Proteins* **2010,** *78*, 1441-1456.

[27] Serrière, J.; Dugua, J.-M.; Bossus, M.; Verrier, B.; Haser, R.; Gouet, P.; Guillon, C., Fab'-induced folding of antigenic N-terminal peptides from intrinsically unstructured HIV-1 Tat protein revealed by X-ray crystallography. *J Mol Biol* **2011,** *405*, 33-42.

[28] Kabsch, W., Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* **1993,** *26*, 795-800.

[29] Navaza, J., AMoRe: an automated package for molecular replacement. *Acta Cryst. A* **1994,** *50*, 157-163.

[30] Collaborative Computational Project, N., The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **1994,** *50*, 760-763.

[31] Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **1998,** *54*, 905-921.

[32] Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **2007,** *372*, 774-797.

[33] Aquilina, J. A.; Robinson, C. V., Investigating interactions of the pentraxins serum amyloid P component and C-reactive protein by mass spectrometry. *Biochem J* **2003,** *375*, 323-328.

[34] Perkins, S. J.; Okemefuna, A. I.; Nan, R., Unravelling protein-protein interactions between complement factor H and C-reactive protein using a multidisciplinary strategy. *Biochem Soc Trans* **2010,** *38*, 894-900.

[35] Ashton, A. W.; Boehm, M. K.; Gallimore, J. R.; Pepys, M. B.; Perkins, S. J., Pentameric and decameric structures in solution of serum amyloid P component by X-ray and neutron scattering and molecular modelling analyses. *J Mol Biol* **1997,** *272*, 408-422.

[36] Hutchinson, W. L.; Hohenester, E.; Pepys, M. B., Human serum amyloid P component is a single uncomplexed pentamer in whole serum. *Mol Med* **2000,** *6*, 482-493.

[37] Craig, G. M.; Evans, S. J.; Brayshaw, B. J., An inverse relationship between serum zinc and C-reactive protein levels in acutely ill elderly hospital patients. *Postgrad Med J* **1990,** *66*, 1025-1028.

[38] Schroeder, J. J.; Cousins, R. J., Interleukin 6 regulates metallothionein gene expression and zinc metabolism in hepatocyte monolayer cultures. *Proc Natl Acad Sci U S A* **1990,** *87*, 3137-3141.

[39] Auld, D. S., Zinc coordination sphere in biochemical zinc sites. *Biometals* **2001,** *14*, 271-313.

[40] Patel, K.; Kumar, A.; Durani, S., Analysis of the structural consensus of the zinc coordination centers of metalloprotein structures. *Biochim Biophys Acta* **2007,** *1774*, 1247-1253.

[41] Hohenester, E.; Hutchinson, W. L.; Pepys, M. B.; Wood, S. P., Crystal structure of a decameric complex of human serum amyloid P component with bound dAMP. *J Mol Biol* **1997,** *269*, 570-578.

[42] Scott, B. J.; Bradwell, A. R., Identification of the serum binding protein for iron, zinc, cadmium, nickel and calcium. *Clin Chem* **1983,** *29*, 629-633.

[43] Tahirov, T. H.; Babayeva, N. D.; Varzavand, K.; Cooper, J. J.; Sedore, S. C.; Price, D. H., Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature* **2010,** *465*, 747-751.

[44] Jeang, K. T.; Xiao, H.; Rich, E. A., Multifaceted activities of the HIV-1 transactivator of transcription, Tat. *J Biol Chem* **1999,** *274*, 28837-28840.

[45] Ansar, W.; Ghosh, S., C-reactive protein and the biology of disease. *Immunol Res* **2013,** *56*, 131-142.

[46] Radhakrishnan, R.; Walter, L. J.; Hruza, A.; Reichert, P.; Trotta, P. P.; Nagabhushan, T. L.; Walter, M. R., Zinc mediated dimer of human interferon-alpha 2b revealed by X-ray crystallography. *Structure* **1996,** *4*, 1453-1463.

[47] Karpusas, M.; Nolte, M.; Benton, C. B.; Meier, W.; Lipscomb, W. N.; Goelz, S., The crystal structure of human interferon beta at 2.2-A resolution. *Proc Natl Acad Sci U S A* **1997,** *94*, 11813-11818.

[48] Liu, M. J.; Bao, S.; Napolitano, J. R.; Burris, D. L.; Yu, L.; Tridandapani, S.; Knoell, D. L., Zinc regulates the acute phase response and serum amyloid A production in response to sepsis through JAK-STAT3 signaling. *PLoS One* **2014,** *9*, e94934.

[49] Gettings, P.; Cunningham, L. W., A unique pair of zinc binding sites in the human alpha 2-macroglobulin tetraler. A 35Cl and 37Cl NMR study. *Biochemistry* **1986,** *25*, 5004-5010.

[50] Wang, F.; Chmil, C.; Pierce, F.; Ganapathy, K.; Gump, B. B.; MacKenzie, J. A.; Mechref, Y.; Bendinskas, K., Immobilized metal affinity chromatography and human serum proteomics. *Journal of Chromatography B* **2013,** *934*, 26-33.

Received: ???? 22, 2014 **Revised: ????? 23, 2014** **Accepted: ????? 4, 2014**

viruses

MDPI

Research article

Crystal structure of the full-length Feline Immunodeficiency Virus capsid protein unveils original features.

Christelle Folio [1], Natalia Sierra [2], Marie Dujardin [1], Guzman Alvarez [2] and Christophe Guillon [1,*]

[1] Equipe Rétrovirus et Biochimie Structurale, Université de Lyon, CNRS, MMSB, UMR 5086 CNRS/Université de Lyon, IBCP, Lyon, France

[2] Laboratorio de Moléculas Bioactivas, Centro Universitario Regional Litoral Norte, Universidad de la República, Paysandú, Uruguay

* Correspondence: christophe.guillon@ibcp.fr; Tel. : +33(0)4-37-65-29-04

Academic Editor: name
Received: date; Accepted: date; Published: date

Abstract: Feline Immunodeficiency Virus (FIV) is a member of the *Retroviridae* family. It is the causative agent of an acquired immunodeficiency syndrome (AIDS) in cats and wild felines. Its capsid protein (CA) drives the assembly of the viral particle, which is a critical step of the viral replication cycle. Here, we determine the first atomic structure of full-length FIV CA to 1.9 Å resolution. The crystallized protein exhibits an original tetrameric assembly, composed of a dimer of dimers which are stabilized by an intermolecular disulfide bridge induced by the crystallogenesis conditions. FIV CA displays a standard α-helical CA topology with two domains, separated by a linker shorter than for other retroviral CA. The β-hairpin motif at its amino terminal end, which interacts with nucleotides in HIV, is unusually long in FIV CA. Interestingly, this functional β-motif is formed in our construct in the absence of the conserved N-terminal proline. FIV CA exhibits a *cis* Arg-Pro bond in the CypA-binding loop, which is absent in known structures of lentiviral CAs. Our structure represents the first tri-dimensional structure of a functional, full-length FIV CA.

Keywords: Feline Immunodeficiency Virus; FIV; Capsid protein; Crystal structure

1. Introduction

Retroviruses are a major concern of public health in humans but also in animals. The Feline Immunodeficiency Virus (FIV) is the causative agent of an acquired immunodeficiency syndrome (AIDS) in felines [1] with a prevalence rate up to 30% of domestic cats in some areas [2, 3]. FIV is a member of the genus *Lentivirus* from the *Retroviridae* family [4], which also contains Human Immunodeficiency Virus (HIV), Equine Infectious Anemia Virus (EIAV) and Simian Immunodeficiency Virus (SIV), among others. Due to their common biological characteristics such as virion morphology, physiology and pathogenesis, FIV has been described as a useful non-primate model for HIV infection, antiretroviral therapy and vaccine development. FIV could also be used as a simple model for rational drug design for HIV [5-7].

Like all infectious retroviruses, FIV genome contains the three genes *gag*, *pol* and *env* encoding for the structural proteins, the viral enzymes and the envelope proteins, respectively [8]. The Gag polyprotein is involved in the architecture of the viral particle [9, 10]. As for HIV and SIV (but not EIAV), FIV Gag protein is myristoylated at its N-terminus [11-14], which allows its targeting to the plasma membrane at the time of particle assembly [15]. During the maturation step of the virus life cycle, the

FIV Gag polyprotein is cleaved by the viral protease into different subunits: the matrix protein (MA), the capsid protein (CA), and the nucleocapsid protein (NC), with a spacer peptide p1 between the CA and NC domains and a C-terminal peptide p2 [16]. The MA protein forms a layer that underlies the viral envelope of the virion, which contains the CA cone-shaped core and the complex formed by the NC and the RNA viral genome [5].

Retrovirus assembly is a critical step in the viral replication cycle and this process is driven by the CA protein [17-19]. Indeed, CA protein forms a protective coat around the viral genome and its assembly into a cone-shaped core is characteristic of a mature lentivirus. In fact, a virus particle without a properly assembled cone-shaped core appears to be non-infectious [20]. Based on this observation, CA protein is a promising therapeutic target for antiretroviral therapy against HIV [21].

Previous crystallographic studies have determined the structure of full-length retroviral CA units as dimer for EIAV [22] and dimer, pentamer or hexamer for HIV [23-26]. Such high-order oligomers are necessary for the formation of a mature capsid core in the viral particle [27, 28]. Cross-linking agents and antibody fragments have often been used to stabilize CA for crystallographic studies [29] and CryoEM studies have deciphered supramolecular assemblies of tubular and cone-shaped CAs [30]. Despite low sequence homologies, retroviral CA proteins harbor a rather similar α-helical topology, with two domains, the amino-terminal domain (NTD) and the carboxy-terminal domain (CTD) connected by a flexible linker [22, 23]. This linker plays an important role in the relative flexibility of the CA_{NTD} and CA_{CTD} during the assembly of CA proteins into pentamers or hexamers [25, 26]. A highly conserved sequence is observed amongst all retroviral CA proteins, called the major homology region (MHR), while the rest of the sequence is less conserved between retroviral species. This MHR is required for the correct folding and stability of the CA_{CTD} domain and thus is essential for viral replication [31, 32].

So far, only the structure of the CTD domain of FIV CA has been described [33]. To better understand the molecular and structural specificities of this protein, we determined its crystal structure at 1.9 Å resolution and observed original features when compared with other lentiviral CAs such as HIV and EIAV. Functional consequences will be discussed.

2. Materials and Methods

2.1. Construction of recombinant plasmid encoding the FIV capsid protein

The full-length native FIV capsid protein was amplified by polymerase chain reaction on the plasmid p34TF10 (Petaluma strain) as described [34]. A truncated form of the CA protein in its C-terminal end of 11 amino acid residues with a mutation of Pro1 in Thr1 (p24EΔCP-T) was then constructed by PCR using the same protocol and a pair of primers, 5′-AGGATCCAATAGAAGGACGA**ACT**ATTCAAACAGT-3′ and 5′-TGAATTC<u>TCA</u>TATTTCTTGACAAGCCCTCAAC-3′, where the Pro1Thr mutation and the introduced stop codon are shown in bold and underlined, respectively. The Pro1Thr mutation was introduced to allow the removal of the 6xHis tag by the Factor Xa protease, which removes all the amino acid of the cleavage site, allowing to have an intact N-terminus of the protein of interest. However, Factor Xa is not active when the first amino acid after the cleavage site is a proline residue. The product was digested with BamHI and EcoRI and then ligated into the BamHI/EcoRI sites of the vector pRSET-B (Invitrogen) to form the recombinant plasmid pRSET-p24EΔCP-T encoding the FIV CA protein with a 6xHis tag at its N-terminal end.

2.2. Expression and purification of FIV CA protein

Escherichia coli cells (BL2I (DE3) pLysS, Novagen) transformed with pRSET-p24EΔCP-T were grown in Lysogenic-Broth medium (Sigma) supplemented with 50 mg/ml of ampicillin, at 37°C. Cell density was monitored by measuring the optical density at 600 nm (OD_{600}). When cells reached an OD_{600} value between 0.3 and 0.4, the expression of CA protein was induced by the addition of IPTG (Isopropyl-β-D-1-thiogalactopyranoside, Euromedex) to a final concentration of 1 mM. Expression

was carried on for an additional 20 h at 25°C, then cells were harvested by centrifugation and the pellets were stored overnight at –20°C.

Purification of CA protein was performed by nickel affinity chromatography as for the native CA protein [34]. Briefly, the lysate was clarified by centrifugation at 10 000 g for 45 min, and the supernatant was filtered through a 0.45 µm membrane. Purification of the protein from the supernatant was done by batch incubation Ni^{2+}-TED resin (Macherey-Nagel) followed by loading onto a gravity column. The column was washed three times with LEW buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH8.5), and the elution was then performed with LEW buffer containing 50 mM of imidazole.

The concentration of CA protein was quantified by spectrophotometry at 280 nm, using a Nanodrop (ThermoFischer). The purity of the protein was evaluated by SDS-PAGE analysis. Buffer exchange, using Vivaspin ultrafiltration devices (10 kD MWCO, Sartorius), was performed against HEPES/NaCl Buffer (50 mM HEPES pH 6.5, 100 mM NaCl).

2.3. Removal of the 6xHis tag

To remove the 6xHis tag, purified CA protein in HEPES/NaCl buffer was digested overnight with 16U of Factor Xa (Qiagen) per mg of CA protein, at 19°C. After proteolysis, the tag-free protein was obtained by loading the sample on a Ni-NTA centrifugation column (Proteus) according to the manufacturer's protocol, and collecting the flowthrough. Purified CA proteins were then concentrated to 7 mg/ml using a Vivaspin centrifugal concentrator (10 kD MWCO, Sartorius).

2.4 Crystallization of the FIV CA protein

Screening of crystallization conditions was performed in 96-well plates using a Mosquito nanopipetter and commercial crystallization screening kits (Hampton Research and Qiagen) with the sitting drop procedure. FIV CA protein at 7 mg/mL in HEPES/NaCl buffer crystallizes in the presence an equal volume of 0.2 M magnesium sulfate, 20% PEG 4000, 10% glycerol (condition E11 of the Qiagen PEGs II Suite) supplemented with 10% DMSO final. Using these conditions with the hanging drop technique and drops of 1 µl of protein with 1 µl of crystallization conditions, plate-shaped crystals grew within fifteen days. Due to the presence of 10% glycerol in the crystallization solution, the cryoprotection step was dispensable and crystals were directly flash frozen in liquid nitrogen prior to data collection.

2.5. X-ray data collection and structure determination

X-ray data were collected at best to 1.9 Å resolution at the European Synchrotron Research Facility (ESRF) beamline ID30-B (Grenoble, France) at 100K with a wavelength of 0.99187 Å and a PILATUS 6M-F detector. Crystals belong to monoclinic space group C2 with cell dimensions a=122.2 Å, b=74.6 Å, c=77.0Å, α=ɣ=90.0°, β= 128.7°. Indexation and scaling were performed using XDS and XSCALE programs [35]. The structure of FIV CA protein was determined by molecular replacement using the program MrBUMP [36] of the CCP4 program suite [37] and the structures of Relik (Rabbit Endogenus Lentivirus) CA_{NTD} fragment (PDB ID: 2XGU) [38] and FIV CA_{CTD} fragment (PDB ID: 5DCK) [33] as search models. One solution was found with two monomers in the asymmetric unit and an R-factor of 48%. The crystallographic refinement was performed with PHENIX [39]. A few residues in the β-hairpin, the cyclophilin binding loop and the C-terminal end were built manually using WinCOOT [40] and 6 molecules of glycerol were positioned in the electron density maps. The structure was refined to a final R_{work} of 19 % and R_{free} of 22.5% and statistics of the X-ray data are showed in Table 1. It shows a good geometry with 98.1% in preferred regions, 1.9% in allowed regions, and no Ramachandran outliers. The omitmap around the cis-peptide of the CypBL loop was generated using the PHENIX software with the annealing method on residues 88 to 92 from chain B. Figures were generated using PyMol (Schrödinger) [41].

Table 1. Summary of X-ray data collection and refinement statistics. Values in parentheses are for the highest-resolution shell.

Data collection	FIV CA
Space group	C121
Unit cell parameters	
a, *b*, *c* (Å)	122 , 74, 77
α, β, γ (°)	90, 128, 90
Resolution range (Å)	59.7 - 1.90 (2.01 - 1.90)
R_{sym} (%)	3.54 (25.8)
I/σI	16.42 (3.33)
Completeness (%)	91.2 (93.0)
Redundancy	2.38
Refinement	
Resolution range (Å)	59.7 - 1.9
Number of unique reflections	39142
R_{work} (%)	19
R_{free} (%)	22.5
Number of proteins atoms	3330
Number of water/glycerol	697/6
Mean B-factor (Å^2)	22.9
Coordinate deviations	
RMSD bond lengths (Å)	0.007
RMSD angles (°)	0.857
PDB ID	5NA2

3. Results

3.1. Asymmetric unit and crystal packing

The 1.9 Å crystal structure of FIV CA was determined in the monoclinic space group C2 with two monomers in the asymmetric unit. The two chains (chains A and B) are quasi equivalent in structure and can be superposed with a Root-Mean-Square Deviation (RMSD) of 0.63 Å on all Cα pairs.

FIV CA appears to be organized as a dimer of dimers in the crystal (Figure 1a), which assembly is predicted to be stable in solution according to the PISA server for detection of biological oligomers [42]. The A-B dimer (Figure 1b) of the asymmetric unit is related with its symmetrical counterpart A'-B' *via* a 2-fold crystallographic axis to form a tetramer A-B-A'-B' (Figure 1a).

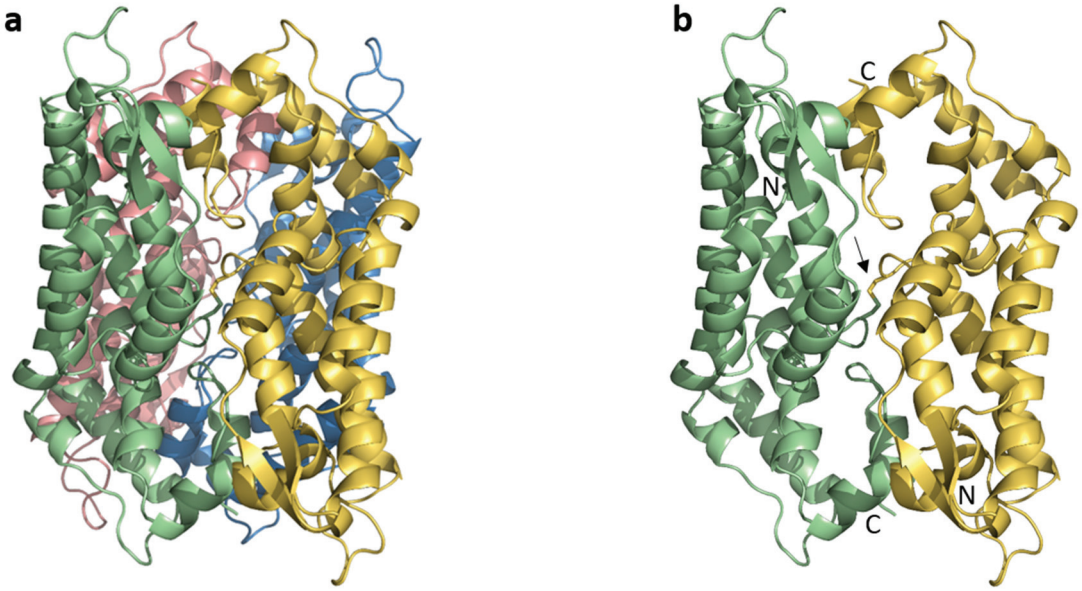

Figure 1. Crystal-packing interfaces and assembly of FIV CA. (a) Tetrameric packing of FIV CA protein (chain A: yellow, chain B: green, chain A': bleu and chain B': pink). (b) Dimer A-B in the asymmetric unit. The disulfide bridge between chains A and B is labeled with an arrow.

Dimeric CA interfaces have been described for other lentiviruses: for EIAV (PDB ID: 2EIA) [22], the CA_{NTD} from one monomer is interacting with the CA_{NTD} from the second monomer in a head-to-tail orientation, resulting in free CA_{CTDs} in opposite directions in the dimer. For HIV (PDB ID: 3NTE) [43], the CA_{NTD} from one monomer (helices α1 and α2) is interacting with the CA_{CTD} from the other monomer (helices 3_{10} and α9), also resulting in a head-to-tail dimer.

In our FIV CA structure, the dimer A-B of the asymmetric unit contains the two CA monomers in a head-to-tail orientation (Figure 1b). The FIV CA_{NTD} of one monomer interacts through loop L4 (between α3 and α4, Figure 2) with the loop L12 (between α10 and α11) of the CA_{CTD} of the other monomer. The average buried surface area in this dimer is ~1100 Å^2 per monomer, with a calculated ΔG of -21.1 kcal/mol. This interface is further stabilized through interactions between the tips of α1 of each FIV CA_{NTD} (Figure 1b).

Intriguingly, the two monomers in the asymmetric unit are covalently linked by a disulfide bridge between the Cys61 of each CA monomer (Figure 1b). However, SDS-PAGE in non-reducing conditions (Figure S1) and mass spectrometry experiments (data not shown) demonstrate the absence of covalently linked monomers in the protein solution used for crystallogenesis experiments. This suggests that this disulfide bridge is a crystallization artifact which likely stabilized the dimer.

3.2. The crystal structure of full length FIV CA

Despite having a low sequence similarity with HIV and EIAV (29% and 39%, respectively), the general fold of FIV CA is similar to that of these lentiviruses with an α-helical, two-domains structure (Figure 2). FIV CA_{NTD} consists of a 12-residues long β-hairpin followed by 7 α-helices (numbered α1 to α7) and contains a cyclophilin-binding loop (CypA-BL in loop L5), which is well defined in the electron density maps. FIV CA_{CTD} follows a short flexible loop (loop L8) linking the two domains, and contains 4 α-helices (α8 to α11) including the highly conserved major homology region (MHR) which is present in all retroviral CA and essential for viral replication.

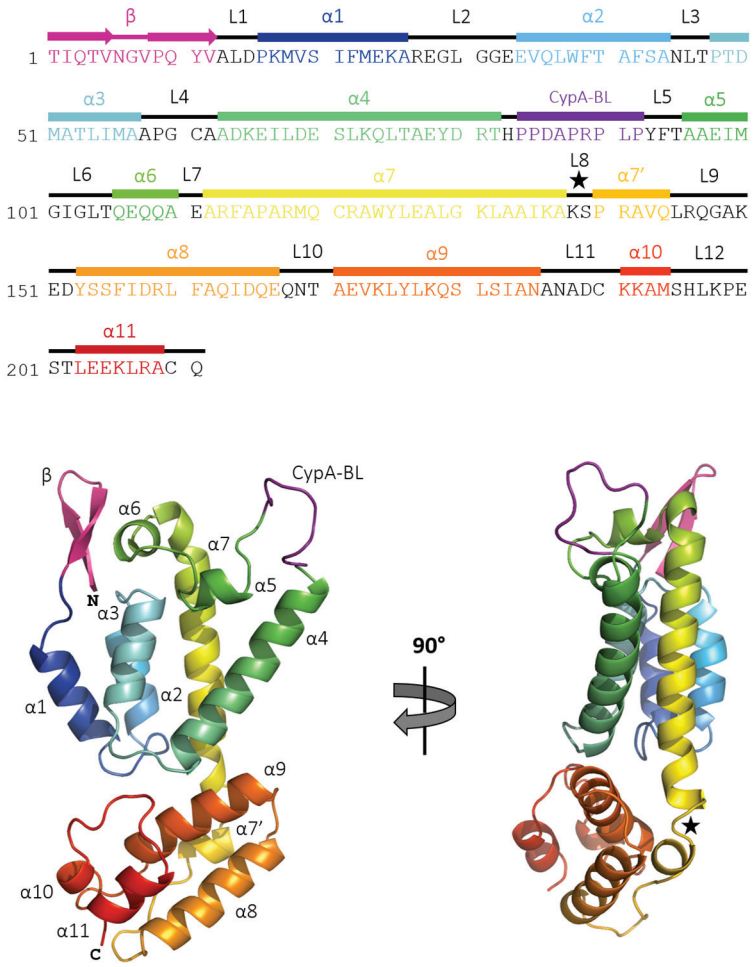

Figure 2. Secondary structures and ribbon diagrams of full-length FIV CA monomer. Helices are colored blue-to-red from N- to C-terminus with helices numbered sequentially from α1 to α11. The β-hairpin is colored in pink and the cyclophilin A-binding loop in purple. The L8 loop, corresponding to the linker between CA$_{NTD}$ and CA$_{CTD}$, is labeled with a black star.

3.3. Comparisons with known structures of lentiviral CAs

Structural comparisons of FIV CA with full-length HIV and EIAV CAs show that the orientation of the CA$_{CTD}$ with respect to the CA$_{NTD}$ is different in FIV CA (Figure 3). These two domains are close to each other, making the FIV CA structure more compact than what is observed for HIV and EIAV (Figure 3). This feature is due to the small linker between FIV CA$_{NTD}$ and CA$_{CTD}$. Still, CAs$_{NTD}$ and CAs$_{CTD}$ can be superposed separately between FIV, HIV and EIAV with an RMSD between 0.5 and 1 Å on Cα pairs (data not shown), demonstrating the global conservation of the lentiviral CA fold in FIV CA. Similarly, a low RMSD value (0.5 Å) is obtained when superposing our FIV CA$_{CTD}$ with the recently solved FIV CA$_{CTD}$ fragment [33].

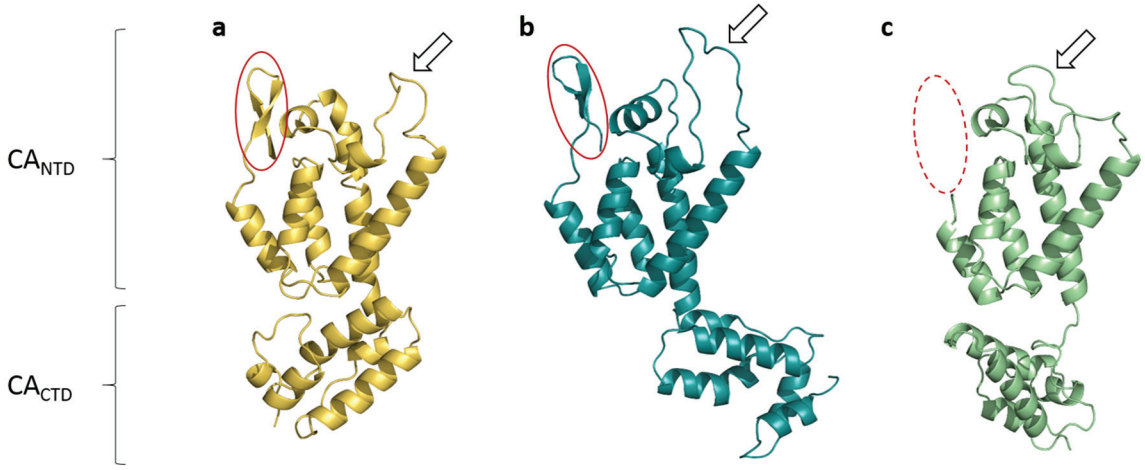

Figure 3. Structural comparison of retroviral capsid protein: (a) FIV in yellow, (b) HIV in teal (PDB ID: 3NTE [43]) and (c) EIAV in green (PDB ID: 2EIA [22]). β-hairpins are circled in red and CypA-BLs are shown with an arrow.

In more details, the first 12 N-terminal residues of FIV CA consist in a β-hairpin motif, which appears longer than that of HIV (Figure 3b). This region is not observed in EIAV CA (Figure 3c). Another difference concerns the CypA-BL of the NTD, which appears to be smaller in our FIV CA structure (Figure 3a, arrow) compared to other retroviral CA proteins such as HIV (Figure 3b).

Noteworthy, the CypA-BL in FIV CA contains a *cis* Arg89-Pro90 peptide bond in both chains in the asymmetric unit, which is 100% in *cis*-conformation (Figure 4). This has no equivalent in published HIV, EIAV or Relik CA structures.

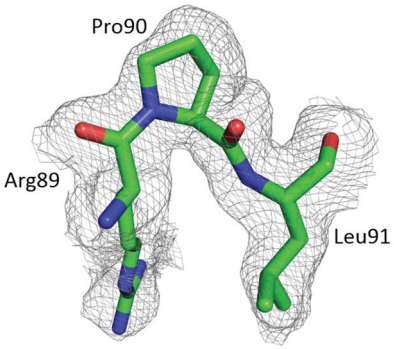

Figure 4. Omitmap of FIV CA cis Arg89-Pro90 peptide bond from chain B CypA-BL, contoured at 1σ. The oxygen and nitrogen atoms are shown in red and blue, respectively.

As concerns the CTD, Cys190 and Cys210 show an alternate conformation resulting in the formation of an intramolecular disulfide bridge with an occupancy of about 75%, which is absent in the published structure of FIV CA$_{CTD}$ [33]. The 3_{10}-helix which is observed in the HIV CA$_{CTD}$ between α7 and α8 [33] appears to be replaced by an α-helix (numbered α7′) in our FIV CA full-length structure (Figure 2).

4. Discussion

To define the specificity of the molecular mechanisms underneath FIV assembly, we investigated the structure of FIV full-length CA protein and compared it with structures of other retrovirus CA

proteins. FIV CA protein is mostly composed of α-helices, like the CA protein of other retrovirus, confirming that the overall α-helical fold of CA protein is highly conserved among retroviruses.

During the assembly, retroviral CA proteins assemble into pentamers and hexamers [44, 45] to form a cone-shaped core, but no pentameric or hexameric assemblies were observed for our FIV CA structure.

The crystal structure of FIV CA contains one dimer of CA protein (chains A and B) in the asymmetric unit. This dimer is probably not functional since the N-terminal ends are oriented in opposite directions while they should be oriented in the same way for proper interaction with the FIV matrix protein (MA). Moreover, this dimer is covalently linked by a disulfide bridge between the Cys61 of each monomer. We showed that this disulfide bridge is a crystallization artifact as it is absent from the protein solution used for crystallogenesis. This is coherent with the observation that Cys61 is not involved in disulfide bridges in functional FIV CA [19]. This crystallization artifact is likely due to the presence of dimethyl-sulfoxide (DMSO) in the crystallization condition. Indeed, DMSO has been reported to promote oxidation of thiol into disulfide at low pH and room temperature [46]. The presence of this artifactual disulfide bridge may have helped stabilizing the CA dimer in the asymmetric unit during crystallization process, resulting in the formation of FIV CA crystals which were not obtained in the absence of DMSO (data not shown). However, as a drawback, the formation of this bridge might also have stabilized the FIV CA dimers in non-relevant interfaces and/or impaired the formation of high order oligomers (pentamers, hexamers) which are necessary for the formation of the retroviral capsid. Crystallogenesis experiments in the absence of DMSO are therefore currently pursued to identify unambiguously the functional oligomeric interfaces.

We observed that Cys190 and Cys210, which are conserved in retroviral CAs, form an intrachain disulfide bridge in 75% of the proteins in the crystal for both chains of the asymmetric unit. This cysteine bond was absent in the isolated FIV CA_{CTD} structure [33]. Although it might have been promoted by the presence of DMSO in our crystal, the presence of this cysteine bond is in agreement with biochemical studies who reported that cysteines Cys190 and Cys210 are involved in an intrachain disulfide bound which is necessary for FIV capsid assembly and FIV infectivity [19]. Moreover, this cysteine bond is highly conserved across several retroviral CA proteins, from HIV to EIAV [22, 47]. Thus, this structural feature is probably relevant for the biology of FIV CA.

As expected from biochemical data [19], the last free cysteine of FIV CA, Cys121 (from α7'), is not involved in any cysteine bond. Interestingly, the sulfur atom of this cysteine participates to a cluster with sulfur atoms of Met51 (from α4) and Met100 (from α6). These three sulfur atoms show an intriguing feature, as they are aligned and distributed at 4 Å one after the other (data not shown). This distribution is unique to FIV CA as other retroviral do not harbor a cysteine residue homologous to FIV Cys121, but its function (if any) remains to be determined.

However, the FIV CA monomeric structure that we have obtained harbors important features to understand FIV assembly. For example, FIV CA contains at its amino-terminal end a β-hairpin motif (Figure 5). This motif could be expected as it has been shown to be required for the formation of the HIV capsid core particle since it participates directly in intermolecular CA–CA interactions [23, 48]. In our structure, this β-hairpin seems to adopt a conformation which corresponds to the "open" conformation described for HIV-1 CA protein [26] which might be important for the import of dNTP in the virus core during reverse-transcription.

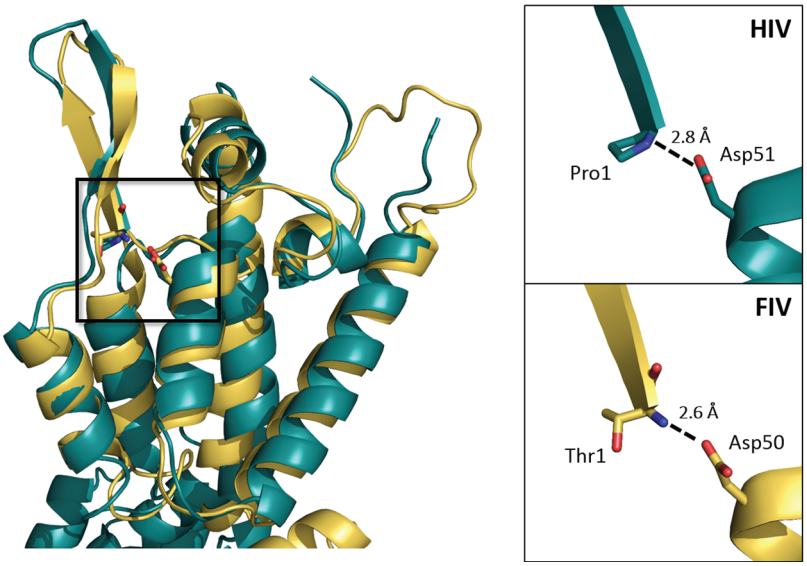

Figure 5. Details of the N-terminal β-hairpin of FIV CA. Superposition of FIV CANTD (yellow) and HIV CANTD in its "open" conformation (teal, PDB ID: 5HGL [26]). A detailed view of the boxed region (right panel) shows the salt bridge of HIV Pro1 with Asp51 (top) and FIV Thr1 with Asp50 (bottom).

The Pro1 of CA has been described as essential for the formation of this N-terminal β-hairpin in HIV CA [49]. Interestingly, this proline is conserved in most lentiviral CA protein such as HIV, SIV and EIAV, which probably reflects a key function for viral replication. So far, the functional role of this proline has been attributed to the formation of the β-hairpin. However, the Pro1 into Thr1 mutation that we have introduced in FIV CA for practical reasons (described in the materiel and method section) did not impair the formation of the β-hairpin nor the assembly of FIV CA *in vitro* (Figure S2). Moreover, the salt bridge between the terminal NH^{2+} group of Pro1 and the side chain carboxyl group of Asp51, which stabilizes this motif in HIV [49], has an equivalent in our FIV CA structure. Indeed, we observe a salt bridge between the terminal NH^{2+} group of Thr1 and the side chain carboxyl group of Asp50 (Figure 5), with about the same bond length than that observed for the Pro-Asp salt bridge in HIV CA (2.6 Å vs 2.8 Å, respectively). Studying the functionality of this mutant Pro1Thr CA mutant for FIV or HIV replication is beyond the scope of our article but could help understanding if the key function of this proline is indeed to induce the formation of the β-hairpin (which is not suggested by our data) or if its main role is for the Gag precursor to adopt the best conformation at the MA-CA junction for efficient protease cleavage during maturation.

The CypA-BL present in all retroviral capsid protein is also observed in the FIV CA_{NTD}. Interestingly, we could detect the presence of a *cis* Arg89-Pro90 peptide bond (Figure 4) in FIV CA CypA-BL. As CypA is a *cis-trans* peptidylprolyl isomerase with a strong specificity for *cis*Pro [50], the presence of this *cis*-proline could explain why the Pro90 residue has been shown to be essential for the binding of cyclophilin A to FIV CA [51]. Notably, no *cis*Pro has been observed in the structures of HIV CypA-BL.

The last two residues of the CA_{CTD} domain are not defined in the electron density. This confirms the high flexibility of the C-terminal end of FIV CA protein, which we had already truncated by 11 residues to avoid problems with crystallogenesis. This C-terminal end might be flexible to allow the correct conformation of the CA-NC cleavage site of the Gag poplyprotein, as what was hypothesize for the flexibility of the FIV MA C-terminus in the MA-CA cleavage process [52].

Altogether, our results show that the various domains which have been involved in key functions of retroviral CA, or which have been observed as important for FIV replication, are present in our FIV CA structure although with their own specificities.

5. Conclusions

We determined the first atomic structure of full-length FIV CA to 1.9 Å resolution. The monomeric FIV CA is a functional capsid protein and displays a standard α-helical CA topology with a short linker between its N- and C-terminal domains. FIV CA also harbors original features like its *cis* Arg89-Pro90 bond, which is visible for the first time in the structure of a retroviral CA. Moreover, despite the absence of the conserved N-terminal proline in our FIV CA construct, the amino-terminal β-hairpin motif is formed in a similar conformation than that of HIV CA, although it is longer in FIV CA. How these features play a role in the differences observed in CA assembly *in vitro* [30], [34] remains to be determined by the elucidation of the structure of high-order oligomers of FIV CA. Altogether, the crystallographic structure of FIV CA represents, in its monomeric form, a functional capsid protein with original features when compared to HIV and EIAV.

Supplementary Materials: The following are available online at www.mdpi.com/link.

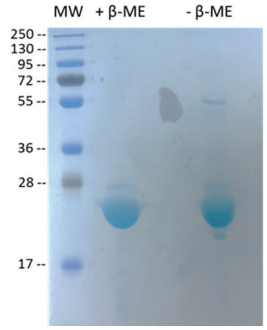

Figure S1: Absence of a disulfide bridge between CA monomers in solution. One microliter of FIV CA protein was loaded in presence (+ βME) or absence (- βME) of reducing agent β-mercaptoethanol on a SDS-PAGE

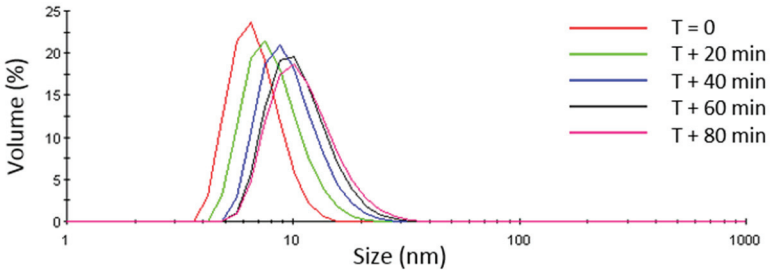

Figure S2: Analysis of FIV CA protein by Dynamic Light Scattering (DLS). Assembly kinetics of FIV p24EΔCP-T protein at 7 mg/mL in presence of NaCl 1M [34].

Acknowledgments: We thank the staff of the Protein Science Facility (PSF) of the SFR BioSciences Gerland-Lyon Sud and especially Isabelle Zanella-Cléon for Mass-Spectrometry analyzes. We also thank the staff at the ID30A-3 beamline of European Synchrotron Radiation Facility (ESRF, Grenoble, France). Finally, we would like to thank L. Ballut and L. Carrique for the data collection at the ESRF. C. Folio benefits from a fellowship from the Ministère de l'Enseignement Supérieur et de la Recherche and M. Dujardin from a fellowship from Agence Nationale de la Recherche (ANR). This work is supported by the ECOS-Sud program (action U14S01).

Author Contributions: C.F. and C.G. designed the studies. C.F. and N.S. purified and analyzed the CA protein. C.F., M.D., G.A. and C.G. performed the crystallographic experiments. C.F. and C.G. analyzed the diffraction data and the CA structure and wrote the manuscript. Finally, all authors reviewed the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, J., J.L. Malmberg, B.A. Wood, S. Hladky, R. Troyer, M. Roelke, M. Cunningham, R. McBride, W. Vickers, W. Boyce, E. Boydston, L. Serieys, S. Riley, K. Crooks, and S. VandeWoude, *Feline immunodeficiency virus cross-species transmission: Implications for emergence of new lentiviral infections*, J Virol, 2016. **91**.

2. Bendinelli, M., M. Pistello, S. Lombardi, A. Poli, C. Garzelli, D. Matteucci, L. Ceccherini-Nelli, G. Malvaldi, and F. Tozzini, *Feline immunodeficiency virus: an interesting model for AIDS studies and an important cat pathogen*, Clin Microbiol Rev, 1995. **8**(1): p. 87-112.

3. Troyer, J.L., J. Pecon-Slattery, M.E. Roelke, W. Johnson, S. VandeWoude, N. Vazquez-Salat, M. Brown, L. Frank, R. Woodroffe, C. Winterbach, H. Winterbach, G. Hemson, M. Bush, K.A. Alexander, E. Revilla, and S.J. O'Brien, *Seroprevalence and genomic divergence of circulating strains of feline immunodeficiency virus among Felidae and Hyaenidae species*, J Virol, 2005. **79**(13): p. 8282-94.

4. Clements, J.E. and M.C. Zink, *Molecular biology and pathogenesis of animal lentivirus infections*, Clin Microbiol Rev, 1996. **9**(1): p. 100-17.

5. Luttge, B.G. and E.O. Freed, *FIV Gag: virus assembly and host-cell interactions*, Vet Immunol Immunopathol, 2010. **134**(1-2): p. 3-13.

6. Burkhard, M.J. and G.A. Dean, *Transmission and immunopathogenesis of FIV in cats as a model for HIV*, Curr HIV Res, 2003. **1**(1): p. 15-29.

7. Elder, J.H., Y.C. Lin, E. Fink, and C.K. Grant, *Feline immunodeficiency virus (FIV) as a model for study of lentivirus infections: parallels with HIV*, Curr HIV Res, 2010. **8**(1): p. 73-80.

8. Kenyon, J.C. and A.M. Lever, *The molecular biology of feline immunodeficiency virus (FIV)*, Viruses, 2011. **3**(11): p. 2192-213.

9. Abdusetir Cerfoglio, J.C., S.A. Gonzalez, and J.L. Affranchino, *Structural elements in the Gag polyprotein of feline immunodeficiency virus involved in Gag self-association and assembly*, J Gen Virol, 2014. **95**(Pt 9): p. 2050-9.

10. Freed, E.O., *HIV-1 gag proteins: diverse functions in the virus life cycle*, Virology, 1998. **251**(1): p. 1-15.

11. Pal, R., M.S. Reitz, Jr., E. Tschachler, R.C. Gallo, M.G. Sarngadharan, and F.D. Veronese, *Myristoylation of gag proteins of HIV-1 plays an important role in virus assembly*, AIDS Res Hum Retroviruses, 1990. **6**(6): p. 721-30.

12. Provitera, P., F. Bouamr, D. Murray, C. Carter, and S. Scarlata, *Binding of equine infectious anemia virus matrix protein to membrane bilayers involves multiple interactions*, J Mol Biol, 2000. **296**(3): p. 887-98.

13. Elder, J.H., M. Schnolzer, C.S. Hasselkus-Light, M. Henson, D.A. Lerner, T.R. Phillips, P.C. Wagaman, and S.B. Kent, *Identification of proteolytic processing sites within the Gag and Pol polyproteins of feline immunodeficiency virus*, J Virol, 1993. **67**(4): p. 1869-76.

14. Brown, L.A., C. Cox, J. Baptiste, H. Summers, R. Button, K. Bahlow, V. Spurrier, J. Kyser, B.G. Luttge, L. Kuo, E.O. Freed, and M.F. Summers, *NMR structure of the myristylated feline immunodeficiency virus matrix protein*, Viruses, 2015. **7**(5): p. 2210-29.

15. Provitera, P., R. El-Maghrabi, and S. Scarlata, *The effect of HIV-1 Gag myristoylation on membrane binding*, Biophys Chem, 2006. **119**(1): p. 23-32.

16. Erickson-Viitanen, S., J. Manfredi, P. Viitanen, D.E. Tribe, R. Tritch, C.A. Hutchison, 3rd, D.D. Loeb, and R. Swanstrom, *Cleavage of HIV-1 gag polyprotein synthesized in vitro: sequential cleavage by the viral protease*, AIDS Res Hum Retroviruses, 1989. **5**(6): p. 577-91.

17. Campbell, E.M. and T.J. Hope, *HIV-1 capsid: the multifaceted key player in HIV-1 infection*, Nat Rev Microbiol, 2015. **13**(8): p. 471-83.

18. Ganser-Pornillos, B.K., U.K. von Schwedler, K.M. Stray, C. Aiken, and W.I. Sundquist, *Assembly properties of the human immunodeficiency virus type 1 CA protein*, J Virol, 2004. **78**(5): p. 2545-52.

19. Nath, M.D. and D.L. Peterson, *In vitro assembly of feline immunodeficiency virus capsid protein: biological role of conserved cysteines*, Arch Biochem Biophys, 2001. **392**(2): p. 287-94.

20. Gelderblom, H.R., M. Ozel, and G. Pauli, *Morphogenesis and morphology of HIV. Structure-function relations*, Arch Virol, 1989. **106**(1-2): p. 1-13.

21. Tang, C., E. Loeliger, I. Kinde, S. Kyere, K. Mayo, E. Barklis, Y. Sun, M. Huang, and M.F. Summers, *Antiviral inhibition of the HIV-1 capsid protein*, J Mol Biol, 2003. **327**(5): p. 1013-20.

22. Jin, Z., L. Jin, D.L. Peterson, and C.L. Lawson, *Model for lentivirus capsid core assembly based on crystal dimers of EIAV p26*, J Mol Biol, 1999. **286**(1): p. 83-93.

23. Momany, C., L.C. Kovari, A.J. Prongay, W. Keller, R.K. Gitti, B.M. Lee, A.E. Gorbalenya, L. Tong, J. McClure, L.S. Ehrlich, M.F. Summers, C. Carter, and M.G. Rossmann, *Crystal structure of dimeric HIV-1 capsid protein*, Nat Struct Biol, 1996. **3**(9): p. 763-70.

24. Pornillos, O., B.K. Ganser-Pornillos, and M. Yeager, *Atomic-level modelling of the HIV capsid*, Nature, 2011. **469**(7330): p. 424-7.

25. Mattei, S., B. Glass, W.J. Hagen, H.G. Krausslich, and J.A. Briggs, *The structure and flexibility of conical HIV-1 capsids determined within intact virions*, Science, 2016. **354**(6318): p. 1434-1437.

26. Jacques, D.A., W.A. McEwan, L. Hilditch, A.J. Price, G.J. Towers, and L.C. James, *HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis*, Nature, 2016. **536**(7616): p. 349-53.

27. Ganser-Pornillos, B.K., M. Yeager, and W.I. Sundquist, *The structural biology of HIV assembly*, Curr Opin Struct Biol, 2008. **18**(2): p. 203-17.

28. Pornillos, O., B.K. Ganser-Pornillos, S. Banumathi, Y. Hua, and M. Yeager, *Disulfide bond stabilization of the hexameric capsomer of human immunodeficiency virus*, J Mol Biol, 2010. **401**(5): p. 985-95.

29. Berthet-Colominas, C., S. Monaco, A. Novelli, G. Sibai, F. Mallet, and S. Cusack, *Head-to-tail dimers and interdomain flexibility revealed by the crystal structure of HIV-1 capsid protein (p24) complexed with a monoclonal antibody Fab*, Embo j, 1999. **18**(5): p. 1124-36.

30. Bharat, T.A., L.R. Castillo Menendez, W.J. Hagen, V. Lux, S. Igonet, M. Schorb, F.K. Schur, H.G. Krausslich, and J.A. Briggs, *Cryo-electron microscopy of tubular arrays of HIV-1 Gag resolves structures essential for immature virus assembly*, Proc Natl Acad Sci U S A, 2014. **111**(22): p. 8233-8.

31. Mammano, F., A. Ohagen, S. Hoglund, and H.G. Gottlinger, *Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis*, J Virol, 1994. **68**(8): p. 4927-36.

32. Bocanegra, R., M.A. Fuertes, A. Rodriguez-Huete, J.L. Neira, and M.G. Mateu, *Biophysical analysis of the MHR motif in folding and domain swapping of the HIV capsid protein C-terminal domain*, Biophys J, 2015. **108**(2): p. 338-49.

33. Khwaja, A., M. Galilee, A. Marx, and A. Alian, *Structure of FIV capsid C-terminal domain demonstrates lentiviral evasion of genetic fragility by coevolved substitutions*, Sci Rep, 2016. **6**: p. 24957.

34. Serriere, J., D. Fenel, G. Schoehn, P. Gouet, and C. Guillon, *Biophysical characterization of the feline immunodeficiency virus p24 capsid protein conformation and in vitro capsid assembly*, PLoS One, 2013. **8**(2): p. e56424.

35. Kabsch, W., *XDS*, Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.

36. Keegan, R.M. and M.D. Winn, *MrBUMP: an automated pipeline for molecular replacement*, Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 119-24.

37. Winn, M.D., C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, and K.S. Wilson, *Overview of the CCP4 suite and current developments*, Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 235-42.

38. Goldstone, D.C., M.W. Yap, L.E. Robertson, L.F. Haire, W.R. Taylor, A. Katzourakis, J.P. Stoye, and I.A. Taylor, *Structural and functional analysis of prehistoric lentiviruses uncovers an ancient molecular interface*, Cell Host Microbe, 2010. **8**(3): p. 248-59.

39. Adams, P.D., P.V. Afonine, G. Bunkóczi, V.B. Chen, N. Echols, J.J. Headd, L.-W. Hung, S. Jain, G.J. Kapral, R.W. Grosse Kunstleve, A.J. McCoy, N.W. Moriarty, R.D. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, and P.H. Zwart, *The Phenix software for automated determination of macromolecular structures*, Methods, 2011. **55**(1): p. 94-106.

40. Emsley, P., B. Lohkamp, W.G. Scott, and K. Cowtan, *Features and development of Coot*, Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

41. L., D.W., *The PyMol molecular graphics systems*, 2008.

42. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state*, J Mol Biol, 2007. **372**(3): p. 774-97.

43. Du, S., L. Betts, R. Yang, H. Shi, J. Concel, J. Ahn, C. Aiken, P. Zhang, and J.I. Yeh, *Structure of the HIV-1 Full-Length Capsid in a Conformationally-Trapped Unassembled State Induced by Small-Molecule Binding*, J Mol Biol, 2011. **406**(3): p. 371-86.

44. Zhao, G., J.R. Perilla, E.L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A.M. Gronenborn, K. Schulten, C. Aiken, and P. Zhang, *Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics*, Nature, 2013. **497**(7451): p. 643-6.

45. Gres, A.T., K.A. Kirby, V.N. KewalRamani, J.J. Tanner, O. Pornillos, and S.G. Sarafianos, *X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability*, Science, 2015. **349**(6243): p. 99-103.

46. Liu, S., L. Zhou, L. Chen, S.G. Dastidar, C. Verma, J. Li, D. Tan, and R. Beuerman, *Effect of structural parameters of peptides on dimer formation and highly oxidized side products in the oxidation of thiols of linear analogues of human beta-defensin 3 by DMSO*, J Pept Sci, 2009. **15**(2): p. 95-106.

47. Gamble, T.R., S. Yoo, F.F. Vajdos, U.K. von Schwedler, D.K. Worthylake, H. Wang, J.P. McCutcheon, W.I. Sundquist, and C.P. Hill, *Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein*, Science, 1997. **278**(5339): p. 849-53.

48. Tang, C., Y. Ndassa, and M.F. Summers, *Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein*, Nat Struct Biol, 2002. **9**(7): p. 537-43.

49. Cortines, J.R., E.B. Monroe, S. Kang, and P.E. Prevelige, *A retroviral chimeric capsid protein reveals the role of the N-terminal β-hairpin in mature core assembly*, J Mol Biol, 2011. **410**(4): p. 641-52.

50. Konno, M., M. Ito, T. Hayano, and N. Takahashi, *The substrate-binding site in Escherichia coli cyclophilin A preferably recognizes a cis-proline isomer or a highly distorted form of the trans isomer*, J Mol Biol, 1996. **256**(5): p. 897-908.

51. Lin, T.Y. and M. Emerman, *Cyclophilin A interacts with diverse lentiviral capsids*, Retrovirology, 2006. **3**: p. 70.

52. Serriere, J., X. Robert, M. Perez, P. Gouet, and C. Guillon, *Biophysical characterization and crystal structure of the Feline Immunodeficiency Virus p15 matrix protein*, Retrovirology, 2013. **10**: p. 64.



© 2017 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

RESUMES

Études fonctionnelle et structurale de deux protéines rétrovirales d'intérêt thérapeutique : la protéine Tax du virus HTLV et la protéine de capside du FIV.

Les rétrovirus sont un enjeu de santé publique, aussi bien humaine qu'animale. La compréhension des déterminants structuraux sous-jacents à la fonction de leurs protéines constitue une étape essentielle dans le développement de stratégies antirétrovirales efficaces.

Ce manuscrit porte sur l'étude des bases structurales des mécanismes moléculaires impliqués dans les fonctions clés des rétrovirus que sont i) la régulation de l'expression des protéines de rétrovirus complexes et ii) l'assemblage des particules virales, à travers l'étude de deux protéines rétrovirales d'intérêt thérapeutique : la protéine Tax du virus T-lymphotrope humain (HTLV) et la protéine de capside du virus de l'immunodéficience féline (FIV). L'étude structurale de ces deux protéines d'intérêt et la compréhension des mécanismes moléculaires nécessaires à leurs fonctions permettraient d'ouvrir la voie à la conception de nouvelles stratégies antirétrovirales.

Malgré de nombreux tests d'expression et de purification, l'étude structurale de la protéine Tax du HTLV n'a pu être réalisée, en raison de son insolubité. Cependant, ce travail doctoral a permis de résoudre, pour la première fois, la structure cristallographique de la protéine de capside entière du FIV. Bien que cette dernière adopte un repliement similaire aux autres capsides rétrovirales dont la structure est connue, elle présente également des spécificités structurales dont les conséquences fonctionnelles seront discutées.

Functional and structural studies of two retroviral proteins of therapeutic interest: the HTLV Tax protein and the FIV capsid protein.

Retroviruses are a major concern of public health in humans but also in animals. A better understanding of the structural determinants underlying the functions of retroviral proteins is a crucial step for the development of efficient antiretroviral therapies.

This manuscript studies the structural basis of the molecular mechanisms implicated in key functions of retroviruses such as, i) the regulation of complex retroviruses protein expression and ii) the assembly of viral particles, through the study of two retroviral proteins of therapeutic interest: the human T-lymphotropic virus (HTLV) Tax protein and the feline immunodeficiency virus (FIV) capsid protein. The functional and structural studies of these two proteins and the understanding of the molecular mechanisms required for their functions will pave the way to the conception of new antiretroviral therapeutic strategies.

Despite several expression and purification assays, no structural studies could be performed for the HLTV Tax protein. However, this study allowed the resolution of the first structure for the full-length FIV capsid protein by X-ray crystallography. Although the FIV capsid protein displays a standard α-helical topology like other retroviral CAs, it also harbors original features whose functional consequences will be discussed.