

**ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE**

**Institut de Biologie Moléculaire et Cellulaire**  
UPR9002 – Architecture et Réactivité de l'ARN

**Institut de Génétique et Biologie Moléculaire et Cellulaire**  
UMR71105 – Biologie Structurale Intégrative

**THÈSE\***

présentée par : **Charlotte BUSSIENNE**

Soutenue le : **27 janvier 2023**

Pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : science de la vie, virologie aspect moléculaire et médicaux

**Etude structurale et fonctionnelle de la protéine Pr55Gag du VIH-1 en association avec le signal d'encapsidation de l'ARN génomique**

**THÈSE dirigée par :**

**Dr Serena BERNACCHI**  
**Dr Marc RUFF**

Directrice de recherches, UPR9002, IBMC, Université de Strasbourg  
Directeur de recherches, UMR7104, IGBMC, Université de Strasbourg

**RAPPORTEURS :**

**Dr Carine TISNE**  
**Dr Olivier MAUFFRET**

Directrice de recherches, UMR8261, IBPC, Université de Paris  
Directeur de recherches, UMR8113, LBPA, Université de Paris Saclay

**AUTRES MEMBRES DU JURY :**

**Dr Maria DIMITROVA**  
**Dr Serge BOUAZIZ**

Professeur, UPR2357, IBMP, Université de Strasbourg  
Directeur de recherches, UMR8038, CiTCoM, Université de Paris

**MEMBRE INVITÉ :**

**Dr Emmanuel BOUTANT**

Maître de conférences, UMR7021, LBP, Université de Strasbourg



À la mémoire de ma mère et de mon père.



# Remerciements

Tout d'abord, je tiens à remercier les membres de mon jury Dr Carine Tisné, Dr Olivier Maufret pour avoir accepté d'évaluer mon travail de doctorat en tant que rapporteur. Un remerciement spécial aux Dr Serges Bouaziz et Dr Emmanuel Boutant pour avoir accepté de suivre mon travail en tant que membres de mon CST au cours de ces trois années. A toi, Maria Dimitrova, pour avoir été un mentor depuis ma licence et d'avoir accepté le rôle de présidente de jury de thèse.

Je remercie Serena, ma directrice de thèse, pour son soutien, sa disponibilité et sa patience et Marc Ruff, mon directeur de thèse, pour ses conseils, sa bonne humeur et sa compréhension.

Merci aux membres de mes deux équipes pour avoir égayé mes journées de travail, répondu à mes questions et m'avoir permis de trouver ma place en tant que chercheuse. Valérie (Lamour) de m'avoir fait confiance pour l'enseignement, c'était une expérience tellement enrichissante. Julien pour ta formation, ta patience, tes conseils, les heures passées à écouter mes bêtises... et à te moquer. Sylvia, d'avoir été une oreille prête à me conseiller en toutes circonstances. Sylvie pour avoir préparé le meilleur APS qui soit. Claire pour toutes les petites attentions et pour avoir remboursé mes dettes en t7. Merci aussi Rui, Marlène, Philippe pour les moments partagés, les secret santa et les labmeeting. Merci Jean-Christophe et Roland de m'avoir posé des questions scientifiques et avoir contribué à développer mon sens critique. Anne pour ta gentillesse et ta bonne humeur. Valérie pour les discussions scientifiques et ta bienveillance à toute épreuve. Benjamin pour avoir été une oreille attentive. Antoine pour les croissants. Anne-Caroline pour ta gentillesse, pour toujours être prête à aider et à trouver des solutions. Je souhaite également beaucoup d'épanouissement aux nouveaux arrivants.

Après presque quatre ans passés entre les deux instituts, je souhaite remercier toutes les personnes que j'ai rencontré et particulièrement celles qui m'ont permis de mettre au point, d'avancer et d'améliorer mon projet de thèse. Lauriane, pour ta patience et ton envie de partager m'ayant fait apprécier chacun de nos échanges. Johanna et Philippe pour m'avoir permis de découvrir les PTM de Gag. Catherine, pour m'avoir tellement appris en biophysique. Pierre pour ta patience. Corinne pour les heures passées à screener les conditions sur le Glacios dans le froid. Mais aussi merci à toutes ces personnes que j'ai croisé, avec qui j'ai partagé un congrès, un café, une formation, un concours et qui m'ont permis de réfléchir, de remettre en question mon travail ou de trouver une nouvelle motivation.

Un merci spécial à Camille et à Auguste. Je suis tellement heureuse d'avoir ouvert la porte de votre bureau et d'avoir rencontré Anna, Laura et Albana. Merci à vous l'équipe pour les moments inoubliables, vivement les prochains.

Thank you Elenia, to become a friend and definitively a style model.

Merci au Gag Crew. Une thèse sans toi, finalement, ça n'aurait pas été la même chose. Pas de reveil à 4h au synchrotron pour voir qu'on a perdu les électrons, pas de saqué, pas vidéos nulles ... Bon courage pour la suite 😊

Un merci spécial à ma Dream Team : Dr Erwan pour m'avoir sauvé la vie plus d'une fois, Dr Orian pour avoir essayé de me tuer plus d'une fois et à toi futur dr Cédric, mon ami de toujours depuis le master. Merci car j'ai passé des moments incroyables, merveilleux et un peu nuls, grâce à vous. PS : Cédric, Orian a tort, n'arrête jamais les blagues nulles.

Merci aux copains pour les randonnées, les pizzas, les raquettes, les verres de vin, Chloé, Adrien, Pierro, Fanny et Valou.

Avant de conclure, un merci à mes amis : Alexia pour les heures passées au téléphone à refaire le monde, Delphine d'être toujours à mes côtés et prête à rendre service à la terre entière. À toi Claire, que je suis tellement heureuse d'avoir rencontrée, reste comme tu es, la personne la plus gentille de la terre. Elenia finalement heureusement qu'il n'y a que les « idiots » qui ne changent pas d'avis. A mon bro, Louis de me prouver que la distance et le temps ne changera jamais notre amitié.

A ma famille pour le soutien et leur amour inconditionnel. Merci aux petites mains secrètes qui ont corrigé et relu mon manuscrit (Magali & Caroline – désolé ce n'est plus un secret). A Cabou, d'être ma béquille et à toi Fred de me soutenir dans mes projets les plus ambitieux.

Merci à mes beaux-parents préférés : Gauthier et Anne, pour m'avoir accueilli à bras ouverts et avoir été d'un grand soutien, particulièrement ces dernières années.

A Virginie pour ta bienveillance et aux deux monstres, merci d'être des monstres, Martin et Emilie.

Merci à toi Lauréliane, d'être toujours là à travers le temps et l'espace.

Catherine, pour ton soutien et ton amour inconditionnel.

Un énorme merci à toi Nicolas, merci d'avoir et d'être ma bequille, merci pour tout, vraiment ! Je ne serais pas là où je suis, ni qui je suis, sans toi à mes cotés.

Et pour finir, je tiens à remercier mon meilleur ami, mon âme sœur, Loïc. Je sais que j'avais dit que je ne te citerai pas ... Mais, merci pour ton support, ton amour, tes petits plats... surtout tes petits plats qui m'ont aidé, fait grandir et permis d'être où je suis aujourd'hui.





# Table des matières

Remerciements.....	5
Table des matières.....	9
Abbréviations.....	13
Liste des tableaux.....	17
Liste des figures.....	19
Liste des annexes.....	23
Introduction.....	25
I. La grande famille des Rétrovirus (Retroviridae).....	27
I.1 Les principales découvertes autour des rétrovirus.....	27
I.2 Classification phylogénique des rétrovirus.....	27
II. Un rétrovirus particulier : le virus de l'immunodéficience humaine (VIH).....	28
II.1 La découverte d'un nouveau virus : le VIH.....	28
II.2 Les Origines du VIH.....	29
II.3 Les chiffres clés dans le monde : données en 2021.....	31
III. Cycle viral du VIH-1.....	32
III.1 La particule virale, structure et composition du génome.....	32
III.2 Les protéines virales essentielles : Gag, Pol et Env.....	34
III.3 Les protéines virales auxiliaires.....	41
III.4 Le cycle de réplication viral.....	43
IV. Progression de la maladie et thérapie.....	50
IV.1 Transmission et manifestation clinique du virus.....	50
IV.2 Restriction cellulaire et mécanisme d'échappement du virus.....	52
IV.3 Prévention, dépistage et traitements.....	55
V. Des étapes particulières : L'Assemblage du virus.....	60
V.1 Sélection spécifique de l'ARNg par Gag.....	60
V.2 Assemblage du virion à la membrane plasmique.....	64
Objectifs.....	69
Matériels et Méthodes.....	73
I. Séquences et cartes des plasmides codants pour les fragments d'ARNg du VIH-1.....	75
II. Production des fragments d'ARNg du VIH-1 <i>in vitro</i> .....	76
II.1 Amplification des plasmides.....	76
II.2 Linéarisation des plasmides amplifiés.....	77
II.3 Transcription des plasmides linéarisés.....	77
II.4 Purification des fragments d'ARN transcrits.....	78
III. Caractérisation de l'ARN du VIH-1 <i>in vitro</i> .....	79
III.1 Repliement de l'ARNg viral et contrôle de la dimérisation.....	79
III.2 Marquage des ARN viraux.....	79
IV. Séquence en acides aminés des protéines recombinantes Gag et des partenaires de Gag.....	80
IV.1 Protéines Gag produites en système bactérien.....	80
IV.2 Protéines Vpr produites par synthèse chimique.....	81
IV.3 Séquences des protéines produites en cellules de mammifères.....	81
V. Production des vecteurs viraux recombinants pour l'expression des protéines en système eucaryote (BHK21). .....	82

V.1 Culture cellulaire .....	83
V.2 Production et sélection des vecteurs recombinants .....	83
V.3 Contrôle de l'expression des nouveaux vecteurs recombinants et production des protéines recombinantes .....	85
VI. Purification des protéines recombinantes en cellules de mammifères .....	86
VI.1 Procédures communes à la purification des protéines produites en cellules de mammifères .....	86
VI.2 Purification des protéines recombinantes en fonction de leur étiquette .....	87
VII. Caractérisation des protéines purifiées .....	89
VII.1 Détection des protéines recombinantes par Western Blot (WB).....	89
VII.2 Contrôle de la séquence et des PTM des protéines par spectrométrie de masse en tandem (MS/MS) ...	90
VII.3 Echange de solutions tamponnées par dialyse des protéines 6His-Gag purifiées .....	92
VII.4 Marquage fluorescent des protéines Gag purifiées .....	93
VIII. Formation des complexes entre Gag et ses partenaires protéiques et nucléiques .....	93
VIII.1 Formation des complexes entre la protéine 6His-Gag et l'ARN111 <sub>WT</sub> .....	93
VIII.2 Formation des complexes entre 6His-Gag et ses partenaires protéiques (ALIX, Vpr).....	94
IX Méthodes biophysiques d'analyse de l'ARNg, de Gag et des complexes .....	94
IX.1 Diffusion dynamique de la lumière (DLS) .....	94
IX.2 Diffusion de lumière multi-angle couplée à une chromatographie à exclusion de taille (SEC-MALS) .....	96
IX.3 Analyse de la dispersion induite par flux (FIDAbio) .....	98
IX.4 Ultracentrifugation analytique (AUC).....	99
IX.5 Nano-fluorimétrie à balayage différentiel (NanoDSF) .....	101
IX.6 Caractérisation des complexes entre 6His-Gag et les ARN viraux par anisotropie de fluorescence .....	102
X. Etudes Structurales .....	103
X.1 Résolution de la structure de 6His-Gag par microscopie électronique à transmission .....	103
X.3 Modélisation de Gag par AlphaFold.....	106
X.4 Tests de cristallisation de 6His-Gag .....	106
Résultats et discussion objectif n°1 .....	107
I. Résultats : caractérisation des interactions <i>in vitro</i> entre l'ARNg et Gag.....	109
I.1. Purification de 6His-Gag produite et sa caractérisation par des méthodes biophysiques .....	109
I.2. Analyse de l'interaction de 6His-Gag avec la région 5'UTR de l'ARNg viral <i>in vitro</i> .....	125
I. 3. Analyse de la structure de Gag en complexe avec l'ARN.....	139
II. Discussion : caractérisation des interactions <i>in vitro</i> entre le précurseur Gag et ses partenaires protéiques .....	151
I.1. Purification de 6His-Gag et sa caractérisation par des méthodes biophysiques .....	151
II.2. Analyse de l'interaction de 6His-Gag avec la région 5'UTR de l'ARNg viral <i>in vitro</i> .....	157
II.3. Analyse structurale de Gag en complexes avec l'ARN.....	161
Résultats et discussion objectif n°2 .....	165
I. Résultats Objectif n°2 : Caractérisation des interactions <i>in vitro</i> entre Gag et ses partenaires protéiques ...	167
I.1. Production des vecteurs viraux permettant l'expression des partenaires de Gag en cellules de mammifères .....	167
I.2. Purification et caractérisations biophysiques des partenaires ALIX et Vpr produits en cellules de mammifères .....	172
I.3 Formation et caractérisation des complexes entre 6His-Gag et ses partenaires protéiques .....	179
II. Discussion Objectif n°2 : Caractérisation des interactions <i>in vitro</i> entre Gag et ses partenaires protéiques	184

I.1. Production des vecteurs viraux permettant l'expression des partenaires de Gag en cellules de mammifères .....	185
I.2 Purification et caractérisations biophysiques des partenaires ALIX et Vpr produits en cellules de mammifères .....	186
I.3. Formation et caractérisation des complexes entre 6His-Gag et ses partenaires protéiques .....	190
Conclusion & perspectives .....	193
Bibliographie.....	201
Annexes .....	215
Charlotte BUSSIENNE.....	253
Résumé .....	253



# Abbréviations

**kDa** : Kilodalton

**+ssDNA** : *Strand strong-stop DNA*

**µg** : Microgramme

**µL** : Microlitre

**µm** : Micromètre

**µM** : Micromolaire

A

**Å** : Angström

**AC@** : Anticorps dirigé contre

**ADN** : Acide desoxyribonucléique

**ADN sb** : ADN simple brin

**ALIX** : *ALG-2 Interacting Protein X*

**ALLINI** : Inhibiteurs allostériques de l'intégrase

**ALV** : Virus de la leucose aviaire

**A<sub>nnn</sub>** : Absorbance

**APOBEC** : *Apolipoproteins B mRNA editing enzyme, catalytic polypeptide-like*

**ARN** : acide Ribonucléique

**ARNg** : ARN génomique

**ARNt** : ARN de transfert

**ARN111<sub>AAA</sub>** : ARN mutant au niveau de la région de dimérisation

**ARN111<sub>WT</sub>** : ARN de 111 nucléotide sauvage

**ARNSL1<sub>35</sub>** : ARN correspondant à la tiges boucle SL1 de 35 nucléotides

**ARNt lys 3** : ARNt de la lysine, isoforme 3

**ART** : Aantiretroviraux

**ATP** : Adenosine triphosphate

**AUC** : Ultracentrifugation analytique

B

**BET** : Bromure d'éthidium

**BHK21** : Cellules de reins de hamsters

**BMH** : *Branched Multiple Hairpins*

**BSA** : *Bovin serum Albumin*

**BTP** : Bacto tryptone phosphate

C

**C** : Cystéine

**c(s)** : Coefficient de sédimentation

**CA** : Capside

**CCR5** : *C-C chemokine receptor type 5*

**CD4** : *Cluster of differentiation 4*

**Cell** : Cellules

**CHAPS** : 3-[(3-cholamidopropyl)diméthylammonio]-1-propanesulfonate

**CHMP4** : Protéine de corps multivésiculaire chargée

**CMH** : Complexe majeur d'histocompatibilité

**Cmol** : Concentration molaire

**CO** : Fractions concentrées

**CPA** : Cellules présentatrices d'antigènes

**CRF** : Forme circulante recombinante

**CRM1** : *Chromosomal maintenance 1*

**CryoEM** : Cryo microscopie électronique

**CXCR4** : *C-X-C chemokine receptor type 4*

**CypA** : Cyclophiline A

D

**DC** : Cellules dendritiques

**DDM** : Dodécyl-β-D-maltoside

**DIS** : Site d'initiation de la dimérisation

**DLS** : *Dynamic light scattering*

**dn/dc** : Incrément d'indice de réfraction spécifique

**dNTP** : Désoxynucléotides triphosphates

**DTT** : Dithiothréitol

E

**e<sup>-</sup>** : Electrons

**EDTA** : Acide éthylènediaminetétraacétique

**Eff.** : Effluent

**ELISA** : Enzyme-linked immunosorbent assay

**Env** : enveloppe

**ESCRT** : Endosomal Sorting Complex Required for Transport

**ET** : Extrait total

**ε<sub>th</sub>** : Epsilon théorique

F

**f/f<sub>0</sub>** : Coefficient de friction

**FBS** : Sérum de veau fœtal

**FDA** : *Food and Drug Administration*

**FIDABio** : Analyse de la dispersion induite par flux

**FTSC** : Fluorescéine 5-thiosemicarbazide

G

**G** : Guanine

**Gag** : *Group-specific antigen*

**GD** : *Glow discharged*, décharges lumineuses

**GFP** : *Green fluorescent protein*

**GMEM** : Milieu minimum essentiel de Glasgow

**Gp** : Glycoprotéine

**GPT** : Guanine phosphoribosyl transferase

**Groupe M** : Majoritaire

**Groupe N** : Non -M, non -O

**Groupe O** : Outlier

H

**HA** : Hémagglutinine

**HBR** : Région hautement basique

**HeNe** : Hélium neon

**His** : Histidine

**HPLC** : Chromatographie en phase liquide haute performance

**HRP** : Peroxydase de raifort

**HSH** : Hommes ayant des rapports sexuels avec d'autres hommes

I

**IF** : inhibiteur de fusion

**Ig-** : Immunoglobuline  
**II** : Inhibiteur d'intégrase  
**IN** : Intégrase  
**INNTI** : Inhibiteurs non nucléosidiques de la transcription inverse  
**INTI** : Inhibiteurs nucléosidiques de la transcription inverse  
**IP** : Inhibiteur de protéase  
**IPTG** : Isopropyl  $\beta$ -D-1-thiogalactopyranoside  
**ISG** : Interferon-Stimulated Gene  
**ITC** : *Isothermal titration calorimetry*

K

**kb** : Kilo base  
**kDa** : Kilo Dalton  
**K** : Lysine

L

**L** : Litre  
**LB** : Lysogeny broth  
**LEDGF** : *Lens epithelium-derived growth factor*  
**LT** : Lymphocyte T  
**LTR** : Longues répétitions terminales  
**Lys, K** : Lysine

M

**M** : Molaire  
**m/z** : Masse-sur-charge  
**MA** : Matrice  
**MALS** : Multi angle light scattering  
**March 8** : *Membrane-Associated RING-CH 8 protein*  
**MgCl<sub>2</sub>** : Chlorure de magnésium  
**MIME** : *Mutational interference mapping experiment*  
**mL** : millilitre  
**mM** : miliMolaire  
**MP** : Membrane plasmique  
**MS/MS** : Spectrométrie de masse en tandem  
**MVA** : virus de la vaccine modifié (souche Ankara)  
**MWCO** : *Molecular weight cutoff*

N

**NaCl** : Chlorure de sodium  
**NanoDSF** : Nano-fluorimétrie à balayage différentiel  
**NC** : Nucleocapsid  
**Nef** : Negative Element Factor  
**NFKB** : Nuclear factor kappa-light-chain enhancer of activated B cells  
**nm** : Nanomètre  
**NTP** : Nucléoside Triphosphate  
**NUP** : Nucléoporine

O

**OMS** : organisation mondiale de la santé  
**UV-VIS** : Ultraviolet - visible

P

**PAGE** : Électrophorèse sur gel de polyacrylamide  
**pb** : Paire de base  
**PBS** : Site de fixation de l'amorce (*primer binding site*)  
**pH** : Potentiel hydrogène  
**PI<sub>(4,5)</sub>P2** : Phospholipides phosphatidylinositol-4,5-bisphosphate  
**PIC** : Complexe de préintégration  
**PKR** : Protéine kinase R  
**PM** : Poids moléculaire  
**pmoles** : picomoles  
**Poly A** : Signal de polyadénylation  
**PPT** : Polytract purine  
**Pr** : Protéase  
**PrEP** : Prophylaxie préexposition  
**PRR** : Région riche en proline  
**p-TEFb** : *Positive transcription elongation factor*  
**PTM** : Modification post traductionnelles  
**PVDF** : Polyfluorure de vinylidène

R

**R** : région répétée  
**R-** : Etiquette tween Strep-tag®  
**Région U** : Uniques  
**Rev** : Regulator of expression of virion proteins  
Rh : rayon hydrodynamique  
**RMN** : Résonance magnétique nucléaire  
**RMSD** : *Root mean square deviation*  
**RNase** : Ribonucléase  
**rpm** : *Revolutions per minute.*  
**RRE** : Élément de réponse Rev  
**RT** : Transcriptase inverse

S

**SAMHD1** : *Sterile Alpha Motif and Histidine Aspartate domain-containing protein 1*  
**SAXS** : Diffraction des rayons X aux petits angles  
**SD** : Site donneur d'épissage  
**SDS** : Dodécylsulfate de sodium  
**SEC** : *Size exclusion chromatographie*  
**SIDA** : Syndrome d'immunodéficience acquise  
**SIV** : Virus de l'Immunodéficience Simienne  
**SL** : *stem loop*  
**Sp** : Espaceur  
**SU** : Surface

T

**tampon** : Solution tamponnée  
**TAR** : Séquence d'activation de la transcription  
**Tat** : *Transactivator of Transcription*  
**TBE** : Tris Borate EDTA  
**TBM** : Tris borate Magnésium  
**TK** : Tyrosine kinase  
**TM** : Température de fusion (*Melting temperature*)

**TM** : Transmembranaire  
**TPE** : Traitement post-exposition  
**TRIM5 $\alpha$**  : *Tripartite motif-containing protein isoform*  
**TROD** : Tests rapides à orientation de diagnostique  
**TSG101** : *Tumor Susceptibility Gene 101*

U

**UEV** : *Ubiquitin E2 variant*  
**URF** : Forme unique recombinante  
**UTR** : Région non traduite

V

**v/v** : Volume pour volume  
**vif** : *Viral infectivity factor*  
**VIH** : Virus de l'immunodeficiency humaine  
**VLP** : Particules de types viral viral like particules  
**Vpr** : Protéine virale R (Viral protein R)  
**Vpu** : Protéine virale u

W

**w/v** : Rapport masse sur volume (weight/volume)  
**WB** : Western blot  
**WT** : Sauvage (*wild type*)

Z

**ZF** : Doigts de zinc  
**Zn** : Zinc

\*

**$\epsilon$**  : Coefficient d'extinction  
 **$\lambda$**  : Longueur d'onde  
 **$\Psi$  / psi** : Site d'encapsidation  
**pLDDT** : *predicted local distance difference test*



# Liste des tableaux

Tableau 1. Les séquences des plasmides correspondant à la région 5' UTR de l'ARNg de la souche NL4-3 produits au laboratoire .....	76
Tableau 2. La séquence de l'ARN correspondant à la tige-boucle SL135, commandé chez IDT .....	76
Tableau 3. Informations sur les protéines Gag produites en système bactérien .....	80
Tableau 4. Informations sur les protéines Vpr produites par synthèse chimique .....	81
Tableau 5. Informations sur les protéines 6His-Gag, R-Alix, R-UEV et R-Vpr produites en cellules de mammifères .....	82
Tableau 6. Solutions tamponnées utilisées au cours des purifications des protéines produites en cellules de mammifères .....	89
Tableau 7. Les PTM identifiées par MS/MS à partir de quatre différentes production et/ou purification de protéines 6His-Gag purifiée dans le tampon A1 .....	115
Tableau 8. Données biophysiques de la protéine recombinante 6His-Gag produite en BHK21 purifiée dans les tampons A1 et A3 .....	123
Tableau 9. Valeurs de Cs mesurées par AUC pour l'ARN111WT dans les tampons A1 et A3 .....	127
Tableau 10. Valeurs de Rh déterminés grâce par DLS de la protéine 6His-Gag, des ARN 111WT et ARNt et des complexes protéine/ARN dans différentes dilutions en NaCl de la solution tamponnée A3 .....	131
Tableau 11. Valeurs Cs mesurées en AUC pour l'ARN111WT les différentes solutions tamponnées A1 et A3 .....	133
Tableau 12. Rh moyens mesurés par FIDABio à partir de mélanges de 6His-Gag et d'ARN 111WT en fonction des ratios protéine-ARN .....	134
Tableau 13. Valeurs moyennes d'anisotropie de fluorescence pour des complexes formés entre Gag et l'ARN 111WT en fonction du ratio protéine/ARN .....	136
Tableau 14. Valeurs moyennes d'anisotropie de fluorescence des complexes entre Gag et l'ARN SL135 en fonction des ratios et du nombre de répétitions .....	137
Tableau 15. Valeurs moyennes d'anisotropie de fluorescence obtenus à partir de mélanges de 6His-Gag et d'ARN 111AAA en fonction de la stœchiométrie .....	138
Tableau 16. Récapitulatif des grilles et des conditions qui sont criblées pour réaliser la structure par CryoEM de 6His-Gag .....	143
Tableau 17. PM théoriques des partenaires de p6 et de leurs coefficients d'extinction molaires .....	168
Tableau 18. Les PTM observées par MS/MS pour ALIX produite en BHK21 .....	174
Tableau 19. Caractéristiques des partenaires de p6 en fonction de la solution tamponnée de purification .....	179
Tableau 20. Valeurs moyennes des Rh mesurées par FIDABio de la protéine 6His-Gag marquée dans les différentes solutions tamponnées .....	179
Tableau 21. Valeurs de Rh mesurées par FIDABio de la protéine 6His-Gag marquée en présence d'ALIX en fonction des ratios protéine-protéine .....	180
Tableau 22. Valeurs de Rh déterminées par FIDABio de la protéine 6His-Gag marquée en présence de R-Vpr en fonction des ratios protéine-protéine et des conditions de formation des complexes .....	181
Tableau 23. Valeurs de Rh déterminées par FIDABio de la protéine 6His-Gag marquée en présence de Vpr produite par synthèse chimique en fonction des ratios protéine-protéine .....	182
Tableau 24. Valeurs de Rh déterminées par FIDABio de la protéine Gag-6His produite en E. coli marquée en présence de R-Vpr en fonction des ratios protéine-protéine .....	183



# Liste des figures

Figure 1. Arbre phylogénétique des origines du VIH-1 .....	30
Figure 2. Distribution de la prévalence du VIH-1 dans le monde .....	32
Figure 3. Représentation de la particule virale du VIH-1 .....	32
Figure 4. Représentation du génome du VIH-1 et de l'ADN proviral .....	34
Figure 5. Les différents domaines fonctionnels de la protéine Gag du VIH-1 et leurs multiples rôles .....	35
Figure 6. Les structures tertiaires du domaine MA dans différentes conformations .....	35
Figure 7. Les structures tertiaires du domaine CA dans différentes conformations .....	37
Figure 8. Représentation de la structure tridimensionnelle de la NC .....	37
Figure 9. Représentation de la structure tridimensionnelle de la RT .....	39
Figure 10. Représentation de la structure des domaines de l'IN .....	40
Figure 11. Représentation de la structure d'un trimère d'Env du VIH-1 .....	40
Figure 12. Cycle réplcatif du VIH-1 .....	44
Figure 13. Représentation de l'entrée du VIH-1 dans les cellules cibles .....	45
Figure 14. Transcription inverse du génome du VIH-1 .....	47
Figure 15. Evolution au cours du d'une infection par le VIH-1 .....	51
Figure 16. Représentation des facteurs de restrictions cellulaires et des protéines virales qui contrecarrent les défenses cellulaires au sein du cycle viral du VIH-1 .....	54
Figure 17. Représentation des étapes du cycle viral du VIH-1 ciblées par les ART disponibles sur le marché .....	59
Figure 18. Représentation des interactions tertiaires de la région 5'UTR de l'ARNg .....	61
Figure 19. Modèle des interactions tertiaires de l'extrémité 5' UTR de l'ARNg impliquées dans la régulation de la liaison spécifique avec Gag .....	63
Figure 20. Représentation du domaine C-terminal de Gag, de sa séquence et de ses partenaires .....	66
Figure 21. Modèle de l'étape de bourgeonnement et de relargage de la particule virale du VIH-1 .....	67
Figure 22. Modèle proposé pour le rôle du domaine p6 dans le recrutement spécifique de l'ARNg .....	71
Figure 23. Représentation des étapes clés de l'encapsidation de l'ARNg .....	72
Figure 24. Représentation du plasmide MAXIPuc18 utilisé pour produire les fragments d'ARNg .....	75
Figure 25. Représentation des étapes de production des virus recombinants .....	84
Figure 26. Représentation des grandes étapes de production des protéines dans les cellules BHK21 .....	85
Figure 27. Représentation des étapes de production et de purification des protéines .....	87
Figure 28. Représentation de l'analyse des macromolécules par MS/MS .....	92
Figure 29. Représentation d'un instrument de diffusion dynamique de la lumière (DLS ; Wyatt) .....	95
Figure 30. Représentation du système de chromatographie d'exclusion de taille couplé à de la diffusion de lumière multi-angle (SEC-MALS ; Wyatt) .....	97
Figure 31. Représentation du principe de la méthode d'analyse de la dispersion induite par un flux (FIDABio) .....	99
Figure 32. Représentation du système d'ultracentrifugation analytique (AUC) .....	101
Figure 33. Représentation du principe de la nano-fluorimétrie à balayage différentiel (NanoDSF ; Prometheus) .....	102
Figure 34. Représentation de la méthode d'anisotropie de fluorescence .....	103
Figure 35. Purification de 6His-Gag dans le tampon A1 .....	110
Figure 36. Purification de 6His-Gag dans le tampon A3 .....	112
Figure 37. Analyse par Western blot (WB) des protéines 6His-Gag purifiées dans les tampons A1 et A3 .....	113
Figure 38. Couverture de la séquence de 6His-Gag par des spectres MS/MS pour les préparations en solution et en gel .....	114
Figure 39. Couverture de séquences des bandes protéines (bande haute dans le gel, encadré en vert) et contaminants (bande basse dans le gel, encadré en orange) pour chaque résidu .....	114
Figure 40. Analyse DLS des protéines 6His-Gag produites en cellules bactériennes .....	116

Figure 41. Analyse DLS des protéines 6His-Gag produites en BHK21 et purifiées dans le tampon A1.....	117
Figure 42. Analyse DLS des protéines 6His-Gag produites en BHK21 et purifiées dans le tampon A3.....	118
Figure 43. Analyse DLS des protéines 6His-Gag purifiées dans le tampon A1 et A3 en comparaison avec le profil d'analyse DLS des protéines 6His-Gag purifiées dans le tampon A1 et dialysées contre le tampon A3.....	119
Figure 44. Analyse SEC-MALS de 6His-Gag purifiée dans le tampon A1 après traitement avec le modèle classique ....	120
Figure 45. Analyse SEC-MALS de 6His-Gag purifiée dans le tampon A1 après traitement avec le model Protein Conjugate Analysis.....	120
Figure 46. Analyse SEC-MALS de 6His-Gag purifiée dans le tampon A1 et dialysée contre le tampon A3.....	121
Figure 47. Analyse des vitesses de sédimentation de la protéine recombinante 6His-Gag dans les différentes solutions tamponnées.....	122
Figure 48. Détermination des températures de dénaturation (TM) de la protéine recombinante 6His-Gag les solutions tamponnées A1 et A3 .....	122
Figure 49. Analyse par gel agarose Tris Borate magnésium (TBM) 1% de la dimérisation des ARNg marqués et non marqués.....	125
Figure 50. Représentation des analyses de vitesse de sédimentation de l'ARN111WT dans le tampon A1.....	126
Figure 51. Représentation des analyses de vitesse de sédimentation SV de l'ARN111 WT dans le tampon A1 .....	127
Figure 52. Comparaison par DLS de l'effet de la dilution de la concentration de NaCl pour la protéine 6His-Gag en solution dans le tampon A3.....	128
Figure 53. Comparaison par DLS de l'effet de la concentration de NaCl sur la formation de complexes entre 6His-Gag et l'ARN111WT .....	129
Figure 54. Comparaison par DLS de l'effet de l'ajout d'ARNt ou d'ARN111Wt à la protéine 6His-Gag dans le tampon A3 .....	130
Figure 55. Représentation des analyses de vitesse de sédimentation de 6His-Gag mélange à l'ARN111WT .....	132
Figure 56. Représentation des Rh moyens de la protéine 6His-Gag marquée par une sonde ALEXA488. en présence d'une quantité croissante de dimères d'ARN 111WT obtenus par FIDABio.....	134
Figure 57. Analyse par gel dénaturants supplémentés en urée révélés au Typhoon et par un bain de BEt du marquage des ARNg.....	135
Figure 58. Représentation des moyennes des valeurs d'anisotropie de l'ARN111WT marqué et dimérique en présence de quantités croissantes de 6His-Gag et ajustement de la courbe par le modèle de Scatchard .....	136
Figure 59. Représentation des moyennes des valeurs d'anisotropie de l'ARNSL135 marqué et dimérique en présence de quantités croissantes de 6His-Gag et ajustement de la courbe par le modèle de Scatchard .....	137
Figure 60. Représentation des moyennes des valeurs d'anisotropie de l'ARN111AAA marqué et replié en présence de quantités croissantes de 6His-Gag.....	138
Figure 61. Analyse par diffraction aux rayons X des cristaux formés en présence de 6His-Gag dans le tampon A3 ....	139
Figure 62. Micrographes des grilles de microscopie électronique à coloration négative chargée avec différentes concentration 6His-Gag dans le tampon A3.....	140
Figure 63. Micrographe d'une grille C-flat chargée avec 0,18 mg/mL de 6His-Gag purifiée dans le tampon A3 supplémenté avec du DDM et du glutaraldéhyde .....	141
Figure 64. Criblages des conditions de préparation des grilles de carbones pour la CryoEM par le Chameleon.....	142
Figure 65. Reproduction des différentes étapes de construction du modèle de structure de la protéine 6His-Gag (1/3) .....	144
Figure 66. Reproduction des différentes étapes de construction du modèle de structure de la protéine 6His-Gag (2/3) .....	144
Figure 67. Reproduction des différentes étapes de construction du modèle de structure de la protéine 6His-Gag (3/3) .....	145
Figure 68. Analyse par DLS de l'effet du DDM couplé ou non avec du glutaraldéhyde, sur la formation de complexes entre 6His-Gag et l'ARN dans le tampon A3 .....	146

Figure 69. Micrographe d'une grille C-flat chargée avec 0,20 mg/mL de 6His-Gag en complexe avec l'ARNg (1 : 10) supplémenté avec du DDM et du glutaraldéhyde .....	147
Figure 70. Visualisation du modèle proposé par Alphafold pour la structure de la protéine Gag ainsi que les scores associés à la prédiction de la position de chaque résidu.....	148
Figure 71. Visualisation du modèle proposé par Alphafold pour la structure de la protéine 6His-Gag ainsi que les scores associés à la prédiction de la position de chaque résidu.....	149
Figure 72. Visualisation du modèle proposé par Alphafold pour la structure de la protéine Gag dépourvue du domaine p6 ainsi que les scores associés à la prédiction de la position de chaque résidu .....	150
Figure 73. Représentation des PTM de la protéine Gag en fonction de ces domaines .....	154
Figure 74. Analyse par SDS-PAGE de la purification de Vpr et détermination de sa séquence par WB dans les tampons A1 et A3 .....	169
Figure 75. Analyse par SDS-PAGE de la purification d'UEV et détermination de sa séquence par WB dans les tampons A1 et A3 .....	170
Figure 76. Analyse par SDS-PAGE de la purification d'ALIX et détermination de sa séquence par WB dans le tampon ALIX .....	171
Figure 77. Purification de R-ALIX dans le tampon ALIX.....	173
Figure 78. Couverture de la séquence de la protéine R-ALIX pour les préparations en solution et en gel .....	173
Figure 79. Comparaison des profils DLS des fractions nommées pic1 et pic2 obtenues à l'issue de la purification par chromatographie de taille d'ALIX .....	175
Figure 80. Purification de R-Vpr dans le tampon A2 .....	176
Figure 81. Couverture de la séquence de la protéine R-Vpr pour les préparations en gel .....	176
Figure 82. Couverture de la séquence de la protéine R-Vpr et mise en évidence des sites de coupures des enzymes de digestion disponibles à la plateforme .....	177
Figure 83. Profils DLS de la protéine Vpr produite en cellules de mammifères .....	178
Figure 84. Profils DLS de la protéine Vpr produite par synthèse chimique .....	178
Figure 85. Analyse par FIDAbio de l'effet de l'ajout de quantité croissante de protéine Vpr synthétisée chimiquement sur le Rh moyens de la protéine 6His-Gag marquée par une sonde fluorescente .....	182
Figure 86. Analyse par FIDAbio de l'effet de l'ajout de quantité croissante de protéine R-Vpr sur le Rh moyens de la protéine Gag-6His, produite en cellules bactériennes, marquée par une sonde fluorescente.....	183
Figure 87. Modèle de la stratégie d'expression de multi protéines en BHK21 par un MVA recombinant .....	192
Figure 88. Stratégie expérimentales proposée pour apporter une meilleure compréhension des interactions entre Gag et l'ARNg permettant l'encapsidation spécifique de l'ARNg et du rôle des partenaires de p6 dans l'encapsidation spécifique de l'ARNg.....	197



## Liste des annexes

**Annexe 1** : Bussienne, C., Marquet, R., Paillart, J., C., and Bernacchi, S. (2021). Post-translational modifications of retroviral HIV-1 Gag precursors: an overview of their biological role. *Int. J. Mol. Sci.* 2021, 22, 2871

**Annexe 2** : Profils AUC obtenus après application du modèle distribution continue avec une variable discrète

**Annexe 3** : Profils FIDAbio obtenu pour déterminer la viscosité de la solution tamponnée A1 et A3



# Introduction



## I. La grande famille des Rétrovirus (Retroviridae)

### I.1 Les principales découvertes autour des rétrovirus

Les rétrovirus sont découverts au début du 20<sup>ième</sup> siècle dans le cadre d'études consacrées aux maladies cancéreuses des poulets. Dans ce but, l'équipe du docteur Ellermann montre que la leucose du poulet est causée par un virus, que l'on appelle aujourd'hui le virus de la leucose aviaire (ALV ; Ellermann and Bang, 1909). Quelques années plus tard, l'équipe du docteur Rous isole le virus responsable d'un sarcome chez les poulets (Rous, 1911). Le virus porte aujourd'hui son nom, le Rous Sarcoma Virus (RSV). C'est en 1936 que la découverte des tumeurs induites par des virus est étendue aux espèces mammifères, commençant par les souris. Mais rapidement, de nombreux virus de ce type sont découverts chez d'autres espèces comme les chats, les bovins et les singes (Coffin et al., 1997). Dans les années 1957, l'infection de moutons par un rétrovirus produisant une maladie neurologique (appelée Visna) met en évidence l'existence d'infections virales lentes. Le nom de lentivirus provenant du mot latin *lentus* est associé à ces virus. Quelques années plus tard, la présence d'ARN dans les particules virales de RSV est démontrée (Crawford and Crawford, 1961). Cependant, les cellules infectées par le RSV présentent des phénotypes qui sont conservés au cours des mitoses. L'hypothèse d'une copie ADN capable de s'intégrer dans le chromosome de l'hôte est alors émise. Le nom de provirus par analogie avec le prophage intégré des bactériophages, est proposé en 1964. Il faudra attendre les années 1970 et la découverte de la transcriptase inverse pour que la présence d'un d'ADN rétroviral intégré dans les cellules infectées, soit acceptée (Temin and Mizutani, 1970). L'isolement et l'identification du premier rétrovirus humain ont lieu en 1980 grâce à la mise au point de culture de cellules lymphocytes T *in vitro*. Ce virus est le virus de la leucémie des cellules T humaines de type 1 (HTLV-1 ; Poiesz et al., 1980). En comparaison avec les autres virus déjà découverts, le génome du HTLV-1 est défini comme complexe. Le clonage et le séquençage de son génome viral montre qu'il code pour des protéines rétrovirales additionnelles d'abord appelées protéines X, aujourd'hui renommées Tax et Rex (Coffin et al., 1997). En 1982, le HTLV-2 est à son tour découvert (Yoshida et al., 1982). Toutes ces découvertes ont rendu nécessaire la création de différentes classifications en fonction des caractéristiques des virus.

### I.2 Classification phylogénique des rétrovirus

De nos jours, un virus appartenant à la famille des Retroviridae est caractérisé comme étant un virus enveloppé qui possède un génome à ARN de polarité positive et qui code pour la transcriptase inverse (RT) et l'intégrase (IN). Ses protéines permettent respectivement la rétrotranscription du génome ARN en ADN double brin et son intégration dans le génome cellulaire. Cette famille contient deux sous-familles les Orthoretrovirinae et les Spumavirinae. Il existe au total 7 genres de rétrovirus qui peuvent être classés en fonction de la complexité de

leurs génomes ou de leurs caractéristiques virales. Les espèces ALV, RSV font partis du genre **Alpharetrovirus**, les virus Mason-Pfizer Monkey Virus (MPMV) et le Mouse Mammary Tumor Virus (MMTV) font partie du genre **Betaretrovirus**, les espèces Murine Leukemia Virus (MLV), Feline Leukemia Virus (FeLV) et Moloney Murine Sarcoma virus (MoMS) appartiennent au genre **Gammaretrovirus**. Les **Epsiloretrovirus** sont représentés par le Walleye Dermal Sarcoma virus. Tous ces genres sont des rétrovirus simples. Leur génome ne code que pour les protéines de structure : groupe d'antigènes spécifique (*group-specific antigen*, Gag) enzymatique : Pol et d'enveloppe : Env (voir section Introduction ; §III.2). Les rétrovirus dits complexes possèdent un génome qui code pour ces protéines, mais aussi pour des protéines supplémentaires appelées auxiliaires. Les **Deltaretrovirus** dont font partie le Bovine Leukemia Virus (BLV), le HTLV, le primate T-lymphotropic virus et les **Lentivirus** dont le VIH, les virus de l'immunodéficience du singe (SIV) et le **Spumaretrovirinae** sont des rétrovirus complexes.

Les **Alpharetrovirus**, **Betaretrovirus**, **Gammaretrovirus**, **Deltaretrovirus** et **Epsiloretrovirus** sont des rétrovirus oncogènes ; les **Lentivirus** sont des virus caractérisés par la lenteur de leur cycle infectieux et enfin les **Spumavirus** dont le nom vient du latin *spuma* à cause de leur aspect mousseux dans le cytoplasme des cellules infectées (d'après ICTV, <https://ictv.global/taxonomy>).

## II. Un rétrovirus particulier : le virus de l'immunodéficience humaine (VIH)

Parmi tous les rétrovirus, le VIH est un virus particulier. Responsable d'une pandémie mondiale depuis les années 1981, il n'a cessé de causer la mort de nombreux patients. En effet le nombre de personnes infectées depuis l'apparition du virus est estimé à 84 millions et le nombre de décès associé une infection par le VIH est estimé à 40 millions (en 2021, selon l'ONUSIDA, <https://www.unaids.org/fr/>). Ainsi, ce virus représente encore actuellement un problème de santé majeur dont l'étude est importante encore de nos jours.

### II.1 La découverte d'un nouveau virus : le VIH

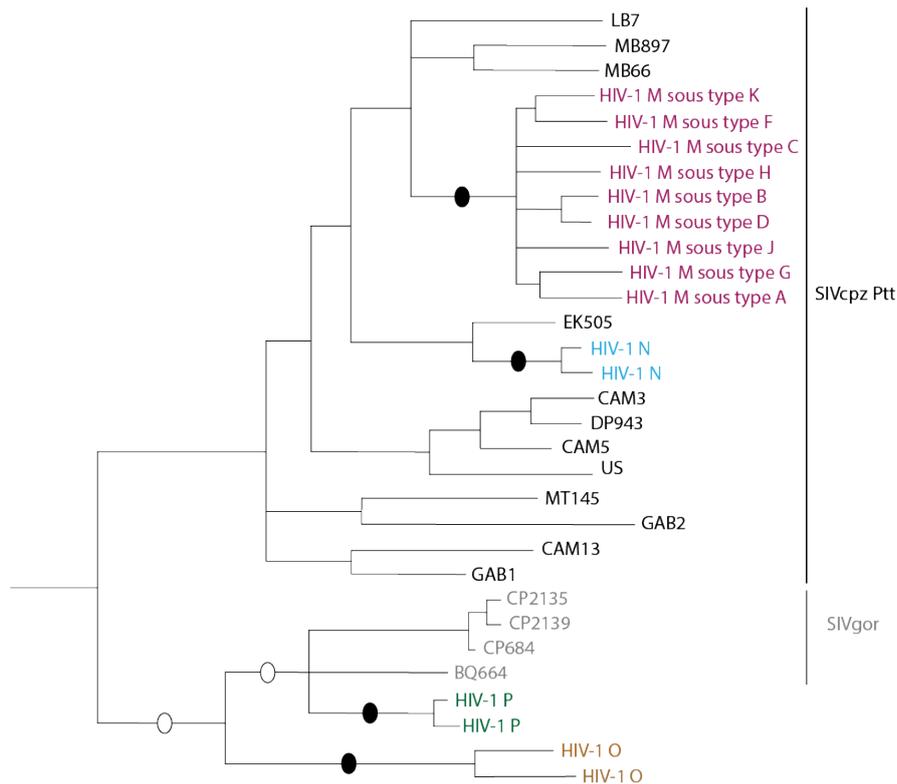
En 1981, aux Etats-Unis, de jeunes patients sont atteints par des maladies telles que sarcome de Kaposi et des pneumonies rares. De plus, ils présentent une incapacité à se défendre contre des maladies infectieuses opportunistes normalement bénignes (Hymes et al., 1981). Ces maladies ont la particularité d'infecter des personnes immunodéprimées. Des contrôles montrent que les patients atteints de ces maladies présentent tous et sans exception un taux de lymphocytes T extrêmement bas. La maladie associée à cet affaiblissement du système immunitaire est alors nommée syndrome de l'immunodéficience acquise (SIDA). En 1983, l'isolement du virus à partir des ganglions lymphatiques des patients a permis à l'équipe du docteur Barré-Sinoussi de montrer que la protéine de capsid est différente de celle du HTLV-1. Ainsi le nouveau virus est nommé lympho-adenopathie virus (LAV ; Barré-Sinoussi et al., 1983). Cependant, après plusieurs années de désaccords au sein de la

communauté scientifique, le virus sera rebaptisé Virus de l'Immunodéficience Humaine (VIH) en 1986. Cette année correspond aussi à la découverte d'un nouveau virus présentant 42% d'homologie avec le VIH de type 1 (VIH-1), c'est le VIH de type 2 (VIH-2 ; Clavel et al., 1986). Les chercheurs ayant isolé et identifié le VIH-1 sont récompensés du Prix Nobel de médecine en 2008.

## II.2 Les Origines du VIH

### II.2.1 Franchissements d'espèces et variabilité génétique du VIH

Le VIH-1 est sous-divisé en 4 groupes et le VIH-2 est sous-divisé en 8 groupes. Les deux espèces de virus, ainsi que chacun de leurs groupes, proviennent d'évènements de franchissement d'espèces indépendants. Ces transmissions entre les espèces ont eu lieu dans la même zone géographique, au Congo et ont été possibles à cause des activités de chasse ou de capture des singes (Cavallari et al., 2011; Peeters et al., 1997). Ainsi, il existe quatre groupes au sein du VIH-1, le groupe majoritaire (M), le groupe *outlier* (O), le groupe non-M, non-O (N) et le groupe P. Le groupe M provient d'un franchissement d'espèces entre les chimpanzés de la sous espèce de *Pan troglodytes* infectée avec un SIVcpzPtt (**figure 1**). Ce groupe est responsable de la pandémie de SIDA, infectant environ 39 millions de personnes dans le monde (2022, ONUSIDA). Au cours de l'histoire, le SIVcpzPtt est transmis à l'homme une seconde fois lors d'un nouvel événement de franchissement d'espèce permettant l'apparition du groupe N (Keele et al., 2006). Le succès épidémiologique est largement différent d'un groupe à l'autre, en effet le groupe N découvert en 1998 (Simon et al., 1998) a infecté 20 personnes au Cameroun (Ayoub et al., 2000; Roques et al., 2004). Il est intéressant de noter que le franchissement d'espèce n'a pas eu lieu qu'entre le singe et l'Homme mais aussi entre différentes espèces de singes. Ainsi, le SIVcpzPtt est aussi à l'origine des SIV infectants les espèces de Gorilles *gorilla* (D'arc et al., 2015). Deux nouveaux évènements de franchissement d'espèce indépendants, du gorille vers l'Homme ont permis l'apparition des VIH de groupe O et P (Keele et al., 2006; Plantier et al., 2009). Le groupe O est responsable de l'infection d'environ 100.000 personnes, (Mourez et al., 2013; Peeters et al., 1997). Enfin, le groupe P n'est reporté que pour deux cas, en 2009, au Cameroun (Vallari et al., 2011). Les raisons de la différence de succès épidémiologique ne sont pas entièrement définies à ce jour.



**Figure 1. Arbre phylogénétique des origines du VIH-1**

Représentation de la relation phylogénétique pour une région du gène Pol parmi les SIVcpzPtt représentés en noir, SIVgor représenté en gris et les quatre groupes du VIH-1. Le groupe M est représenté en violet le groupe N en bleu, le groupe P en vert et le groupe O en marron. Les cercles blancs représentent les possibles transmissions entre chimpanzés et gorilles et les cercles noirs représentent les possibles transmissions du chimpanzé ou du gorille à l'homme (modifiée depuis Sharp and Hahn, 2011).

### II.2.2 Origine et variabilité génétique du VIH de type 2

Le VIH-2 à une autre origine phylogénétique. En effet il provient d'un franchissement d'espèce à partir du SIVmmb infectant le mangabey fuligineux (Gao et al., 1999). L'épidémie du VIH-2 est limitée à l'Afrique de l'Ouest (Tchounga et al., 2016). Comme pour le VIH-1 plusieurs franchissements d'espèces entre le SIVsmm et l'Homme sont responsables de l'apparition de 8 groupes de A à H (Sharp and Hahn, 2011). Le VIH-2 est beaucoup moins infectieux que le VIH-1, seuls les groupes A et D se sont propagés et circulent toujours de nos jours (Visseaux et al., 2016). De plus, leurs histoires naturelles diffèrent au niveau de la progression de la maladie, en effet la plupart des personnes infectées par le VIH-2 ne développent pas le SIDA (Rowland-Jones and Whittle, 2007).

Le VIH-1 et le VIH-2 ne présentent que 42% d'homologie de séquence (Dilley et al., 2011). Ainsi, la suite de mon travail de thèse portera sur le VIH-1.

### II.2.3 Diversité génétique au sein des groupes du VIH-1

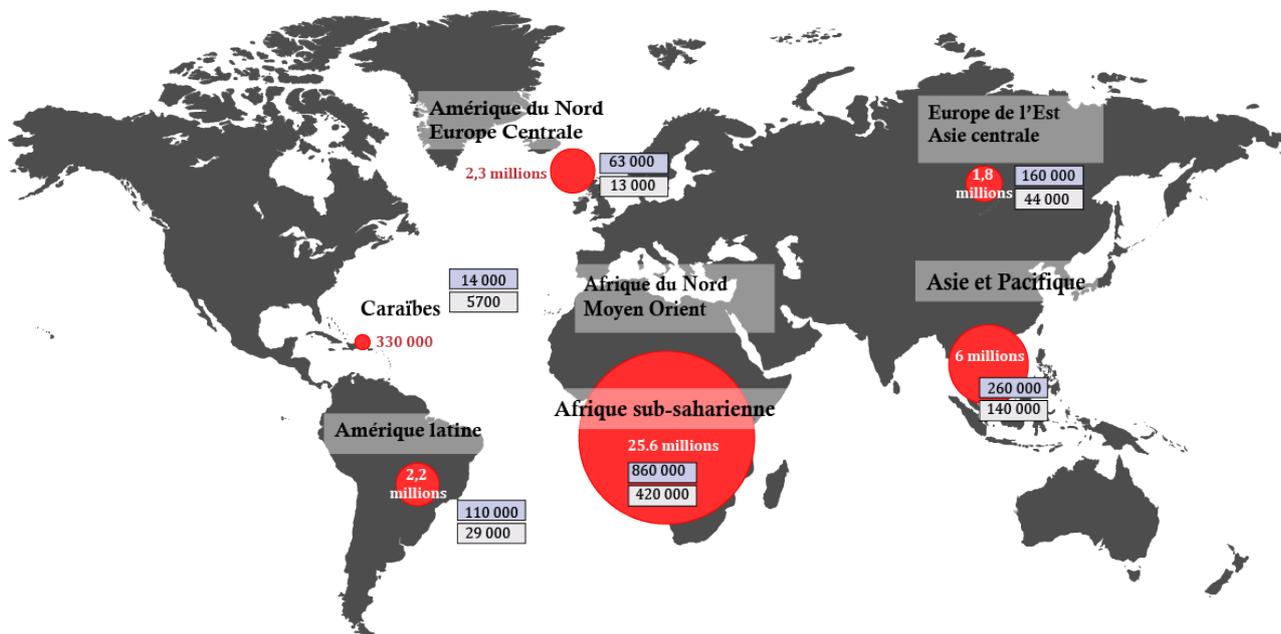
Le VIH a une grande capacité évolutive et de variabilité génétique. Cette particularité est directement liée à son cycle viral et notamment aux erreurs produites lors de la transcription inverse (environ de  $10^{-5}$  à  $10^{-3}$  erreurs par

paire de base par cycle (Yeo et al., 2020). Ces erreurs sont dues à l'absence de mécanisme de relecture et à la recombinaison pendant la transcription inverse (détaillée section introduction ; §III.4.1). Ces erreurs augmentent la diversité génétique du VIH-1 au sein des patients (Korber et al., 2001). Cette variabilité a permis l'apparition de variants présentant jusqu'à 35% de différence au sein du groupe M (Hemelaar et al., 2006). Les virus de ce groupe sont donc subdivisés en 10 sous-types de A à L (Bbosa et al., 2019). Plus rarement, la recombinaison pendant la transcription inverse peut impliquer des virus de sous-types différents qui auraient coïnfecté la même cellule (Peeters et al., 1999; Rousseau et al., 2007). Ce phénomène s'est produit suffisamment pour qu'une nouvelle classification des formes recombinantes soit créée. Si les recombinants sont capables de circuler d'un individu à un autre on parle de formes recombinantes circulantes (CRF). Au contraire, si les recombinants ne sont pas capables de circuler entre les individus, on parle alors de formes recombinantes uniques (URF ; Robertson et al., 2000). Actuellement, 98 CRF sont caractérisées et ce est en augmentation (Bbosa et al., 2019). La formation de recombinaison entre sous types différents semble préférentiellement se produire lorsque les ARN présentent un fort degré d'identité (Baird et al., 2006). Néanmoins, des recombinaisons entre isolats de groupes M et O ont déjà été caractérisées (De Oliveira et al., 2018 ; De Oliveira et al., 2018). Les erreurs et les recombinaisons lors de l'étape de transcription inverse présentent un problème de santé majeur, en effet cette variabilité favorise les mécanismes d'échappement du virus face au système immunitaire, mais aussi face aux traitements.

### **II.3 Les chiffres clés dans le monde : données en 2021**

En 2021, 38,4 millions de personnes vivaient avec le VIH, 650 000 personnes décédaient du virus. Les nouvelles infections touchaient 1,5 million de personnes. Malheureusement, 9,7 millions de personnes n'ont pas eu accès aux traitements antirétroviraux (*AntiRetroviral Treatment, ART* ; ONUSIDA, <https://www.unaids.org/fr/resources/fact-sheet>). Il est important de noter que l'accès aux ART a un fort impact sur la mortalité. En 2021, les populations clés fortement infectées sont les travailleuses du sexe et leurs clients, les hommes ayant des rapports sexuels avec d'autres hommes (HSH), les consommateurs de drogues injectables et les femmes transgenres. Ces populations représentent 70% de toutes les nouvelles infections au VIH. En d'autres chiffres, le risque de contracter le VIH est évalué à 35 fois plus élevé chez les personnes utilisant des drogues injectables que chez des adultes qui ne s'injectent pas de drogue ; 30 fois plus élevé pour les femmes professionnelles du sexe que chez les autres femmes ; 28 fois plus élevé chez les hommes ayant des rapports sexuels avec d'autres hommes que chez les autres hommes et 14 fois plus élevé pour les femmes transgenres. Néanmoins, il est important de noter que les tendances varient entre l'Afrique subsaharienne, où l'épidémie sévit le plus fort et les autres pays (**figure 2**). En Afrique, les femmes âgées de 15 à 24 ans sont deux fois plus

susceptibles de vivre avec le VIH que les hommes du même âge. En 2021, elles représentent 63% de toutes les nouvelles infections (ONUSIDA, 2022, <https://www.unaids.org/fr/resources/>).



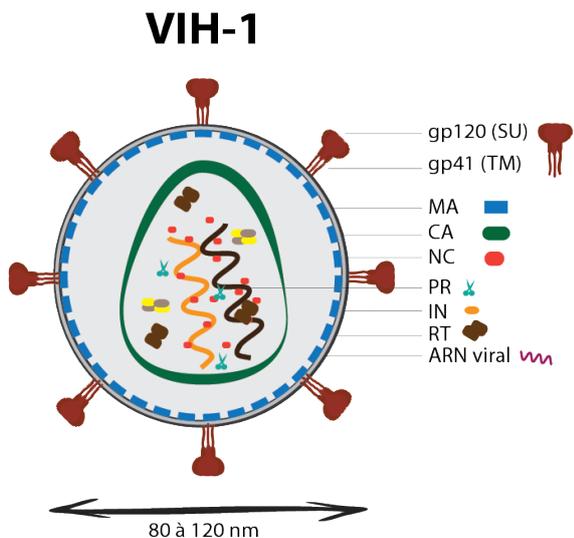
**Figure 2. Distribution de la prévalence du VIH-1 dans le monde**

Encadré en bleu le nombre de nouvelles infections et en blanc le nombre de décès associés au VIH-1 en 2020 (ONUSIDA, 2022, <https://www.unaids.org/fr/resources/>).

### III. Cycle viral du VIH-1

#### III.1 La particule virale, structure et composition du génome

Le VIH-1 est un virus sphérique de 80 à 120 nm de diamètre enveloppé d'une bicouche lipidique externe acquise lors de son bourgeonnement (**figure 3**).



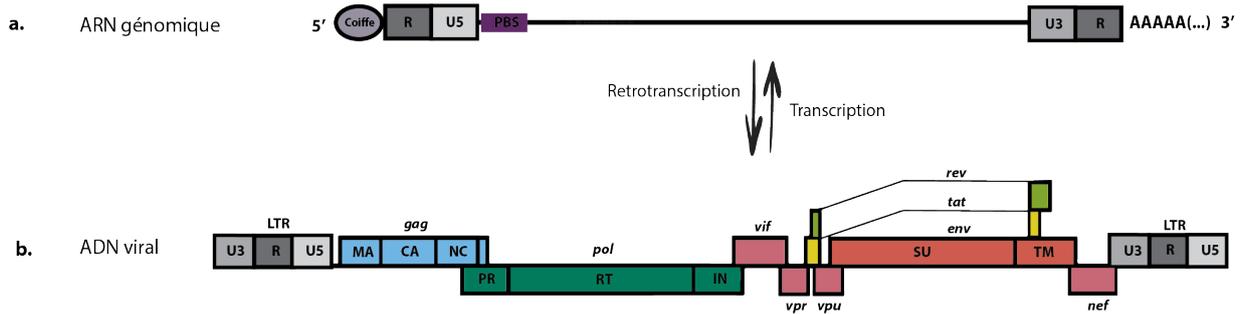
**Figure 3. Représentation de la particule virale du VIH-1**

La particule du VIH-1 est enveloppée d'une bicouche lipidique provenant du bourgeonnement du virus lors d'une infection antérieure et est recouverte de glycoprotéines d'enveloppes virales (gp120 et gp41). La face interne de cette bicouche lipide est tapissée de protéines de matrice (MA) qui entoure le core conique formé de protéines de capsid (CA). A l'intérieur se trouvent deux copies d'ARNg dimériques recouverts de nucléocapside (NC) et les protéines enzymatiques de transcriptase inverse (RT), protéase (PR) et d'intégrase (IN). La figure est adaptée de Sundquist and Kräusslich, 2012.

Cette bicouche est recouverte de glycoprotéines de surface cellulaires et virales. Les protéines d'enveloppes virales sont nécessaires à la reconnaissance du récepteur cellulaire et donc à l'internalisation du virus. A l'intérieur du virion, se trouve une structure conique en fullerène composée de protéines de capsid (CA) enfermant deux copies d'ARN génomique (ARNg) simple brin de polarité positive. L'ARNg est protégé par de multiples copies de nucléocapsides. D'autres macromolécules cellulaires et protéines virales sont aussi retrouvées à l'intérieur de la particule virale (Sundquist and Kräusslich, 2012).

L'ARNg a un poids moléculaire d'environ 9,3 kb. De manière semblable aux ARN messagers cellulaires, l'ARNg viral est coiffé et polyadénylé. Il est flanqué de régions non codantes ou régions non traduites (*untranslated region*, UTR) au niveau des extrémités 5' et 3'. Ces régions sont composées de séquences régulatrices intervenant dans différentes étapes du cycle viral. L'extrémité 5' est importante pour la dimérisation et l'encapsidation de l'ARNg (détaillée section Introduction ; §V1 ; Smyth et al., 2015). Cette région hautement structurée est composée d'une région répétée (R) et d'une séquence unique U5. L'extrémité 3' est importante pour la régulation de la transcription (**figure 4a**). Elle contient la région R et une région unique U3 (Das et al., 1999). Les régions R des extrémités 5' et 3' contiennent la séquence d'activation de la transcription (*Trans-Activation Response element* ; TAR) nécessaire à la transactivation du promoteur viral et à la réplication du virus ainsi que le signal de polyadénylation (polyA). La conformation de l'extrémité 5' de l'ARN, ne permet pas la fonctionnalité de la région R alors que à l'extrémité 3' cette région est fonctionnelle (Das et al., 1999). De plus, en aval de la région U5 de l'extrémité 5' se trouve le site de fixation de l'amorce de la transcription inverse (*Primer Binding Site*, PBS) permettant la fixation de l'ARNt lysine cellulaire nécessaire à l'initiation de la transcription inverse (Kleiman, 2002; Rhim et al., 1991) et le site d'encapsidation (*Packaging Signal*, psi ;  $\Psi$ ). Décrit plus en détails section Introduction ; §V.1.

Après la transcription inverse et la formation de l'ADN viral double brin, les régions 5' UTR et 3' UTR sont remplacées par deux longues répétitions terminales (*Long Terminal Repeat*, LTR) toutes deux composées de région uniques U3 (~450 nucléotides) et U5 (~80 nucléotides) séparées par R (~100 nucléotides ; **figure 4b** ; Resnick et al., 1984; Zhang and Crumpacker, 2022).



**Figure 4. Représentation du génome du VIH-1 et de l'ADN proviral**

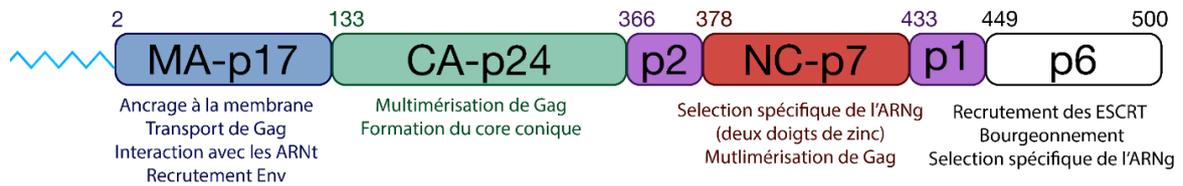
a. Représentation de l'ARNg et notamment des extrémités 5' et 3' UTR. L'ARNg viral est coiffé à l'extrémité 5' et polyadénylé à l'extrémité 3'. b. Représentation de l'ADN viral rétro-transcrit et des gènes qui le composent. Ainsi, les différentes phases codantes sont représentées encadrées par les longues régions terminales répétées (LTR) composées de U3, R et U5. Les trois gènes caractéristiques des rétrovirus gag, pol et env sont respectivement représentés en bleu, en vert et en orange. Chaque gène codant pour une polyprotéine est subdivisé pour représenter les différents domaines. Les gènes codants pour des protéines auxiliaires sont représentés en rose pour les gènes vif, vpr, vpu et nef, jaune pour tat et vert pour rev (modifiée à partir de Meissner et al., 2022).

Ainsi, la région U3 contient des domaines modulateurs, activateurs essentiels à la transcription du provirus intégré (Zhang and Crumpacker, 2022). La région 3'LTR joue un rôle dans le contrôle de l'expression (Malim et al., 1989). Entre les deux LTR se trouvent les régions codantes pour les protéines virales, caractéristiques des rétrovirus (gag, pol et env), organisées en trois cadres de lectures ouverts et des gènes individuels codant pour des protéines auxiliaires, spécifiques des rétrovirus complexes. L'ARNg contient différents sites donneurs et accepteurs d'épissage qui permettent l'obtention d'ARN plus ou moins épissés (Ocwieja et al., 2012). Ainsi l'ARNg peut être multi-épissé, permettant de synthétiser les protéines *Transactivator of Transcription* (Tat), *Regulator of expression of virion proteins* (Rev) et *Negative Element Factor* (Nef). Ce sont les premières protéines à être synthétisées lors du cycle viral. L'ARNg peut aussi être partiellement épissés permettant la traduction des protéines *Viral Infectivity Factor* (Vif), protéine virale R (Vpr), protéine virale U (Vpu) et l'enveloppe (Env). Finalement, l'ARN non épissé permet la traduction des polyprotéines Gag et GagPol et peut être également encapsidé dans les particules de type virales (*viral like particles* ; VLP).

### III.2 Les protéines virales essentielles : Gag, Pol et Env

Le VIH-1 est un rétrovirus complexe et son génome possède les trois principaux cadres de lectures ouverts codant pour les protéines structurales et fonctionnelles du virus (Gag, Pol et Env). Le gène gag code pour la polyprotéine de 55 kDa Pr55Gag ou Gag. Elle est composée de toutes les protéines structurales du virus et nécessite une étape de maturation protéolytique par la protéase virale (PR) pour libérer ces protéines et former ainsi un virus mature. Cependant, sous forme de polyprotéine, ce précurseur joue un rôle crucial dans l'encapsidation de l'ARNg et dans le bourgeonnement et le relargage du virus. Ainsi, la protéine Gag du VIH-1, comme toutes les protéines Gag

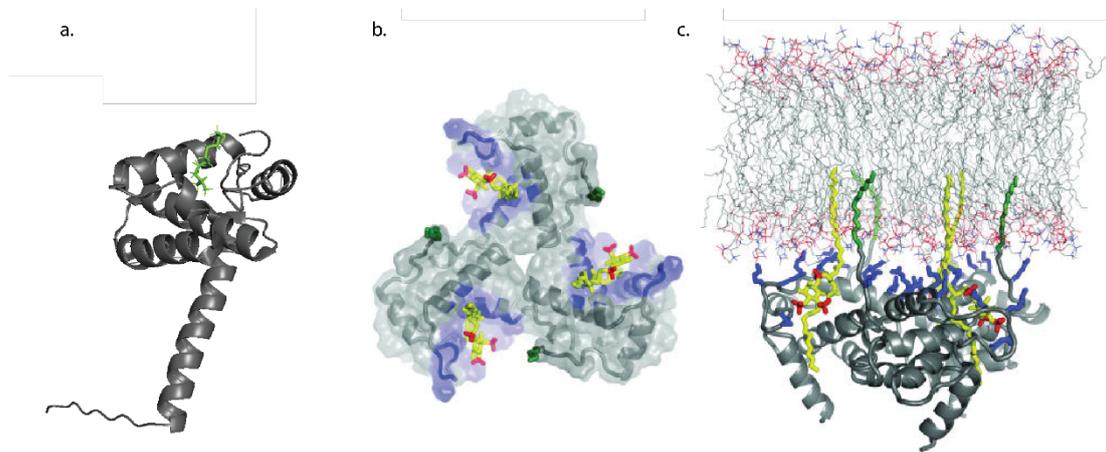
rétrovirales est composée de quatre domaines appelés matrice (MA), capsid (CA), nucléocapsid (NC), p6 et de deux peptides espaceurs (p2 ou sp1 et p1 ou sp2). Chaque domaine de Pr55Gag joue un rôle spécifique dans le cycle réplcatif (**figure 5**).



**Figure 5. Les différents domaines fonctionnels de la protéine Gag du VIH-1 et leurs multiples rôles**

*De Bussienne et al., 2021.*

À partir de l'extrémité N-terminale, le domaine MA de 17 kDa permet d'ancrer la protéine Gag dans la membrane plasmique (MP) au niveau de régions enrichies avec des phospholipides phosphatidylinositol-4,5-bisphosphate (PI(4,5)P<sub>2</sub>). L'absence de PI(4,5)P<sub>2</sub> dans les cellules induit une réduction du nombre de Gag à la MP et inhibe l'étape de relargage viral. De plus, dans ce cas particulier, les précurseurs seraient relocalisés vers les membranes intracellulaires (Chukkapalli et al., 2008). Cet ancrage est possible grâce à un signal bipartite composé de la myristoylation de la glycine en position 2 (voir bussienne et al., 2021 **annexe 1**) et d'une région hautement basique (*Highly Basic Region*, HBR ; Olety and Ono, 2014; Wen et al., 2020). L'acide myristique peut être exposé ou séquestré dans la poche hydrophobe du domaine de MA (**figure 6**).

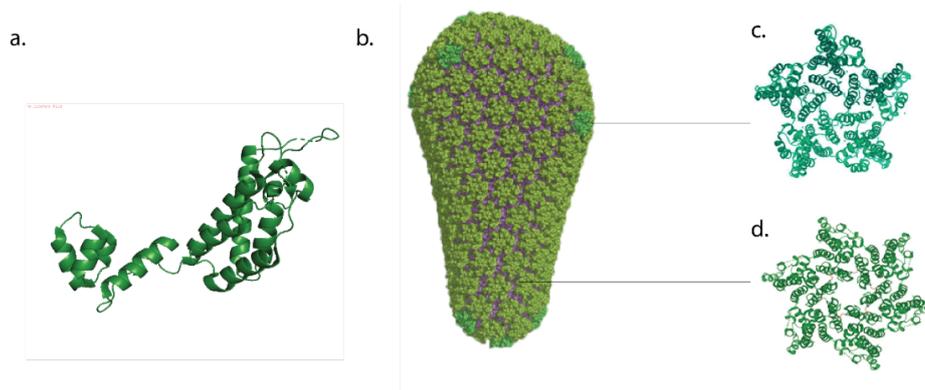


**Figure 6. Les structures tertiaires du domaine MA dans différentes conformations**

*a. Le domaine MA dans sa conformation monomère (en gris) présente un acide myristique séquestré en vert (PDB : 1UPH ; d'après Bussienne et al., 2021). b. Représentation du trimère de MA et des groupes myristoylés exposés représentés en vert. Les phosphatidylinositides de la MP sont représentés en jaune avec les phosphates colorés en rouge. Enfin, les résidus de MA en contact avec la MP sont représentés en bleu (figure issue de Saad et al, 2006). c. Modèle de liaison de MA à la MP. Le trimère de MA est représenté en gris avec les groupements myristoylés exposés colorés en vert. Les résidus en contact avec le phosphatidyl inositide représenté en jaune avec les phosphates colorés en rouge, sont colorés en bleu (figure issue de Saad et al., 2006).*

Ce contrôle de l'exposition de l'acide myristique permet la prévention des interactions aberrantes avec les membranes intracellulaires et assure ainsi l'insertion du groupement dans la MP. L'exposition du myristate est déclenchée par les interactions entre Gag et PI(4,5)P2 (Saad et al., 2006) et des études de résonance magnétique nucléaire (RMN) démontrent que l'exposition au myristate serait également régulée par la trimérisation du domaine MA. En effet la protéine MA dans sa forme mature présente une affinité plus faible pour les membranes par rapport au précurseur complet (Hogue et al., 2012; Saad et al., 2007). De plus, le domaine MA dans le précurseur lie des acides nucléiques tels que les ARNt de l'hôte. Cette interaction facilite le recrutement spécifique de Gag à la MP en inhibant les interactions avec les phosphatidylsérines chargées négativement contenues dans les membranes endocytiques intracellulaires (Chukkapalli et al., 2008). Des études montrent que cette interaction avec les ARN cellulaires ne change pas l'affinité du précurseur pour l'ARNg viral (Kutluay et al., 2014; Olson and Musier-Forsyth, 2019). Enfin, la MA assure l'incorporation de la protéine d'enveloppe, Env, lors de l'assemblage des particules virales en interagissant avec la queue cytoplasmique de la protéine gp41 (Saad et al., 2006; Tedbury et al., 2015).

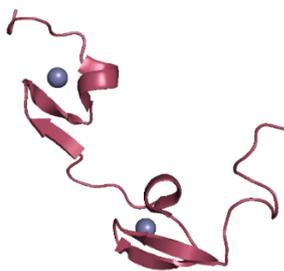
Le domaine de CA est un domaine de 24 kDa qui dirige la multimérisation de Gag (**figure 7**). Une fois mature, la protéine de CA s'auto-assemble dans une structure fullerène composée de 250 hexamères et 12 pentamères formant le core viral (Klingler et al., 2020a). L'interaction avec les protéines cyclophiline A encapsidées dans les particules virales néoformées permet d'échapper au système de défense TRIM5 $\alpha$  lors de l'infection d'une nouvelle cellule cible (Kim et al., 2019; Klingler et al., 2020a ; section Introduction, §IV.2). Dans les étapes précoces du cycle viral, la CA permet le transport du complexe de pré-intégration (PIC) vers le noyau cellulaire en interagissant avec les protéines associées aux microtubules (MAP1A et MAP1S ; Carnes et al., 2018; Klingler et al., 2020a). Ensuite ce domaine permet l'import du PIC au niveau du pore nucléaire en interagissant avec la nucléoporine nommée Nup358/RanBP2 (Dharan et al., 2016; Klingler et al., 2020a). Dans le précurseur Gag le domaine de CA est séparé du domaine de NC par un espaceur de 14 acides aminés appelé Sp1 ou p2. Cet espaceur permet les changements de conformation du précurseur et favorise la multimérisation de Gag (Datta et al., 2016, 2011; Klingler et al., 2020a) et semble également jouer un rôle dans l'empaquetage spécifiques des ARNg viraux par rapport aux ARN cellulaires ou viraux épissés (Kaye and Lever, 1998; Klingler et al., 2020a).



**Figure 7. Les structures tertiaires du domaine CA dans différentes conformations**

**a.** Structure d'un monomère de CA représenté en vert (PDB : 3H47). **b.** Représentation du cône fullerène pour la capsid du VIH en représentant les hexamères en vert clair et les pentamères en vert foncé (adapté depuis Pornillos et al., 2011). **c.** Structure d'un pentamère de CA en vert foncé (PDB : 3P05). **d.** Structure d'un hexamère de CA en vert clair (PDB : 3H47).

Ensuite, la NC (domaine de 7 kDa), est crucial pour l'interaction spécifique avec l'ARNg viral et son empaquetage grâce à la présence de deux motifs en doigt de zinc (CCHC) qui interagissent spécifiquement avec la région Psi de l'ARNg viral (Boutant et al., 2020; Webb et al., 2013 ; **figure 8**). Ce domaine permet également l'incorporation de l'amorce de la transcription inverse, l'ARNt lysine 3 et la multimérisation de Gag, favorisant ainsi l'assemblage viral (Cimarelli et al. 2000; Ott, Corenet Shatzer 2009). Ces résultats montrent bien qu'au-delà de leur fonction primaire les différents domaines de Gag sont impliqués dans plusieurs étapes du cycle réplcatif. Enfin, dans sa forme mature, la NC joue plusieurs rôles de chaperonne d'ARN (Godet and Mély, 2010; Rein et al., 1998) et favorise la processivité de RT et l'activité RNase H (Poljak et al., 2003).



**Figure 8. Représentation de la structure tridimensionnelle de la NC**

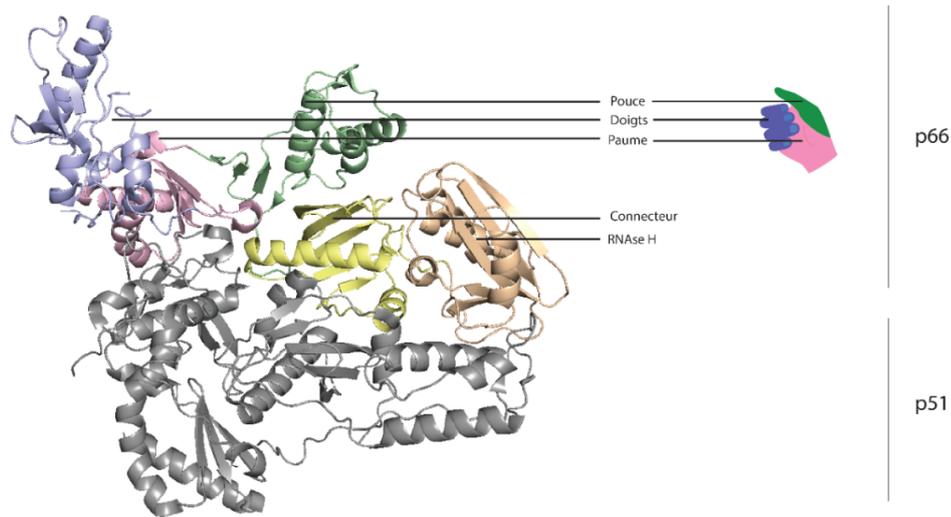
La NC est représentée avec deux molécules de zinc, en gris (PDB : 1MFS).

Dans le précurseur Gag, la NC est séparée du dernier domaine par un deuxième espaceur de 16 acides aminé (sp2 ou p1). Son rôle est peu décrit dans la littérature, bien qu'il pourrait apporter la flexibilité nécessaire aux changements de conformation du précurseur Gag nécessaires aux fonctions du précurseur (Sundquist and Kräusslich, 2012).

Finalement, à l'extrémité C-terminale de Gag se trouve le domaine p6 non structuré de 6 kDa. Ce domaine est largement connu pour réguler le bourgeonnement des virions à la MP en recrutant les éléments de la machinerie ESCRT (*Endosomal Sorting Complex Required for Transport*). De plus, des analyses de RMN ont montré que p6 assure l'incorporation de la protéine virale R (Vpr) dans les particules virales (Salgado et al., 2009a). Récemment, le laboratoire d'accueil a montré que ce domaine joue également un rôle dans l'interaction spécifique entre Gag et l'ARNg (Dubois et al., 2018a).

Le gène *pol* code pour les toutes les enzymes nécessaires au cycle viral, la PR, la RT et l'IN. Ces enzymes sont exprimées sous la forme d'une poly protéine de 160 kDa Pr160 Gag-Pol ou Gag-Pol. Gag-Pol est produit lors d'un évènement de décalage de cadre de lecture lors de la traduction de l'ARN viral. Ce décalage apparaît avec une fréquence de 5% permettant l'expression d'un précurseur Gag-Pol pour vingt précurseurs Gag (Jacks et al., 1988). Le premier domaine qui compose Pol est la PR. C'est une protéine de 15 kDa qui permet de séparer toutes les protéines présentes dans les précurseurs Gag et Gag-Pol. La dimérisation du précurseur Gag-Pol dans les virions immatures permet la dimérisation du domaine de PR, qui va alors être capable de s'auto-exciser. Sa forme active est un homodimère symétrique, tandis que son site actif est composé des résidus acides aspartiques, thréonine et glycine (Asp25, Thr26, Gly27), caractéristique de la famille des aspartyl-protéases. La protéolyse catalysée par la PR se déroule de manière séquentielle et ordonnée (Freed, 2015a; Lee et al., 2012).

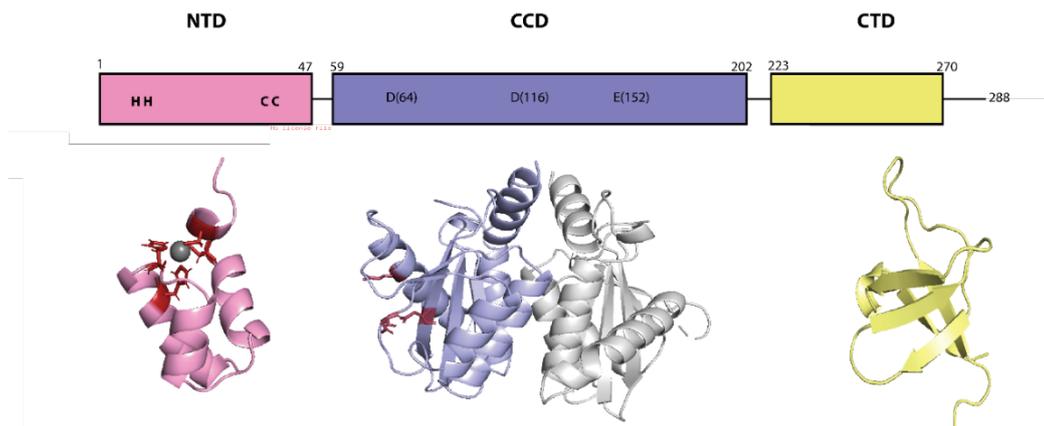
La RT est un domaine de 66 kDa. La forme mature de l'enzyme est composée d'un hétérodimère entre une sous-unité de 51 kDa (p51) et une sous unité de 66 kDa (p66). La sous unité p51 est issue du clivage du domaine RNase H de la p66. Cette sous-unité n'a plus d'activité catalytique et ne possède qu'un rôle structural (Hu and Hughes, 2012; Jacobo-Molina et al., 1993). Seule p66 possède une activité catalytique d'ADN polymérase ARN/ADN dépendante et une activité RNase H qui dégrade spécifiquement l'ARN dans les hybrides ADN/ARN formés lors de la transcription inverse. L'analyse cristallographique de la RT montre que le domaine polymérase de la p66 peut être comparé à une main droite, où se loge le complexe matrice/amorce dans les sous domaines « doigts », « paume » et « pouce » (**figure 9**). La paume représente le site actif de l'enzyme, composé de la triade d'acides aspartiques (Asp111, Asp185 et Asp186) qui catalysent l'attaque nucléophile du nucléotide entrant.



**Figure 9. Représentation de la structure tridimensionnelle de la RT**

La RT est composée de deux sous unités, une de 66 kDa (p66) et une de 51 kDa (p51). La sous unité p51 est représentée en gris. Le domaine RNase H de p66 est représenté en orange et le connecteur en jaune. Les domaines polymérases de p66 sont subdivisés en doigts (en violet), la paume (en rose) et le pouce (en vert) adapté de Hu and Hughes, 2012.

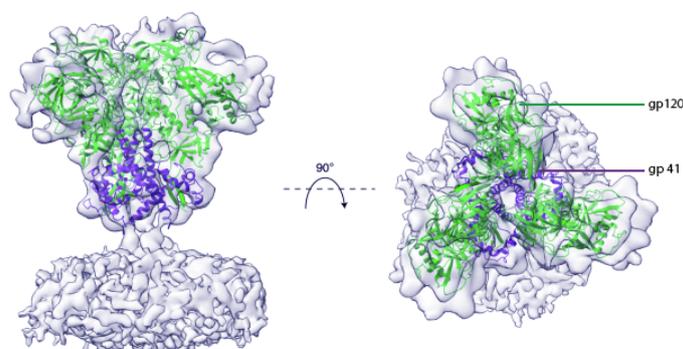
Finalement, l'IN est une protéine de 32 kDa qui permet l'intégration de l'ADN viral dans le génome de la cellule hôte. Elle est constituée de trois domaines, le domaine N-terminal (NTD), le domaine central ou cœur catalytique (CCD) et le domaine C-terminal (CTD ; **figure10**). Le NTD contient un motif en doigt de zinc capable de lier du zinc (HHCC) qui est aussi important pour l'oligomérisation de l'IN. Le CCD est un domaine très conservé chez les rétrovirus. Il comporte la triade catalytique (Asp64, Asp116 - 35 - E152) qui est impliquée dans la liaison spécifique à l'ADN viral et cellulaire. Il contient aussi la surface pour l'interaction avec le cofacteur cellulaire LEDGF/p75 et la région de dimérisation de l'enzyme mature (Serrao et al., 2012). Enfin, le CTD est impliqué dans la fixation non spécifique à l'ADN et permet l'interaction avec la RT nécessaire à l'accomplissement optimal de l'étape de transcription inverse (Lusic and Siliciano, 2017).



**Figure 10. Représentation de la structure des domaines de l'IN**

L'IN est composée de 3 domaines. Le NTD contenant un doigt de zinc est représenté en rose, avec les résidus HHCC, représentés en rouge sur la structure tridimensionnelle, coordonnant un Zn représenté par une boule grise (PDB : 6VRG). Il est relié par un connecteur au CCD représenté sous forme dimérique avec une sous unité en violet et une sous unité en gris (PDB : 4LH4). Les résidus appartenant à la triade catalytique (D64, D111 et E152) sont représentés en rouge sur la structure tridimensionnelle. Finalement, le CTD est représenté en jaune (PDB : 1IHV ; adaptée de Engelman and Cherepanov, 2014).

Le gène *env* code pour les glycoprotéines de l'enveloppe virale. Elles sont exprimées sous la forme d'une polyprotéine appelée gp160 (**figure 11**). Elle est d'abord glycosylée dans le réticulum endoplasmique, puis clivée dans l'appareil de Golgi par la protéase cellulaire à sérine, appelée furine. Cette étape de maturation génère ainsi une protéine de surface (SU), la glycoprotéine de 120 kDa (gp120) et la glycoprotéine transmembranaire (TM) de 41 kDa (gp41 ; Zhu et al., 2006). D'une part, la gp120 est composée de cinq domaines conservés (C1-5) internalisés et de cinq régions variables (V1-5) à la surface (Starcich et al., 1986). Les régions conservées C1 et C5 sont les sites d'interaction avec gp41 alors que les régions de C2-4 sont impliquées dans la liaison avec le récepteur cellulaire *cluster of differentiation* (CD4). D'autre part la gp41 est constituée de trois domaines, le domaine extracellulaire N-terminal, le domaine transmembranaire et la partie cytoplasmique C-terminale. C'est le domaine extracellulaire qui contient le peptide de fusion permettant la fusion entre les membranes cellulaires et virales (Barteseighi et al., 2013).



**Figure 11. Représentation de la structure d'un trimère d'Env du VIH-1**

Représentation d'un trimère d'Env obtenu par cryotomographie électronique. Le trimère de protéines est composé des sous-unités gp120, représentées en vert et gp41 représenté en violet (PDB : 4ZMJ ; adaptée Li et al., 2020).

### III.3 Les protéines virales auxiliaires

Le génome du VIH est complexe car il code pour six protéines supplémentaires par rapport aux rétrovirus simples. Il code pour deux protéines régulatrices de la réplication virale et du métabolisme cellulaire, le facteur de régulation de l'épissage (Rev) et la protéine Trans activatrice (Tat). Il code aussi pour quatre autres protéines auxiliaires, le facteur régulateur négatif (Nef), le facteur d'infectivité virale (Vif), la protéine virale R (Vpr) et la protéine virale unique (Vpu) qui ont des rôles dans la pathogénèse, le contrôle du cycle cellulaire mais aussi dans les mécanismes d'échappement aux facteurs de restrictions.

#### III.3.1 Protéines régulatrices Rev, Tat et Nef

Les protéines Rev, Tat et Nef sont traduites à partir de l'ARN multi-épissé.

La protéine Rev est une protéine de 19 kDa. Cette protéine possède un signal de localisation nucléaire (NLS) et un signal d'export nucléaire (NES), qui facilite son transfert nucléo-cytoplasmique. Elle a pour rôle le transport des ARN non-épissés (ARNg) ou partiellement épissés, en se fixant à l'élément de réponse Rev (RRE) et en permettant leur export dans le cytoplasme (Kalland et al., 1994). Les ARN non-épissés ainsi exportés sont soit traduits en précurseurs Gag et Gag-Pol, soit incorporés sous forme de dimère dans les particules virales naissantes.

La protéine Tat est une protéine de 9,8 kDa, composée de deux domaines, le domaine d'activation et le domaine de liaison à l'ARN. Cette protéine active la transcription des ARN messagers viraux en se liant au niveau de l'élément TAR dans l'ARN messenger viral naissant (Dingwall et al., 1989). Elle joue aussi un rôle dans le recrutement du facteur d'élongation positif de la transcription b (P-TEFb) au niveau de l'ARN viral. Grâce à une cascade de réactions, ce facteur phosphoryle la queue C-terminale de l'ARN polymérase induisant la phase d'élongation de la transcription (Cujec et al., 1997). De plus, Tat semble également favoriser la suppression immunitaire, l'apoptose et le stress oxydatif en facilitant ainsi la progression de l'infection du VIH (El-Amine et al., 2018).

La protéine Nef est une protéine de 23 kDa exprimée majoritairement dans les phases précoces du cycle viral. Elle est essentielle pour maintenir les niveaux élevés de charge virale et joue un rôle dans la destruction des cellules LTCD4+. En effet elle favorise l'absorption de ces cellules immunitaires par le compartiment endosomal et dans les lysosomes (Basmaciogullari and Pizzato, 2014). Nef régule aussi la communication entre les cellules présentatrices d'antigènes (CPA) et les lymphocytes T, en inhibant la voie des récepteurs des cellules T (TCR ; Bell et al., 1998; Münch et al., 2007). De plus, elle régule négativement l'expression des molécules du complexe majeur d'histocompatibilité-I (CMH-I) en protégeant les cellules infectées contre la destruction induite par les cellules LT cytotoxiques (Collins et al., 1998) et interfère aussi avec l'expression du CMH-II (Schindler et al., 2003).

### III.3.2 Protéines auxiliaires Vif, Vpr et Vpu

Les protéines Vif, Vpr et Vpu sont traduites à partir de l'ARN multi-épissé.

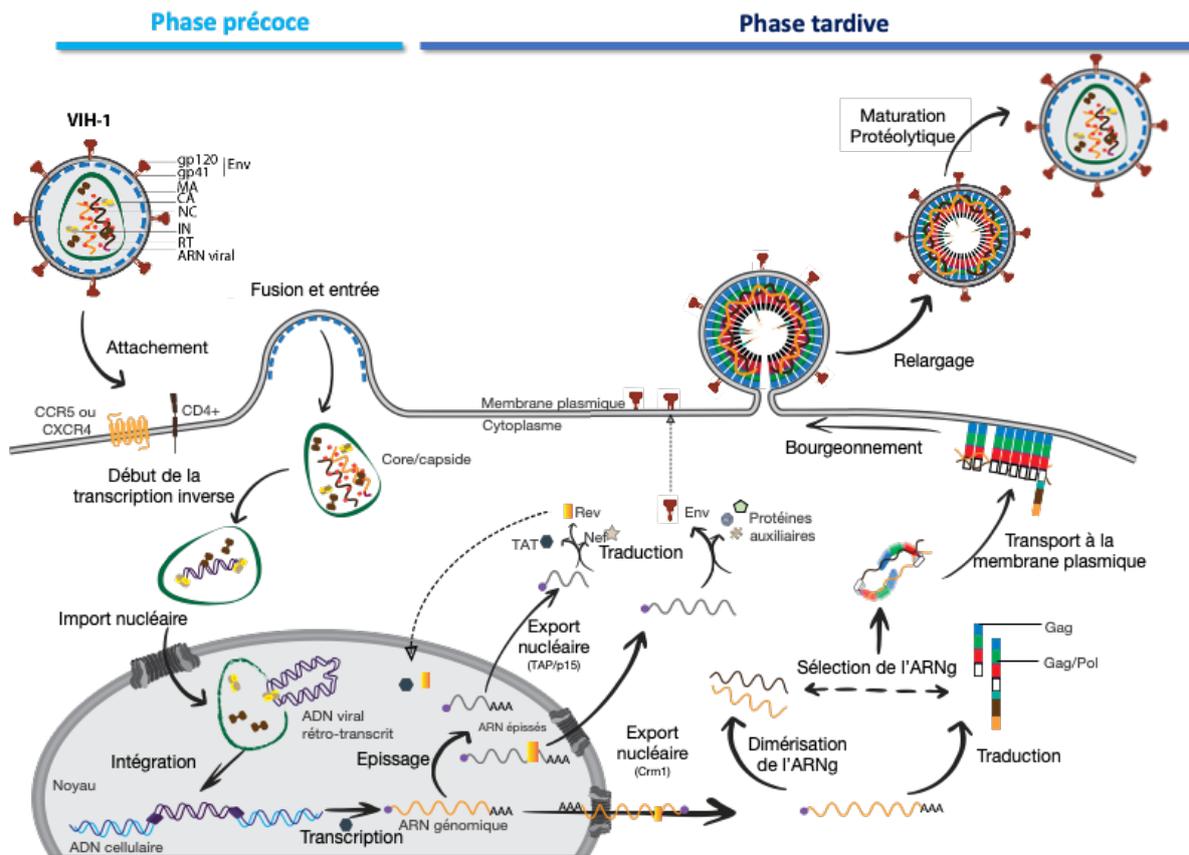
La protéine Vif est une protéine de 23 kDa qui joue un rôle dans les mécanismes d'échappement au système immunitaire notamment en supprimant la réponse antivirale médiée par les enzymes d'édition d'ARNm de l'apolipoprotéine B, qui se comporte comme un polypeptide catalytique (*APolipoproteine B mRNA-Editing enzyme Catalytic polypeptide-like* APOBEC3G et 3F). Ce facteur de restriction cellulaire sera décrit plus en détails dans la section Introduction, §IV.2.

La protéine Vpr est une protéine de 14 kDa qui interfère avec de nombreuses fonctions cellulaires et virales tout au long du cycle viral. Tout d'abord elle régule les étapes précoces du cycle viral. En effet elle joue un rôle dans le transport de PIC vers le noyau grâce à son interaction avec des facteurs de transport nucléaire et des composants du complexe du pore nucléaire. Elle permet aussi l'activation de la transcription et l'arrêt du cycle cellulaire à la transition G2/M. C'est la liaison entre Vpr et le facteur associé à la cycline 4A, DCAF1 qui permet l'arrêt du cycle cellulaire (pour revue voir Morellet et al., 2009.). Cet arrêt a pour conséquence l'augmentation de la transcription virale, puisque la transcription est plus active pendant cette phase cellulaire (Le Rouzic and Benichou, 2005). Vpr induit l'apoptose en perméabilisant la membrane des mitochondries. De plus, elle est également essentielle pour l'infection des macrophages (Kogan and Rappaport, 2011). L'incorporation de Vpr dans les particules virales est assurée par l'interaction avec la séquence riche en leucines (LXXLF) du domaine p6 du précurseur Gag. Des études en utilisant des micelles de dodecylphosphocholine (DPC) mettent en évidence que l'affinité de Vpr pour p6 augmente dans un environnement membranaire (Salgado et al., 2009b). Dans les virus bourgeonnants, un rapport de 7 molécules de Gag pour 1 molécule de Vpr est identifié (Müller et al., 2000). Avec ce rapport, le domaine de p6 reste disponible pour interagir avec ses autres partenaires cellulaires, tels que la protéine *ALG-2 Interacting Protein X* (ALIX) et *Tumor Susceptibility Gene 101* (Tsg101 ; Fossen et al., 2005).

La protéine Vpu est une protéine membranaire de 16kDa. Elle stimule l'exposition d'un grand nombre de glycoprotéines d'enveloppe sur les particules virales néoformées, en réduisant l'expression du corécepteur CD4. En effet en l'absence de Vpu, les protéines Env et CD4 transitent par les mêmes compartiments cellulaires ce qui induit la séquestration des protéines Env en faveur de la formation des complexes avec CD4. Ainsi, Vpu recrute le complexe ubiquitine ligase qui induit la dégradation protéasomique des récepteurs CD4. Elle favorise aussi le relargage des particules virales néoformées en inhibant l'expression du facteur restriction cellulaire tétherine (Van Damme and Guatelli, 2008). Ce facteur sera décrit plus en détails dans la section IV.2

#### III.4 Le cycle de réplication viral

Les cellules cibles du VIH-1 sont majoritairement les cellules de l'immunité adaptative, les cellules lymphocytes T auxiliaires et régulatrices (LT CD4+). Cependant, le virus est aussi capable d'infecter les cellules de l'immunité innée comme les macrophages, les monocytes et les cellules dendritiques (Weiss, 2002). Dans ces cellules le virus réalise un cycle de réplication qui peut être divisé en deux phases (**figure 12**). La phase précoce commence par l'entrée du virus dans la cellule cible et se termine par l'intégration du génome proviral dans l'ADN de l'hôte et la phase tardive qui commence par la transcription et la traduction des protéines virales et se termine par l'assemblage du virus au niveau de la MP et par son bourgeonnement. Une étape supplémentaire de maturation protéolytique est nécessaire pour rendre le virus mature et infectieux.



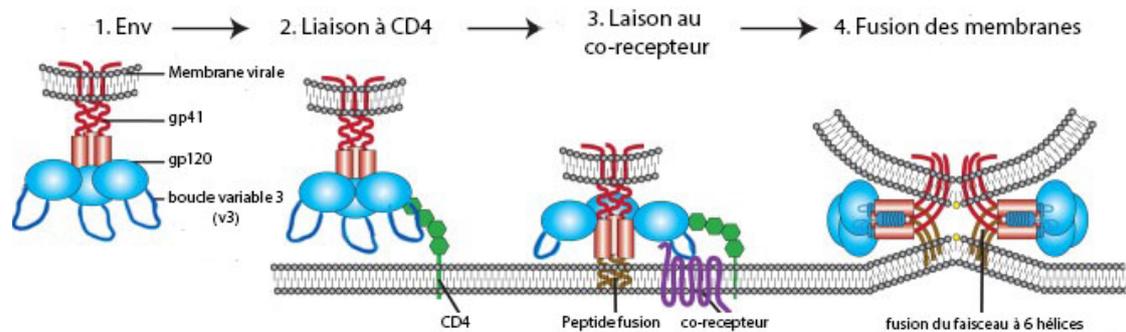
**Figure 12. Cycle répliatif du VIH-1**

Le cycle du virus peut être séparé en deux phases, tout d'abord la **phase précoce** qui commence par l'attachement du virus au récepteur cellulaire CD4 et à ses corécepteurs CCR5 ou CXCR4. Il est suivi de la fusion des membranes virales et cellulaires permettant la libération de la capsid dans le cytoplasme. La capsid permet le transport du PIC le long des microtubules jusqu'au pore nucléaire ou il est internalisé. Une fois la transcription inverse terminée, ce complexe permet l'intégration de l'ADN proviral dans le chromosome cellulaire. Cette étape marque le début de la **phase tardive** du cycle viral. Elle débute par la transcription du provirus. Les ARN sont alors exportés dans le cytoplasme et vont soit, servir de matrice pour la traduction des protéines virales, soit être sélectionnés et encapsidés sous forme de dimères dans les particules virales en formation. La sélection et l'encapsidation de l'ARNg sont orchestrées par le précurseur Gag qui permet également le transport des dimères d'ARNg aux sites d'assemblage à la MP permettant le bourgeonnement de la particule virale. Cette étape est possible grâce au recrutement de partenaires cellulaires comme les protéines ESCRT qui permettent son relargage. D'abord sous forme immature, une étape supplémentaire de maturation protéolytique assurée par la protéase virale est nécessaire pour permettre aux protéines virales de constituer un virus mature et infectieux.

#### III.4.1 Phase précoce – de l'entrée du virus à l'intégration de son ADN viral

Le virus s'attache soit de manière aspécifique grâce à l'interaction des spicules d'enveloppe avec les protéoglycanes à la surface de la cellule (Saphire et al., 2001), soit de manière plus spécifique *via* l'interaction des glycoprotéines d'enveloppe avec les intégrines  $\beta_4$ , ou les récepteurs des cellules dendritiques de type DC-SIGN (Geijtenbeek et al., 2000). Une fois attachée la protéine d'enveloppe gp120 interagit avec le récepteur CD4 des cellules (Klatzmann et al., 1984). Cette interaction induit un changement de conformation d'Env, permettant le repositionnement de la boucle V3 qui peut alors se lier soit, au récepteur de chimiokine C-C de type 5 (CCR5) ; soit,

au récepteur de chimiokine C-X-C de type 4 (CXCR4) qui sont les corécepteurs de l'entrée virale (Raja et al., 2003 ; **figure 13**).



**Figure 13. Représentation de l'entrée du VIH-1 dans les cellules cibles**

**1.** L'enveloppe du VIH-1 est composée de trimères de la sous-unité gp120 représentés en bleu et gp41 représentés en rouge. **2.** Attachement du VIH-1 à la cellule par liaison de la gp120 au récepteur CD4 représenté en vert. **3.** Réarrangement conformationnel permettant la libération du peptide fusion, représenté en marron et la liaison au corécepteur, représenté en violet. **4.** Fusion membranaire suite à la formation du faisceau à six hélices du peptide fusion (adaptée de Wilen et al., 2012).

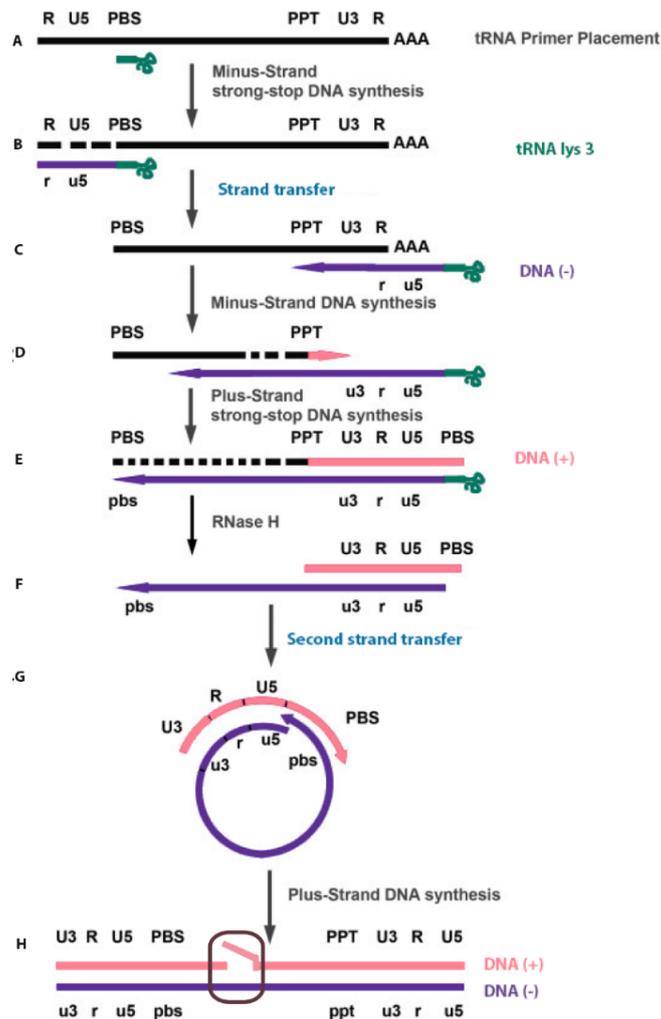
L'affinité pour l'un ou l'autre corécepteur définit le tropisme du virus. On parle alors de tropisme R5 ou macrophagique lorsque l'Env se lie préférentiellement à CCR5 et de tropisme X4 ou lymphocytaire lorsque l'affinité est majoritaire pour CXCR4. Au cours de l'infection, les phases de primo-infection et asymptomatique (voir section Introduction : §IV.1) sont caractérisées par la dissémination des souches de tropismes R5 contrairement à la phase SIDA pendant laquelle le patient est infecté par une majorité de souches de tropisme X4 (Bleul et al., 1997; Esté et al., 1999). Les virus à tropisme R5 représentent la majorité des souches transmises par voie sexuelle. Elles sont peu répliquatives et peu cytopathogènes. De plus, 1% de la population présente une mutation dans le corécepteur CCR5 leurs conférant une auto-immunité face au virus. Cependant au cours du temps et de l'évolution génétique du virus, les souches ayant un tropisme R5 et X4 vont coexister jusqu'à ce qu'une majorité de souches à tropismes X4 subsistent. Elles sont hautement répliquatives et ont un effet cytopathogène prononcé.

L'interaction entre les glycoprotéines d'enveloppe et les récepteurs et corécepteurs cellulaires induit le changement de conformation des deux sous-unités gp120 et gp41 permettant l'insertion dans la MP du peptide de fusion de gp41. La formation du faisceau à six hélices permet alors de rapprocher les deux membranes (virale et cellulaire) et de former le pore de fusion par lequel la capsid virale rentre dans la cellule cible (Killian et al., 2006). A l'issue de l'internalisation de la capsid, il était décrit que la transcription inverse avait lieu dans le cytoplasme avant l'entrée dans le noyau et nécessitait une décapsidation partielle de la capsid. Néanmoins ce dogme a récemment été remis en question en faveur d'un nouveau modèle selon lequel la capsid intacte pénètre

dans le noyau. Ainsi, c'est seulement dans le noyau que l'étape de décapsidation a lieu au niveau des sites d'intégration (Burdick et al., 2020; Dharan et al., 2020; Rensen et al., 2021) et que la transcription inverse se terminerait (Burdick et al., 2020; Francis et al., 2020; Selyutina et al., 2020). Le modèle actuel décrit donc qu'à l'issue de la fusion membranaire, la capsid est libérée intacte dans le cytoplasme, puis elle est transportée activement grâce au système de microtubules jusqu'au complexe du pore nucléaire où elle est activement importée. Pour cela, les protéines de CA interagissent avec les protéines du complexe du pore nucléaire comme Nup153 et Nup358 mais aussi la Cyclophiline A (CypA), la transportine 3 (TNPO3 ; Dharan et al., 2020; Klingler et al., 2020b). Une fois accrochées au pore nucléaire les CA interagissent avec le facteur de clivage et d'épissage 6 (CPSF6) facilitant l'entrée dans le noyau. La transcription inverse initiée dans le cytoplasme se termine à ce stade. A ce moment, la capsid virale perd alors son intégrité et l'ADN retro-transcrit peut être intégré dans le chromosome de l'hôte.

Une étape clé du cycle viral, la transcription inverse permet la rétrotranscription de l'ARN viral en ADN double brin. Elle est initiée par la RT qui utilise comme matrice, l'ARN viral et comme amorce l'ARNt Lysine 3 hybridée à la région PBS de la 5' UTR (**figure 14**). Cette liaison est facilitée et stabilisée par la NC (Auxilien et al., 1999; Tisné et al., 2005). La RT reconnaît spécifiquement le complexe matrice d'amorce ARN/ARN et commence à synthétiser le brin (-) de l'ADN simple brin (ADN sb) à partir de l'extrémité 3' de l'ARNt Lys 3. En parallèle, son domaine RNase H dégrade l'ARN matrice dans l'hybride ARN/ADN créé par la polymérisation de l'ADN. La rétrotranscription s'interrompt lorsque la RT atteint l'extrémité 5' de l'ARNg, après avoir copié les régions U5 et R. La présence de régions complémentaires au niveau de l'extrémité 3' de l'ADN sb (-) et la dégradation de l'ARN matriciel permettent le transfert de la RT (ou saut de brin) vers la région R de l'extrémité 3' de l'ARNg. La RT redémarre alors pour retro transcrire le reste du génome, jusqu'à atteindre la séquence PBS, qui était le point de départ de la rétrotranscription. A l'issue de la synthèse de l'ADN sb (-), la matrice d'ARN est complètement dégradée à l'exception de la région *PolyPurine Tract* (PPT) et de la région PPT centrale (cPPT) résistantes à la RNase H. Ces deux régions servent alors d'amorces de la transcription inverse pour former l'ADN sb (+). La synthèse à partir du PPT à l'extrémité 3' donne naissance à l'ADN sb (+) nommé *strand strong-stop DNA* (+ssDNA). Après le clivage par la RNase H de l'ARNt Lys 3 hybridé au PBS, le +ssDNA s'hybride sur la région PBS complémentaire du brin (-) après un deuxième événement de saut de brin. La synthèse du brin d'ADN (+) reprend et se termine par la formation de l'ADN bicaténaire double brin. L'ADN proviral qui en résulte est flanqué par deux LTR à chaque extrémité. De plus une zone n'est pas reliée au niveau de cPPT, ce court chevauchement est appelé « flap central » (**figure 14**). Comme mentionné précédemment, l'absence d'activité de relecture induit l'apparition de nombreuses erreurs au cours de

la transcription inverse. De plus, les événements de transfert de brin, en particulier, lorsqu'ils se produisent dans des virus hétérozygotes, peuvent amplifier la diversité génétique, en combinant plusieurs polymorphismes différents présents dans les deux copies de l'ARNg présentes dans le virion (Chen et al., 2009; Hu and Temin, 1990).



**Figure 14. Transcription inverse du génome du VIH-1**

L'ARN est représenté en noir. **A.** L'amorce ARNt Lys 3 représentée en vert s'hybride à sa séquence complémentaire, le site de liaison de l'amorce (PBS). **B.** La synthèse du brin d'ADN (-) représenté en violet commence par l'ARNt et continue jusqu'à atteindre l'extrémité 5'. **C.** Après la dégradation du brin d'ARN hybridé à l'ADN par la RNaseH, le brin d'ADN (-) est libéré et peut s'hybrider sur la séquence complémentaire R à l'extrémité 3' par un saut de brin. La synthèse du brin (-) d'ADN continue et l'ARN hybridé à l'ADN est dégradé en parallèle. **D.** Au fur et à mesure de la synthèse du brin de polarité (-), la synthèse de l'ADN du brin de polarité (+) est initiée à partir des séquences PPT, résistantes à la RNaseH. La synthèse à partir du PPT en 3' donne naissance au plus strand strong-stop DNA (+ssDNA) représenté en rose. **E.** Elongation de la synthèse du +ssDNA. **F.** dégradation de l'ARN matriciel par la RNase H. **G.** Après le clivage de l'ARNt Lys,3 hybridé à la région PBS par la RNase H, le +ssDNA s'hybride sur la région PBS complémentaire du brin de polarité positive à l'aide d'un second saut de brin, afin de permettre la synthèse complète des LTR. **H.** La synthèse aboutit à un ADN bicaténaire avec un LTR à chaque extrémité. Ainsi qu'une zone non liée en amont de la région du PPT central (cPPT) qui sera réparée par des protéines cellulaires (Adapté de Basu et al., 2008).

Lorsque les deux étapes de transcription inverse et de décapsidation sont terminées dans le noyau, l'ADN viral néoformé est associé au complexe de pré-intégration (PIC) en interagissant avec un tétramère d'IN. Ce complexe est composé, en plus de l'IN, de plusieurs protéines virales et cellulaires dont la NC, la MA, la RT et Vpr. L'ADN viral est ainsi intégré par l'IN dans le génome de la cellule hôte (Craigie, 2012). Liés à l'ADN viral, les tétramères ordonnés d'IN forment ce qu'on appelle l'intasome. L'IN catalyse alors l'intégration de l'ADN en deux étapes, d'abord le traitement de l'extrémité 3' (ou *3' processing*) et le transfert de brin. Ces deux réactions sont des réactions de transestérifications dépendantes d'une molécule de magnésium. L'IN liée aux LTR de l'ADN viral hydrolyse deux nucléotides de chaque extrémité 3' en utilisant une molécule d'eau, créant deux extrémités 3'OH réactives (Engelman et al., 1991). Suite à cette réaction les extrémités 3'OH libérées restent en interaction avec les acides aminés du site actif de l'IN. Elles sont utilisées pour attaquer les groupes 5'-phosphate dans un sillon majeur de l'ADN chromosomique hôte, conduisant à la transestérification et à la liaison covalente de l'ADN viral au chromosome (Engelman et al., 1991). Il en résulte un intermédiaire d'intégration qui doit être réparé par la machinerie de réparation des jonctions non homologues de l'hôte et abouti à l'intégration complète du génome proviral dans celui de la cellule (Maertens et al., 2022). Une fois intégré l'ADN viral est nommé provirus et son intégration marque la fin de la phase précoce du cycle répliatif.

#### III.4.2 Phase tardive : de l'expression du provirus à la formation d'une nouvelle particule virale.

Une fois intégré, le provirus se comporte comme un gène cellulaire. Cependant, l'endroit où il est intégré va conditionner son devenir. En effet soit son intégration a lieu dans des régions transcriptionnellement actives et il sera efficacement transcrit ; soit dans des régions transcriptionnellement inactives et il restera par conséquent silencieux (Eisele and Siliciano, 2012; Shan et al., 2017). Il est suggéré que le réservoir latent des virus proviennent d'infections de cellules LT CD4+ actives qui passeraient ensuite à l'état de repos (Shan et al., 2017). Néanmoins, les mécanismes régulant la latence ne sont pas encore complètement élucidés à ce jour. La transcription du provirus commence par le recrutement des facteurs de transcription cellulaire au niveau des régions promotrices dans la séquence U3 à l'extrémité 5' NTR. D'autres régions sont reconnues par les facteurs NF- $\kappa$ B et AP-1 qui activent la transcription. La tige boucle TAR joue un rôle important de régulation positive de la transcription. Elle est liée par la protéine virale Tat, qui est l'une des premières protéines virales à être traduite. Suffisamment accumulée, elle rentre dans le noyau où elle se lie aux séquences TAR présentes sur les ARN viraux ce qui permet ainsi de recruter le facteur positif d'élongation transcriptionnelle (P-TEFb). P-TEFb phosphoryle alors la partie C-terminale de l'ARN polymérase induisant sa transition vers la phase d'élongation (Karn and Stoltzfus, 2012). Le produit de transcription du provirus est un ARN messager d'une taille d'environ 9 kb, qui contient de nombreux sites donneurs et accepteurs

d'épissage qui vont moduler le devenir de cet ARN qui pourrait être complètement épissé, partiellement ou non-épissé (ARNg). Dans un premier temps, les ARN messagers viraux totalement épissés sont exportés dans le cytoplasme par la voie canonique qui permet l'expression des protéines Tat, Rev et Nef. Les produits non épissés et partiellement épissés nécessitent l'interaction de Rev avec le RRE pour être exportés du noyau. L'export de ces transcrits utilise la voie *chromosomal maintenance 1* (CRM1 ; Cullen, 1998). En effet Rev contient un signal d'export nucléaire (NES) qui permet de recruter le complexe CRM-1/RAN-GTP (Farjot et al., 1999).

En parallèle, les précurseurs polyprotéiques Gag et Gag-Pol s'accumulent dans le cytoplasme et vont recruter l'ARNg. Je reviendrais plus en détails sur ces étapes du cycle viral section Introduction ; §V. Brièvement, les précurseurs se ressemblent sous forme de petits oligomères (Kutluay et al., 2014) et recrutent l'ARNg sous sa forme dimérique (pour une revue voir (Mailler et al., 2016)). C'est le domaine de NC qui est capable d'interagir spécifiquement avec la région UTR Psi de l'ARNg qui assure son recrutement (Boutant et al., 2020; Ferrer et al., 2016; Lu et al., 2011). Les complexes entre Gag et l'ARNg sont ensuite dirigés vers la MP par un transport actif ciblant particulièrement les inositols phosphates PI(4,5)P2 (Rein et al., 2011). A la MP un changement conformationnel de Gag et notamment de la MA sous forme trimérique permet l'exposition du myristoyl induisant l'ancrage des précurseurs (section Introduction, §III.2). Ce phénomène est appelé le « *myristoyl switch* » (Bussienne et al., 2021; Paillart and Göttlinger, 1999). En outre, l'interaction de la MA avec la MP produirait un changement conformationnel du précurseur à la MP (Rein et al., 2011). A partir de ce moment, seul le domaine NC de Gag n'interagirait avec l'ARNg. De plus, des études *in vitro* avec des membranes modèles montrent que le domaine p6 pourrait aussi interagir avec une forte affinité avec les MP (Salgado, et al. 2009).

Suite à l'ancrage des complexes Gag/ARNg à la MP, d'autres précurseurs Gag sont ensuite rapidement recrutés de manière à former un réseau immature (Ivanchenko et al, 2009; Jouvenet et al, 2008; Kutluay & Bieniasz, 2010). Ce réseau force une courbure au niveau de la MP cellulaire induisant la formation d'une particule sphérique qui finit par bourgeonner à la surface de la cellule (Rose et al., 2020). Le rapport entre Gag et l'ARNg est d'environ 2000 précurseurs Gag dans les particules virales en formation pour un dimère d'ARNg. Pour bourgeonner, Gag et notamment son domaine p6 recrute les éléments ESCRT et notamment les facteurs Tsg101 appartenant aux ESCRT-I et ALIX associé aux ESCRT-III (Garrus et al., 2001; Martin-Serrano et al., 2001). Pendant et/ou après la libération de la particule virale nouvellement formée, la multimérisation du précurseur Gag-Pol active la protéase virale, qui procède à la maturation protéolytique de manière séquentielle et ordonnée des précurseurs Gag et Gag-Pol (Pettit et al., 2005). Ainsi, la première coupure se produit entre SP1 et NC, permettant la formation du complexe ribonucléoprotéique entre le dimère d'ARNg et la NC. Ensuite, la MA et CA sont dissociées ce qui leur permet de

former respectivement le réseau matriciel et le core de la capside, qui contiennent les copies de l'ARNg, ainsi que les protéines RT et IN. Le virus est alors capable de commencer un nouveau cycle viral (Pornillos and Ganser-Pornillos, 2019).

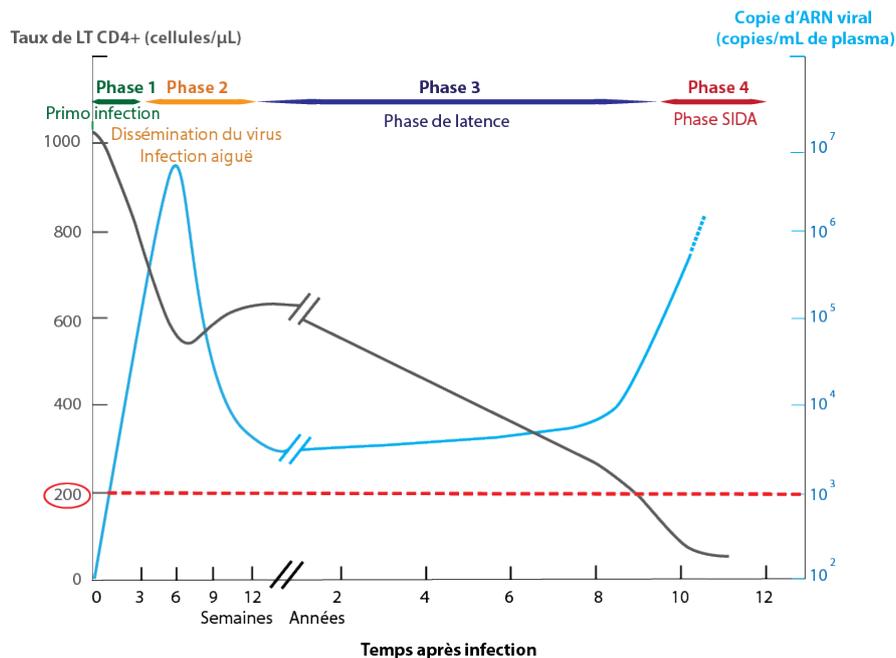
#### **IV. Progression de la maladie et thérapie**

##### **IV.1 Transmission et manifestation clinique du virus**

Le virus est présent dans tous les fluides biologiques des patients qui contiennent des cellules dendritiques, des macrophages ou des lymphocytes T CD4+ (LT CD4+ ;détaillé section Introduction ; §IV.1) comme le sang, mais aussi les liquides séminaux, les sécrétions vaginales et le lait maternel (Levy, 1993).

La voie principale de transmission du virus est lors des rapports sexuels, bien qu'en comparaison avec les autres voies de transmission, la probabilité d'infection est la plus faible allant de 1 sur 200 à 1 sur 3000 (Shaw and Hunter, 2012). Ainsi, en absence d'ART, le risque de transmission de la maladie est modulé par différents facteurs comme le type d'épithélium, le taux de virémie dans les sécrétions cervico-vaginales et rectales et dans le liquide séminal. Certains facteurs induisant le recrutement de cellules immunitaires favorisent le risque de transmission du virus, comme la présence de lésions génitales ou la co-infection par d'autres maladies sexuellement transmissibles (Hladik and McElrath, 2008). Au contraire, la circoncision masculine médicale, par exemple, réduit le risque de contracter la maladie de 60% (Auvert et al., 2005). La seconde voie de transmission du VIH est de la transmission verticale de la mère vers l'enfant. Elle est responsable de la majorité des infections chez les enfants. En absence de traitement, le risque de transmission du virus est de 49%. Elle peut avoir lieu au cours des deuxième et troisième trimestres de la grossesse (8% des enfants infectés), de l'accouchement (15%) ou de l'allaitement (12% à 26% ; Cerveny et al., 2021). Enfin, le partage de seringues ou les accidents médicaux peuvent aussi être vecteurs de la maladie. Le partage de seringue représente 1,5% des transmission du virus en France (Santé publique France, 2021).

L'infection par le VIH induit quatre différents stades de la maladie (**figure 15**).



**Figure 15. Evolution au cours du d'une infection par le VIH-1**

En bleu la courbe du nombre de copie d'ARN viral par millilitres de plasma, en gris le taux de LT CD4+ au cours du temps. Son représenté les 4 phases d'infection du VIH allant de la primo-infection à la phase SIDA. En rouge pointillé la valeurs seuil de LTCD4+ minimale à partir de laquelle le système immunitaire n'est plus capable de fonctionner normalement, en se défendant contre des infections opportunistes par exemple (figure inspirée de Coffin and Swanstrom, 2013).

Tout d'abord, une phase d'éclipse a lieu pendant une à deux semaines, à la suite de la primo-infection. Cette phase est asymptomatique et correspond à la circulation et à la réplication du virus à un niveau indétectable. Elle se déroule avant la mise en place d'une réponse immunitaire adaptative (Coffin and Swanstrom, 2013). Elle est suivie de la phase aiguë qui commence deux à quatre semaines après la primo infection. Cette phase est caractérisée par une augmentation de la charge virale ( $10^7$  ou plus de copies d'ARN par mL de sang), la destruction des LT CD4+ et la mise en place d'une réponse immunitaire adaptative contre le virus. Au moment du pic de virémie, cette réponse correspond à l'expression d'anticorps dirigés contre toutes les protéines virales et d'une réponse des cellules lymphocytaires cytotoxiques CD8+ (LT CD8+) contre les antigènes du VIH-1 exprimés par les cellules infectées. Cette phase s'accompagne généralement de symptômes semblables à un état grippal. A la fin de la phase aiguë, le niveau de virémie diminue fortement. Cette baisse résulte à la fois du contrôle partiel de l'infection par le système immunitaire et par l'épuisement des cellules cibles LT CD4+ (Coffin and Swanstrom, 2013). Une fois la charge virale stabilisée, commence la phase chronique ou phase de latence. En absence de traitement, elle peut durer jusqu' à une vingtaine d'années. Cette phase est quasiment asymptomatique et se caractérise par la multiplication lente du virus et la destruction des LT CD4+ induisant la diminution du taux de ces cellules jusqu'à un seuil critique de 200

cellules/ $\mu\text{L}$  de sang. En dessous de ce seuil, la charge virale augmente considérablement et le système immunitaire est compromis. Le patient n'est alors plus en mesure de se défendre face à des pathogènes opportunistes et/ou au développement de cancer (Coffin and Swanstrom, 2013). En absence de traitements, les patients décèdent suite à des infections opportunistes comme des infections à *Candida albicans* ou des pneumonies mortelles.

## **IV.2 Restriction cellulaire et mécanisme d'échappement du virus**

Le VIH infecte les cellules immunitaires et perturbe profondément l'ensemble du système afin d'échapper aux défenses immunitaires.

### **IV.2.1 La première barrière de défense antivirale : l'immunité innée**

Lorsque le VIH infecte une cellule, il doit surmonter ou désarmer les défenses antivirales cellulaires. Ainsi, une des premières lignes de défense face au virus est la réponse immunitaire innée. Elle est caractérisée par la production de cytokines pro-inflammatoires qui conduisent au recrutement de cellules présentatrices d'antigènes comme les cellules dendritiques (*dendritic cells*, DC). Ces cellules sont impliquées dans le déclenchement de la réponse immunitaire adaptative. En effet elles sont capables de reconnaître et de capturer les antigènes. Les DC sont ensuite activées et maturées en migrant vers les ganglions lymphatiques, où elles présentent les antigènes aux cellules de l'immunité adaptative, les lymphocytes T naïfs, à l'aide des molécules du complexe majeur d'histocompatibilité (CMH) de classe I et II présents à leur surface. La réponse immunitaire innée est essentielle pour activer et diriger les réponses immunitaires adaptatives. Ainsi, la détection des agents pathogènes est la première étape d'une cascade de réactions aboutissant à la réponse adaptative. Les facteurs de restriction en particulier, jouent un rôle important dans les mécanismes de défenses innées contre les virus. Un facteur de restriction antiviral peut être défini comme des protéines naturelles limitant la multiplication d'un virus en inhibant directement ou indirectement certaines étapes de son cycle répliatif. Ces protéines peuvent être constamment produites après induction par la voie interféron. Leur action antivirale induit une forte pression de sélection sur le virus ayant pour conséquence la mise en place de mécanismes d'échappement viral. On observe souvent une compétition entre le facteur de restriction et le mécanisme d'échappement viral se traduisant par une signature d'évolution rapide dans la séquence du gène cellulaire (Malim and Bieniasz, 2012).

Au cours de son cycle répliatif, le VIH est soumis à différents facteurs de restriction cellulaire (**figure 16**). L'une des premières étapes du cycle viral ciblée est l'entrée. Ainsi, les protéines de la famille des sérines incorporeurs (SERINC) qui sont des protéines transmembranaires capables d'incorporer des sérines dans les phosphatidylsérines et les sphingolipides modulent la composition lipidique de la MP. Cette modulation inhibe la fusion correcte de

l'enveloppe virale avec la membrane plasmique, empêchant ainsi le virus de pénétrer dans une nouvelle cellule hôte (Seissler, 2017). La protéine virale Nef empêche l'incorporation de SERINC5 dans les virions en médiant sa relocalisation dans les endosomes tardifs par une interaction avec l'adaptateur de clathrine AP-2 (Trautz et al., 2016).

Une fois que le virus est entré dans la cellule, la protéine 5 $\alpha$  contenant un motif tripartite (*Tripartite motif-containing protein isoform 5 $\alpha$* , TRIM5 $\alpha$ ) inhibe les étapes de transcription inverse et d'import nucléaire en accélérant la décapsidation hôte (Seissler et al., 2017). TRIM5 $\alpha$  forme des trimères qui se lient à la CA et induit des lésions dans la capsid virale. Pour contrecarrer l'activité de TRIM5 $\alpha$ , le VIH recrute la CypA. Cette protéine entre en compétition avec TRIM5 $\alpha$  pour la liaison à la CA diminuant donc l'effet restrictif du facteur (Neil and Bieniasz, 2009). Un autre facteur de restriction est SAMHD1 (*Sterile Alpha Motif and Histidine Aspartate domain-containing protein 1*). Il cible également la phase précoce de l'infection virale et notamment la transcription inverse. Ce facteur exprimé dans les cellules myéloïdes (DC et monocytes) est une désoxynucléotide-triphosphohydrolase dépendante du GTP qui épuise le pool de désoxynucléotides triphosphates cellulaires (Lahouassa et al., 2012). Elle inhibe la transcription inverse en maintenant le taux de dNTPs dans la cellule tellement bas que la transcription inverse est défavorisée. Ce facteur n'est exprimé que dans des types cellulaires précis, car il régule fortement la synthèse cellulaire. Pour l'heure, il n'est pas décrit de défense virale chez le VIH-1 pour contrecarrer SAMHD1, contrairement à Vpx exprimée chez le VIH-2 ou Vpr chez certaines souches de SIV capables de dégrader le facteur de restriction (Matreyek et al., 2014).

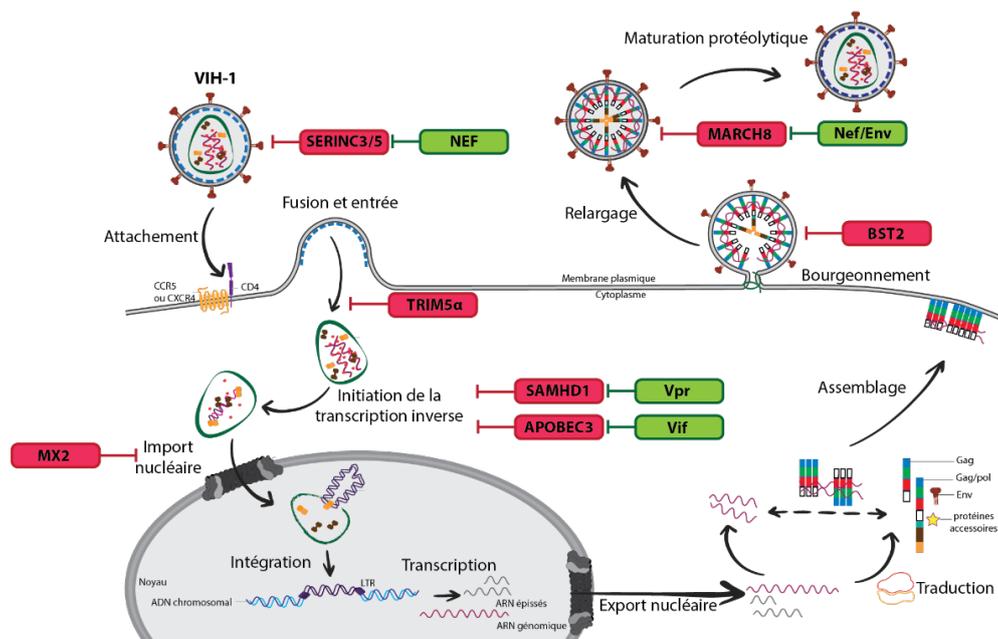
La transcription inverse est aussi ciblée par un autre facteur de restriction, APOBEC. Cette enzyme cellulaire fait partie de la famille des cytidines désaminases capable de catalyser la mutation des C en U (Marquet et al., 2012; Stupfler et al., 2021; Verriez et al., 2020). Elle est composée de sept membres (de A à H), dont seules APOBEC3G (A3G) et APOBEC3F (A3F) ont une activité antivirale contre le VIH-1. C'est la protéine virale Vif qui empêche l'incorporation d'A3G/F dans les particules virales naissantes. Pour cela, Vif interagit avec les protéines A3G ou A3F et recrute le complexe E3 ubiquitine ligase grâce à son site de liaison à la Culline 5, aboutissant à la dégradation d'A3G et/ou A3F.

Les étapes de la phase tardive du cycle du VIH-1 sont aussi ciblées par des facteurs de restriction. Notamment la tetherine aussi appelée BST2 pour *Bone marrow stromal antigen 2*, inhibe la libération de nouvelles particules virales. Cette protéine possède deux domaines transmembranaires qui s'ancrent à la fois dans la MP et dans la membrane virale lors de l'étape de bourgeonnement. Elle empêche alors le relargage des particules virales à la

surface de la cellule. Cette protéine est contrée par Vpu qui induit l'ubiquitinylation de la tetherine, induisant son endocytose et finalement sa dégradation dans les lysosomes (S. Neil et Bieniasz 2009 ; Seissler, 2017)

Finalement, un dernier facteur de restriction est March 8 (Membrane-Associated RING-CH 8 protein), qui diminue l'incorporation des protéines Env dans les virions nouvellement produits, diminuant ainsi leur infectivité (Tada et al., 2015). En plus des facteurs de restriction, il existe d'autres mécanismes cellulaires qui peuvent inhiber la réplication virale. Par exemple la formation de complexes entre la protéine Env et le récepteur CD4 à la sortie du réticulum endoplasmique. Mais une fois de plus, ils existent des mécanismes d'échappement pour le VIH-1. Ainsi, la protéine virale Vpu contrecarre ce mécanisme antiviral en empêchant la formation de ces complexes.

Ainsi, les facteurs de restriction sont parmi les premiers effecteurs de l'immunité innée contre les pathogènes viraux. Ils sont aussi considérés comme des gènes stimulés par la réponse interféron (*Interferon-Stimulated Gene*, ISG) et peuvent donc aussi être induits par les voies de signalisation de l'immunité innée. De plus, ils régulent la signalisation immunitaire et favorisent la reconnaissance des cellules infectées (Malim and Bieniasz, 2012).



**Figure 16. Représentation des facteurs de restrictions cellulaires et des protéines virales qui contrecarrent les défenses cellulaires au sein du cycle viral du VIH-1**

Les principaux facteurs de restriction du VIH-1 sont représentés en rouge et les protéines auxiliaires virales qui contrecarrent ces facteurs sont représentées en bleu (figure adaptée de Seissler et al., 2017).

#### IV.2.3 La deuxième barrière de défense antivirale : l'immunité innée

Une fois que les cellules de l'immunité innées sont présentes dans les organes lymphoïdes secondaires, ils permettent de mettre en place d'une réponse spécifique, la réponse immunitaire adaptative. C'est d'ailleurs pendant la migration des DC dans les ganglions lymphatiques que le VIH-1 infecte les cellules CD4+. Lorsqu'une cellule T naïve reconnaît spécifiquement les protéines associées au CMH-II présentées par des cellules de

l'immunité innée, elle est activée et subit l'expansion clonale. Les cellules T CD4+ représentent une pièce centrale dans la mise en place de réponse immunitaire adaptative. Elles sont nécessaires à l'activation des cellules lymphocytaires T cytotoxiques CD8+ ainsi que des cellules B productrices d'anticorps. Leurs destructions par le VIH expliquent donc l'incapacité du système immunitaire du patient à être actif et l'apparition d'une immunodéficience. La réponse lymphocytaire B n'est pas efficace contre le VIH-1. Le rôle des lymphocytes B est de produire des anticorps qui se lient aux particules virales et qui peuvent inhiber la fixation et l'entrée des virus dans les cellules cibles (Deeks and Walker, 2004). Dans le cas d'une infection par le VIH-1, la réponse anticorps est rarement assez spécifique, permettant l'échappement viral par des mutations des antigènes ciblés. Néanmoins, la réponse immunitaire cytotoxique semble efficace contre le virus et est responsable de la baisse de la charge virale après l'infection aiguë. Les cellules T CD8+ cytotoxiques reconnaissent les cellules infectées grâce aux antigènes viraux présentés à leur surface, sur le CMH-I. Les cellules T CD8+ sécrètent alors des molécules cytotoxiques comme la perforine et la granzyme qui vont tuer la cellule cible. Cependant, la protéine Nef du VIH-1 permet également l'évasion du virus au système immunitaire adaptatif. Elle cible les CMH-I empêchant l'exposition des peptides viraux à la surface de cellules infectées qui sont normalement reconnues et dégradées par les lymphocytes T cytotoxiques (Collins et al., 1998). Néanmoins, les cellules dépourvues de CMH-I sont reconnues et tuées par les cellules de l'immunité innée, les cellules tueuses naturelles (*Natural Killer*, NK ; Deeks and Walker, 2004). Ainsi le succès du VIH-1 réside sa capacité à évoluer et à s'adapter.

### **IV.3 Prévention, dépistage et traitements**

Depuis la découverte du virus, de nombreuses stratégies pour le combattre ont été mises au point. Ces stratégies vont de la prévention, au dépistage massif, passant par la mise en place de traitements non curatifs. Elles ont permis de réduire fortement le taux de mortalité passant de 1,7 million dans les années 2000 à 650 000 en 2021 et le nombre de nouvelles infections, qui est passé de 2,9 millions dans les années 2000 à 1,5 millions en 2021. Ces chiffres sont directement liés à l'augmentation du nombre de personnes ayant accès aux ART, passant de 560 000 dans les années 2000 à 28,7 millions en 2021. Malheureusement, cela ne suffit pas à endiguer la maladie, ainsi l'ONUSIDA a mis au point un plan de lutte pour mettre fin à l'épidémie d'ici 2030. Ce plan a trois objectifs principaux, tout d'abord que 95% des personnes vivant avec le VIH connaissent leur statut sérologique. Ensuite, que 95% des personnes connaissant leur séropositivité reçoivent un traitement. Et finalement que 95% des personnes sous traitement aient une charge virale supprimée. L'absence de détection de la charge virale assure la non-transmission du virus d'un individu à l'autre (« indétectable = intransmissible »). Le deuxième objectif du plan de l'ONUSIDA est de réduire le nombre de nouvelles infections à 200 000 dans le monde. Enfin, le dernier objectif est

de faire disparaître les discriminations qui empêchent les personnes infectées ou à risques de faire le nécessaire en réalisant des dépistages et en mettant au point des stratégies préventives et/ou curatives (ONUSIDA, 2022). En 2021 seulement 85% des personnes vivants avec le VIH connaissent leur statut sérologique, 88% de ces personnes avaient accès à un traitement et 92% avaient une charge virale indétectable (ONUSIDA, 2022).

#### IV.3.1 La première étape vers la fin de l'épidémie, le dépistage

La connaissance exacte du statut sérologique du patient est un élément clé dans la mise au point des stratégies de prévention. Dans ce sens, seules les personnes diagnostiquées peuvent bénéficier d'un traitement ART réduisant le risque de propagation du virus et de développement de symptômes graves. De nos jours, il existe différents moyens de dépister la présence du virus, en réalisant des tests sérologiques, en faisant des tests rapides à orientation de diagnostic (TROD) ou en réalisant des auto-tests.

Les tests sérologiques sont réalisés exclusivement en laboratoires d'analyses médicales. Ils consistent à rechercher la présence d'anticorps dirigés contre des antigènes du VIH à partir d'une prise de sang, et plus récemment d'antigènes (p24). Pour cela un test immunologique *enzyme-linked immunosorbent assay* (ELISA) dit de 4<sup>ième</sup> génération est réalisé. Ce test détecte les immunoglobulines G et M (IgG et IgM) dirigés contre les protéines d'enveloppe du VIH-1 (Gp41) et du VIH-2 (Gp36 ; (Hurt et al., 2017). En France, un délai de 6 semaines après l'exposition est demandé pour valider un résultat négatif afin d'être dans la phase aigüe ou de latence (**figure 15**). Si le test est positif, un western-Blot est réalisé sur le même prélèvement sanguin pour confirmer le résultat et discriminer une infection par le VIH-1 d'une infection par le VIH-2. Si les deux tests réalisés sur le même prélèvement sont positifs, alors un second prélèvement devra obligatoirement être réalisé afin d'éliminer une erreur d'identité (HAS-santé, 2019).

Le TROD donne un résultat dans un délai court allant de 10 à 30 minutes en fonction des tests et ne nécessite pas de matériels d'analyses biologiques particuliers. Ils sont réalisés à partir d'une goutte de sang total (Vikia, Deterine) ou de fluide oral (OraQuick). Il existe deux grandes stratégies de détection par immuno-chromatographie ou par immuno-filtration. La majorité des tests mis sur le marché actuellement ne détectent que les AC dirigés contre les protéines d'enveloppe du VIH-1 (Gp41) et du VIH-2 (Gp36). Si le test est positif, le résultat doit être confirmé par un test sérologique et suivre les voies de confirmation décrites précédemment. Si le test est négatif, la personne testée est informée des limites des TROD qui sont moins sensibles que les tests sérologiques. Les TROD peuvent être réalisés dans des cadres précis (associations, pharmacie, planning familial).

Finalement, des auto-tests peuvent être réalisés dans un environnement domestique par des profanes. Ces derniers n'ont pas de valeurs juridiques (HAS-santé, 2019).

### IV.3.2 La prévention comme arme contre la propagation du virus

En absence de traitements curatifs contre la maladie, le seul moyen d'empêcher la propagation du VIH est de prévenir l'infection en réduisant l'exposition aux facteurs de risques. Il existe différentes solutions au niveau individuel permettant la protection face au virus (ONUSIDA, 2022). Tout d'abord pour réduire les risques liés à une exposition sexuelle, l'utilisation de préservatifs externes/masculins ou internes/féminins (ONUSIDA, 2022) et la circoncision médicale volontaire de l'homme peuvent réduire fortement les risques de transmission (Auvert et al., 2005). En parallèle, la prévention et le dépistage régulier du VIH et des autres IST est un point important et non négligeable pour se protéger et protéger les autres (Simon et al., 2006). Afin de réduire les risques liés au partage de seringues entre consommateurs de drogues injectables, des centres d'accueil et d'accompagnement à la réduction des risques des usagers de Drogues sont ouverts (par exemple, la salle de consommation à moindre risques de Strasbourg). Ces centres permettent la distribution de matériels injectables stériles, la sensibilisation des consommateurs et l'accompagnement vers les soins des personnes dépistées positivement (AIDES, 2022 <https://www.aides.org/>). De plus, la transmission verticale de la mère vers l'enfant peut être totalement évitée en prenant des ART en continu (Gupta et al., 2020).

En outre, il existe des traitements médicamenteux préventifs, nommé prophylaxie préexposition (PrEP). Cette protection repose sur l'utilisation préventive d'agents antirétroviraux pour empêcher les nouvelles infections à VIH chez les personnes séronégatives ayant un comportement à risque (Phanuphak and Gulick, 2020; Spinner et al., 2016). Actuellement, il existe deux moyens de prendre ce traitement. D'une part, de manière continue, en prenant 1 comprimé tous les jours à la même heure pendant une période donnée. Dans ce cas, le traitement est efficace après une prise continue pendant 7 jours pour les hommes et 21 jours pour les femmes. D'autre part, le traitement peut être pris à la demande. Cette stratégie n'est possible que pour les hommes et consiste à prendre deux comprimés entre 24 heures et 2 heures avant le rapport sexuel à risque et un comprimé après 24 heures et un autre après 48 heures (Spinner et al., 2016). En France, seul le médicament TRUVADA et ses génériques sont autorisés pour la PrEP (HAS, 2022). Le Truvada contient deux principes actifs, l'emtricitabine qui est un inhibiteur nucléosidique et le ténofovir disoproxil qui est un inhibiteur nucléotidique de la transcriptase inverse (Phanuphak and Gulick, 2020; Spinner et al., 2016). En plus de la PrEP qui permet de prévenir l'infection, il existe aussi un traitement d'urgence post-exposition (TPE). Ce traitement doit être consommé dans les 48 heures après exposition d'une personne séronégative. C'est une tri-thérapie composé de TRUVADA associé à une autre molécule qui cible d'autres protéines virales comme l'intégrase (Raltegarvir) ou la protéase (Darunavir ou Atazanavir).

En conclusion, le virus ne peut pas se propager si la charge virale est indétectable. Par conséquent, les mesures pour améliorer l'accès au dépistage et favoriser l'accès au traitement antirétroviral sont une stratégie non négligeable de la prévention du VIH.

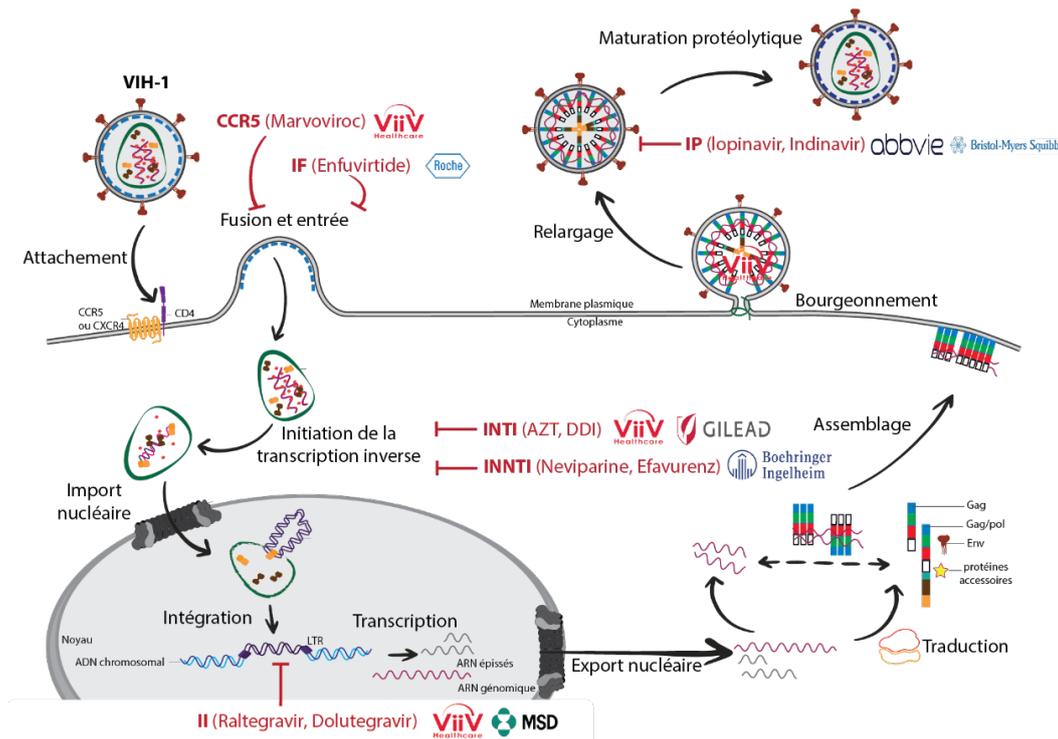
#### IV.3.3 Les traitements comme dernières barrières contre la propagation du virus

La stratégie la plus efficace contre le virus serait de trouver un traitement curatif. L'intégration de l'ADN proviral dans des cellules immunitaires forme un réservoir latent empêchant l'éradication complète du virus. Si de nombreuses stratégies sont encore en voie expérimentale, comme par exemple la thérapie antivirale séquentielle à libération lente et à action prolongée (LASER ART) couplée à CRISPR-Cas9, il n'existe pour l'heure aucun traitement curatif sur le marché (Dash et al., 2019).

Néanmoins, les ART sont efficaces pour inhiber la réplication virale à un niveau indétectable. Pour cela, plusieurs médicaments dirigés contre le VIH-1 sont approuvés et mis sur le marché, depuis le début de la pandémie. Ces composés peuvent être divisés en cinq groupes, ciblant cinq grandes étapes du cycle viral comme l'entrée virale, la fusion, la transcription inverse, l'intégration et l'étape de maturation protéolytique (**figure 17**).

Premièrement, les inhibiteurs de l'entrée virale bloquent la liaison de virus avec le récepteur et le corécepteur cellulaire. Ce sont des inhibiteurs CCR5 qui bloquent les virus de tropismes CCR5 (Kuritzkes, 2009). Ensuite, les inhibiteurs de fusion empêchent la fusion membranaire. L'enfuvirtide est le seul représentant de cette famille et s'administre sous forme de deux injections par jour sous la peau. Il existe aussi les inhibiteurs de la transcription inverse qui sont regroupés en deux catégories, les inhibiteurs nucléosidiques de la transcription inverse (INTI) et les inhibiteurs non nucléosidiques de la transcription inverse (INNTI). D'une part, les INTI sont des analogues nucléosidiques terminateurs de chaîne. Ils entrent en compétition avec les désoxynucléotides naturels normalement incorporés dans la chaîne ADN naissante par la transcriptase inverse. Une fois incorporés, les INTI mettent fin à la synthèse de l'ADN (Menéndez-Arias, 2008). D'autre part, les INNTI se lient à la RT et bloquent les changements de conformation nécessaires à la polymérisation (Kohlstaedt et al., 1992). Une association de deux INTI et d'un INNTI est le plus couramment donné en ART de première attention, notamment dans le cadre de la PrEP et du TPE (Phanuphak and Gulick, 2020). L'étape d'intégration de l'ADN viral est aussi ciblée par des ART. Dans ce cas, les inhibiteurs de l'intégrase peuvent être séparés en deux catégories, les inhibiteurs de transfert de brin (INSTI) qui se lient au site catalytique et les inhibiteurs allostériques de l'intégrase (ALLINI) qui bloquent la dimérisation de l'IN en se liant à la surface de dimérisation entre deux monomères de la protéine. Finalement, les inhibiteurs de la protéase (IP) imitent le substrat naturel de la protéase et entrent en compétition pour se lier au site actif de l'enzyme, empêchant ainsi la maturation des particules virales (Sundquist and Kräusslich, 2012). Le

traitement thérapeutique le plus courant aujourd’hui est la thérapie antirétrovirale combinée (cART). Elle consiste à cibler différentes étapes du cycle viral en même temps (Cihlar and Fordyce, 2016). Actuellement, la combinaison d’INSTI de deuxième génération (dolutegravir ou bictegravir) avec deux INTI est le plus souvent donnée comme cART (Saag et al., 2020). À ce jour, plus de 23 associations de médicaments contre le VIH-1 ont été approuvées par la *Food and Drug Administration* (FDA). Ils se présentent sous la forme d’un seul comprimé quotidien ayant peu d’effets secondaires.



**Figure 17. Représentation des étapes du cycle viral du VIH-1 ciblées par les ART disponibles sur le marché**

Les type d’ART sont écrit en rouge au niveau des étapes du cycle viral qu’ils ciblent. Des exemples de molécules sont donnés pour chaque traitement ainsi que les laboratoires les produisant.

Cependant, tous les médicaments actuellement disponibles peuvent induire une pression de sélection sur le virus et permettre l’apparition de variants résistants. Ces mécanismes d’échappement sont la principale cause d’échec des traitements et justifient la recherche constante de nouveaux médicaments et des nouvelles stratégies. Pour cela, il est nécessaire de comprendre l’intégralité du cycle viral et de supprimer les parts d’ombres.

## V. Des étapes particulières : L'Assemblage du virus

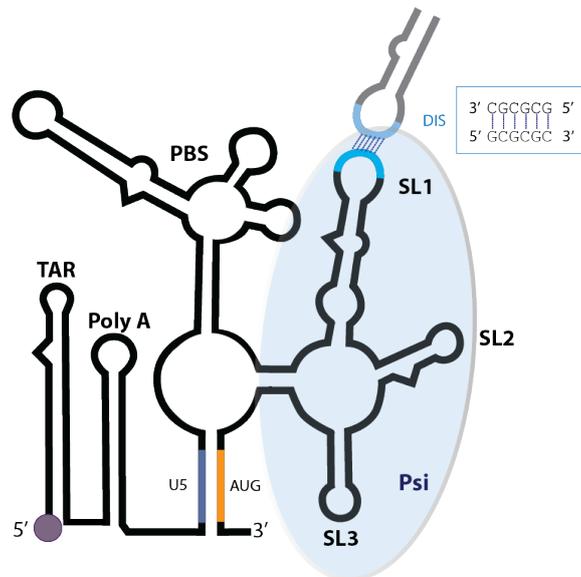
Parmi les étapes du cycle viral, les étapes conduisant à l'encapsidation de l'ARNg ne sont pour l'heure ciblées par aucun des ART disponibles sur le marché. Or, ces étapes sont cruciales pour la production d'un virus infectieux. Cependant, les mécanismes de reconnaissances spécifiques de l'ARNg sont mal compris de nos jours. Les comprendre favoriserait le développement de stratégies antirétrovirales innovantes.

### V.1 Sélection spécifique de l'ARNg par Gag

Pour que l'assemblage de la particule virale soit possible, il est nécessaire que l'ARNg soit spécifiquement sélectionné, parmi un grand nombre d'ARN cellulaires et d'ARN viraux épissés. C'est une étape clé pour la production de virus infectieux.

#### V.1.1 Dimérisation de l'ARNg

Au sein de mon laboratoire d'accueil, il a été démontré que la région 5' UTR de l'ARNg contient tous les éléments nécessaires au recrutement de la polyprotéine Gag (Aldovini and Young, 1990; Berkowitz et al., 1993; Clavel and Orenstein, 1990; Harrison et al., 1998; Lever et al., 1989; Luban and Goff, 1991). Comme décrit dans l'introduction (§III), cette région est hautement structurée et présente de nombreuses tiges boucles (**figure 18**). De l'extrémité 5', vers l'extrémité 3' se trouvent les tiges boucles TAR, puis poly A, qui font partie de la région R. Elles sont séparées par la région U5 de la tige boucle PBS. Enfin, la région  $\Psi$  est composée de quatre tige boucles (*Stem Loop ; SL*) SL1 à SL4 (Blakemore et al., 2021). SL1 et SL3 correspondent aux sites majeurs de l'encapsidation (Paillart et al., 1996; Skripkin et al., 1994), SL2 contient le site majeur d'épissage (*Splicing Donnor, SD*) et empêche le clivage et la polyadénylation au niveau de la région 5' du site de polyadénylation. SL4 contient le codon d'initiation AUG du gène *gag*. De plus, SL1 est le site d'initiation de la dimérisation de l'ARN viral (*Dimerization Initiation Site, DIS*). Elle initie cette dimérisation grâce à une séquence auto-complémentaire palindromique (GUGCAC) dans sa boucle apicale, capable de former un complexe ARN/ARN par une interaction appelé « *kissing loop* » (Paillart et al., 1996). De plus, la région 5' UTR de l'ARNg peut adopter plusieurs conformations qui favorisent soit l'étape de traduction, soit le recrutement de Gag (Gilmer et al., 2021; Keane et al., 2016).



**Figure 18. Représentation des interactions tertiaires de la région 5'UTR de l'ARNg**

Représentation de la structure secondaire de la région 5' non traduite (UTR) de l'ARNg viral. De gauche à droite se trouve le site d'activation de la transcription (TransActivation Response element, TAR) et le signal de poly-Adenylation (polyA), faisant partie de la région répétée (R). Elle est suivie de la région unique U5, en mauve. Ensuite, se trouve le site de fixation de l'amorce (Primer Biding Site, PBS), la tige boucle SL1 dont le site d'initiation de la dimérisation (Dimerization Initiation Site, DIS) est représenté en bleu, les tiges boucles de SL2 et SL3. Les tiges boucles de SL1 à SL3 appartiennent à la région d'empaquetage (Packaging Signal initiation, Psi) entourée en bleu (d'après (Gilmer et al., 2021).

Plusieurs études montrent que la dimérisation de l'ARNg est un prérequis à son encapsidation et au maintien de l'intégrité du génome du VIH (Jouvenet et al., 2009; Moore et al., 2009; Paillart et al., 2004). Elle est nécessaire pour le recrutement de la protéine Gag dans le cytoplasme (Bernacchi et al., 2017; El-Wahab et al., 2014). Le passage d'une conformation monomérique à dimérique, serait contrôlé par des changements conformationnels importants (Keane et al., 2016; Lu et al., 2011). Des études par RMN montrent que la structure monomérique de l'ARN est favorisée par le repliement de l'AUG sous forme d'une tige boucle. La région U5 est alors libre et forme une interaction en pseudo-nœuds avec SL1 (Blakemore et al., 2021; Kenyon et al., 2013). SL1 étant séquestrée, elle ne peut plus former de *kissing loop* (Ye et al., 2022). L'ARN est alors utilisé comme modèle pour la traduction de la protéine Gag (Lu et al., 2011). Au contraire, la dimérisation de l'ARN est favorisée par un appariement de bases à longue distance entre l'AUC et U5 (Song et al., 2008; Yang et al., 2022; Ye et al., 2022). De plus, le site d'initiation de la transcription de l'ARNg présente une hétérogénéité produisant des produits de transcription commençant par un nombre variable de résidus guanines (1G, 2G ou 3G). Le nombre de G aurait un impact sur la structure de l'ARN et jouerait un rôle dans la dimérisation de ce dernier (Brown et al., 2020; Kharytonchyk et al., 2016). De cette manière, les produits de transcription avec 1G favoriseraient l'exposition du DIS et donc la dimérisation et en parallèle la coiffe 5' serait séquestrée (Ye et al., 2022). Tandis que les variantes 3G dissimulent le DIS tout en

exposant la coiffe améliorant la traduction (Brown et al., 2020; Ye et al., 2022). Une fois dimérique, l'ARNg s'associe avec des multimères de petites tailles de Gag (Jouvenet et al., 2008; Kutluay et al., 2014). Ils forment alors une ribonucléoprotéine virale.

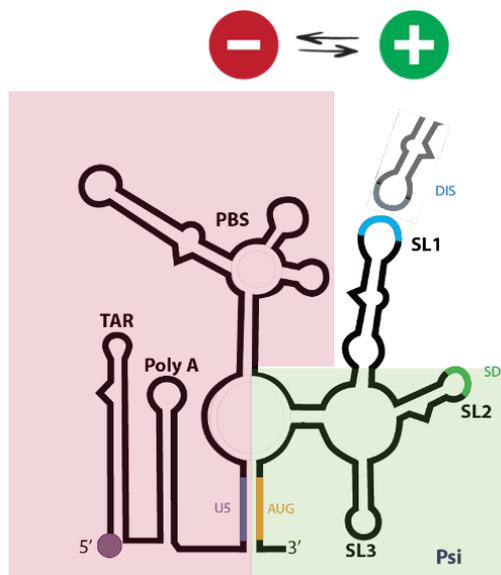
### V.1.2 Interaction entre Gag et l'ARNg

L'association entre Gag et l'ARNg a lieu dans le cytoplasme (Chouaib et al., 2020; Ferrer et al., 2016; Hendrix et al., 2015) et implique un nombre limité de Gag (Kutluay et al., 2014). En effet, lorsque des mutations empêchent la multimérisation de Gag, l'encapsidation de l'ARNg est drastiquement diminuée. Des études *in vitro* identifient que Gag est probablement trimérique en solution (Dubois et al., 2018a) et que six molécules de Gag s'associent avec une grande affinité à l'ARNg dans le cytoplasme (Bernacchi et al., 2017). De plus, le domaine de NC est identifié comme responsable de l'oligomérisation de Gag dans le cytoplasme (Hendrix et al., 2015; Olson and Musier-Forsyth, 2019). Ce domaine est aussi fortement impliqué dans l'interaction du précurseur avec l'ARNg. Cette interaction est régulée par les régions basiques de la NC et par les deux motifs en doigt de zinc, CCHC. La NC interagit avec des régions fortement enrichies en purines et particulièrement en résidus G (Gilmer et al., 2021). Cependant, l'action des deux motifs en doigts de zinc n'est pas équivalente (Boutant et al., 2020; Retureau et al., 2019). Une étude démontre que le deuxième doigt de zinc (ZF2) jouerait un rôle plus important dans l'accumulation des complexes Gag/ARNg à la MP (Boutant et al., 2020). En accord avec ces résultats, il a été également montré que ZF2, en étant plus accessible, contribue à stabiliser le complexe entre l'ARNg et Gag (Retureau et al., 2019). De plus, les protéines NCp7 et Gag possèdent une activité chaperonne vis-à-vis des acides nucléiques. Le précurseur Gag présente une meilleure affinité pour l'ARNg que la protéine mature NCp7 (Godet et Mély 2010; Rein, Henderson et Levin 1998). De plus, les propriétés de liaison à l'ARNg pendant la formation du virion seraient finement modulées par les différents produits de maturation de l'extrémité C-terminale du précurseur (Mouhand et al., 2020, 2018). De manière intéressante, d'autres domaines de Gag sont capables d'interagir avec les acides nucléiques comme le domaine MA (Kutluay et al., 2014; Olson and Musier-Forsyth, 2019; Thornhill et al., 2019). En effet il est proposé que la MA recrute des ARNt dans le cytoplasme (Kutluay et al., 2014). Ces interactions empêcheraient les complexes ribonucléoprotéiques de se lier à des membranes dépourvues de PI(4,5)P<sub>2</sub> (Olson and Musier-Forsyth, 2019). De plus, l'interaction avec le dimère d'ARNg pourrait induire la multimérisation du précurseur en oligomères de petite taille, probablement grâce à des interactions entre les domaines de CA (Kutluay and Bieniasz, 2010). Le domaine Sp1 favorise l'encapsidation de l'ARNg (Kaye and Lever, 1998) et module la spécificité de sélection des ARNg par rapport aux ARN viraux (Roy et al., 2006). Finalement, des analyses de mutagenèses montrent que le domaine riche en proline, p6, intervient aussi dans le recrutement spécifique de l'ARNg (Dubois et al., 2018a; Tanwar et al., 2017).

### V.1.3 Discrimination entre les ARN viraux et cellulaires.

En plus de l'ARNg, les ARN viraux épissés et des ARN cellulaires comme l'ARN 7SL, l'ARN sn U6 et l'ARNm GAPDH sont retrouvés dans les particules virales.

Le rôle de ces ARN n'est pour autant pas encore élucidé (Houzet et al., 2007). Des analyses de mutagenèses montrent que les régions en amont de la tige boucle SL1 (les régions TAR, le poly-A, PBS) sont des signaux négatifs pour l'encapsidation de l'ARN (El-Wahab et al., 2014). Or ces domaines sont présents sur l'ARNg mais aussi sur les ARN épissés. Tandis que, lorsque la région allant de SL2 au codon d'initiation de *gag* est délétée, l'encapsidation de l'ARNg est fortement réduite (Aldovini and Young, 1990) ce qui montre bien l'importance de l'intégrité de la région Psi. Il en est de même lorsque les 40 premiers nucléotides du gène *gag* sont mutés (Luban and Goff, 1994). En parallèle, il est montré que les ARN mono-épissés et complètement épissés sont enrichis à des niveaux similaires dans les virions. Cela montre que ni le domaine REV, ni la tige boucle SL1 ne jouent de rôle dans leurs encapsidations. Néanmoins, la délétion de SL1 réduit significativement l'efficacité d'encapsidation de l'ARNg et s'accompagne de l'augmentation d'incorporation des ARN viraux épissés (Houzet et al., 2007). Ces résultats montrent que les ARNg et ARN épissés sont en compétition pour la fixation de Gag et que l'encapsidation des ARN viraux épissés n'est pas déterminée par SL1. Ainsi, le modèle impliquant des interactions à longue distance entre les régions en amont et en aval de Psi permettrait d'expliquer la régulation de la spécificité de recrutement de l'ARNg par le précurseur (**figure 19**).



**Figure 19. Modèle des interactions tertiaires de l'extrémité 5' UTR de l'ARNg impliquées dans la régulation de la liaison spécifique avec Gag**

*L'interaction tertiaire à longue distance entre les régions en amont et en aval de Psi, induisent le recrutement Gag. La région en amont de SL1, présente sur les ARN épissés et non épissés, régule négativement l'interaction avec le précurseur Gag. Son effet est annulé par la présence des régions en amont de SL4 uniquement présente dans l'ARNg (El-Wahab et al., 2014).*

Selon ce modèle, des interactions à longues distances impliquent les régions génomiques en amont et en aval de SL1 permettant la régulation du recrutement de l'ARNg par Gag. D'une part, la région en amont de SL1, présente sur tous les ARN viraux (épissées et non épissés), inhibe le recrutement de la protéine. D'autre part, la région en aval du gène Gag, uniquement présente sur l'ARNg non épissé, a un contrôle positif sur le recrutement de Gag (El-Wahab et al., 2014).

De plus, l'expression de Gag dans des cellules eucaryotes en absence d'ARN viraux aboutit à la formation de VLP dans lesquelles des ARN cellulaires sont encapsidés (Freed, 2015b). De manière intéressante, des analyses dans lesquelles les domaines NC et sp2 sont remplacés par un motif *leucine zipper* dans un contexte viral, inhibant le recrutement spécifique de l'ARN par Gag, n'empêche pas la formation des particules virales (Mattei et al., 2015). Ainsi, les ARN cellulaires peuvent être encapsidés et cela se ferait par des mécanismes différents (Houzet et al., 2007).

## **V.2 Assemblage du virion à la membrane plasmique**

### **V.2.1 Transport du complexe Gag/ARNg à la membrane plasmique et assemblage à ces sites**

De nombreux résultats tendent à prouver que l'étape de sélection de l'ARNg dimérique et son encapsidation sont couplées lors du cycle viral (El-Wahab et al., 2014; Freed, 2015b; Mailler et al., 2016; Sundquist and Kräusslich, 2012). Ainsi, une fois les complexes formés entre un oligomère de petite taille de Gag et le dimère d'ARNg dans le cytoplasme (Kutluay et al., 2014), ils sont transportés vers la MP jusqu'aux sites d'assemblages (Boutant et al., 2020).

Bien qu'un réseau de filaments d'actine se retrouve aux niveaux des particules virales et des sites de bourgeonnement du virus, le mécanisme de transport n'est pas clairement compris à ce jour (Jouvenet et al., 2006; Mailler et al., 2016). En parallèle, il est suggéré que la protéine Gag peut utiliser la machinerie de trafic vésiculaire intracellulaire de l'hôte (Lehmann et al., 2009; Perlman and Resh, 2006). Néanmoins, en absence de Gag, des analyses en microscopies à fluorescences montrent que l'ARN n'est pas activement transporté dans le cytoplasme (Chen et al., 2014). De plus, dans ces conditions, l'ARNg n'apparaît que de manière transitoire à la MP (Jouvenet et al., 2006). Au contraire, sous forme de complexes avec Gag, l'ARNg est ancré à la MP et cet ancrage est toujours suivi de l'accumulation massive du précurseur (Jouvenet et al., 2008). De plus de nombreuses protéines cellulaires sont identifiées dans les complexes Gag/ARN comme l'*ATP binding cassette subfamily E member 1* (ABCE1)

(Lingappa et al., 2006), Staufen1 (Mouland et al., 2000) et l'hélicase 6 DEAD-box (DDX6) (Reed et al., 2012), leurs rôles exacts ne sont pas connus.

Il est intéressant de noter que la dimérisation de l'ARNg a été considérée par une partie de la communauté scientifique comme se déroulant à la MP (Chen et al., 2016). Cependant, de plus en plus de preuves montrent que l'ARNg est recruté sous forme dimérique par la protéine Gag dans le cytoplasme comme discuté précédemment (Boutant et al., 2020; Hendrix et al., 2015; Jouvenet et al., 2006). Néanmoins, il serait possible que la protéine Gag stabilise le dimère d'ARNg dans le cytoplasme jusqu'au recrutement d'un grand nombre de Gag à la MP (Ferrer et al., 2016). A ces sites d'assemblage, jusqu'à 2500 protéines Gag sont recrutées (Sundquist and Kräusslich, 2012) et les interactions entre Gag et l'ARNg deviennent non spécifiques pour permettre à la particule virale de se former au tour du génome viral (voir section Objectifs). Ce changement de spécificité dans les interactions Gag-ARNg, entre la sélection spécifique de l'ARNg dans le cytoplasme (Ferrer et al., 2016) impliquant quelques copies du précurseur (Kutluay et al., 2014), pourrait, selon notre hypothèse, être assuré par les partenaires cellulaires et viraux du domaine p6 (voir section Objectifs). Le bourgeonnement peut alors débuter. La multimérisation de Gag est médiée par son domaine CA (Olson and Musier-Forsyth, 2019). Dans la particule immature, les molécules Gag sont alignées et encapsidées radialement.

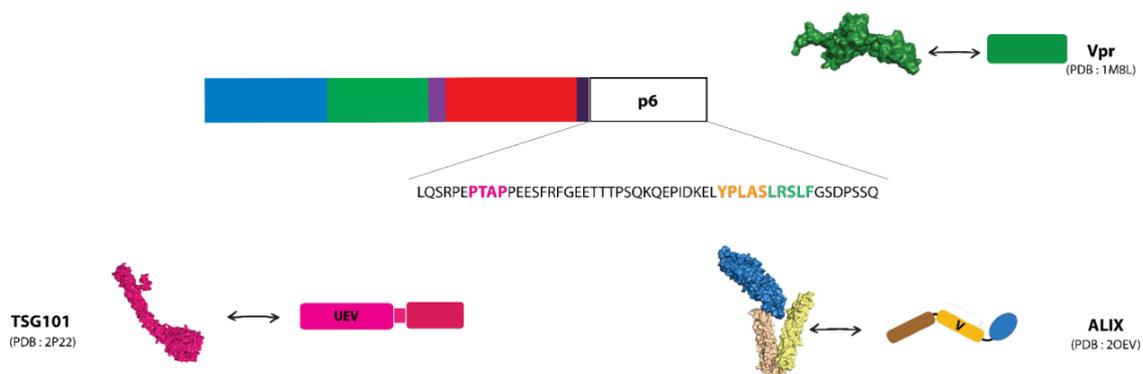
### V.2.2 Interaction avec les facteurs ESRT et bourgeonnement de la particule virale

Aux sites d'assemblage, la MA est enchâssée dans la MP, la CA forme est liée au feuillet interne de la MP et le domaine p6 est orienté vers le centre de la particule (Freed, 2015b). Ce domaine est particulièrement important pour le relargage viral, puisqu'il recrute la machinerie ESCRT. Ce petit domaine est composé de deux régions basiques appelées « domaines tardifs » car leur fonction est requise à un stade tardif du cycle de réplication (Welker et al., 2021). Ainsi, le motif PTAP est impliqué dans le recrutement de la protéine TSG101 (Garrus et al., 2001; Martin-Serrano et al., 2001) et le motif LYPXnL permet le recrutement d'ALIX (Martin-Serrano et al., 2001; Strack et al., 2003). Ces partenaires jouent un rôle dans la biogenèse des vésicules qui bourgeonnent dans les endosomes tardifs pour former des corps multivésiculaires (CMV) (Sundquist and Kräusslich, 2012).

TSG101 est une protéine de 44 kDa composée de 3 domaines (**figure 20**). Elle fait partie de la machinerie ESCRT-I. Couplée avec les autres complexes ESCRT (-0, -II et -III) et le complexe AAA ATPase VPS4, cette protéine joue un rôle essentiel dans le tri des protéines cargo ubiquitinylées dans les vésicules. Tsg101 se lie uniquement au domaine p6 (Garrus et al., 2001; Martin-Serrano et al., 2001). Cependant, des analyses de mutagenèse sur le domaine NC et notamment sur les doigts de zinc montrent une baisse du relargage viral (Chamontin et al., 2015). Ces résultats tendent à prouver un rôle coopératif de la NC et du domaine PTAP dans le recrutement de Tsg101 (Watanabe et al.,

2020). La délétion des sites de liaison des ubiquitines de Tsg101 dans le domaine *Ubiquitin E2 variant* (UEV) diminue son interaction avec p6. Dans ce sens, des analyses de fusion dans lesquelles le domaine p6 est couplé avec une ubiquitine montrent que l'affinité de Tsg101 est renforcée. De plus, l'ubiquitination du domaine p6 permet le recrutement de Tsg101 (Bussienne et al., 2021; Watanabe et al., 2020). Le rôle de l'ubiquitination de p6 dans l'interaction avec Tsg101 est discuté dans ma revue (**annexe 1**).

**ALIX** est une protéine de 96 kDa composée de 3 domaines (Fujii et al., 2009; Göttlinger et al., 1991). Le premier domaine, Bro1 contient un site de liaison aux facteurs *charged multivesicular body protein 4* (CHMP4) appartenant au groupe ESCRT-III. Le domaine central, V est le domaine de liaison aux motifs YPXnL de p6. Finalement, le dernier domaine une région est riche en proline, (PRR). La protéine ALIX est importante pour la biogénèse des corps multivésiculaires et joue aussi un rôle pour les voies de l'apoptose, d'endocytose et de cytokinèse. La délétion du domaine tardif de fixation à ALIX dans p6 n'a pas d'effet majeur sur la production et le bourgeonnement des particules virales (Göttlinger et al., 1991). Néanmoins, ces mutations induisent une réduction de l'infectivité des particules libérées (Martin-Serrano et al., 2003).

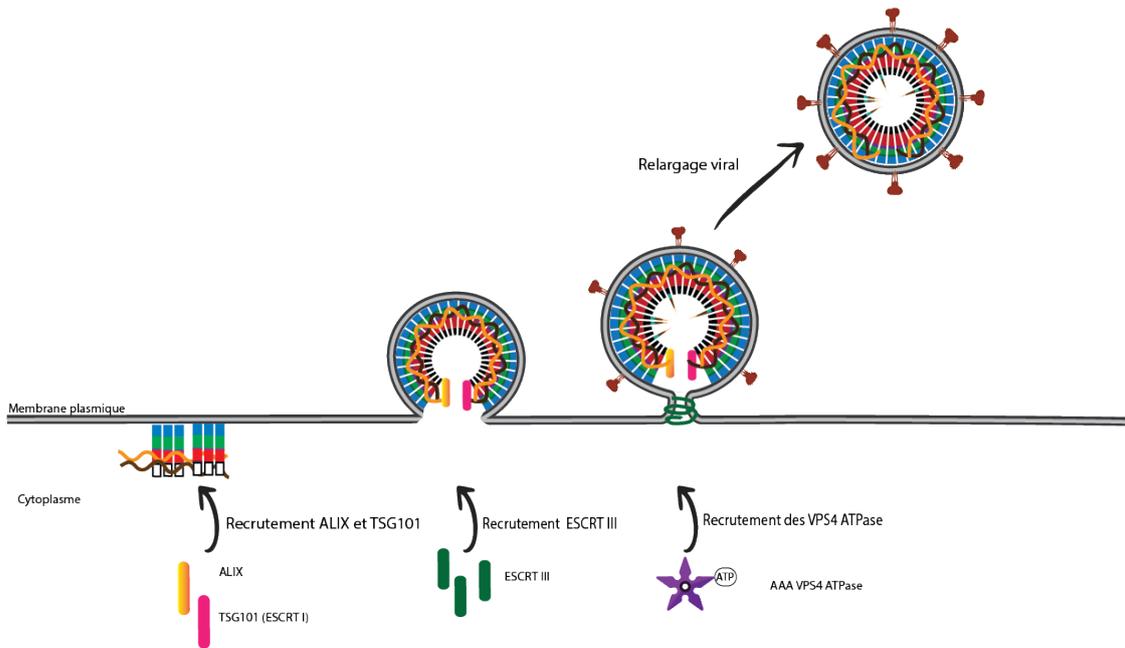


**Figure 20. Représentation du domaine C-terminal de Gag, de sa séquence et de ses partenaires**

**TSG101** est représentée en rose, elle interagit avec le late domaine « PTAP » représenté de la même couleur dans la séquence en acides aminés de p6. Le domaine UEV est le domaine d'interaction avec le late domaine (PDB : 2P22). **ALIX** est représenté en jaune, marron et bleu (PDB : 2OEV). Chaque couleur correspond à un domaine. Le domaine V, représenté en marron est le domaine d'interactions avec le late domaine de p6, YPXnL représenté de la même couleur. Les domaines Bro et PRR sont représentés respectivement en bleu et en marron. La protéine virale **Vpr** est représentée en vert. Elle interagit avec un domaine riche en leucine représenté de la même couleur dans la séquence de p6 (PDB : 1MBL).

Brièvement, les différentes étapes de bourgeonnement et relargage viral commencent par le recrutement des partenaires de p6 à la MP (**figure 21**). Pour cela, le domaine p6 recrute à la fois le domaine V d'ALIX, conduisant à la dimérisation et à l'activation d'ALIX et le domaine UEV de Tsg101. Les deux partenaires sont alors capables de recruter le complexe ESCRT-III, comme les protéines CHMP qui polymérisent en forme de filaments qui favorisent la fermeture du col de la membrane. Ainsi CHMP4 recrutent les protéines CHMP2 et CHMP1, qui, à leur tour

recrutent les ATPases AAA VPS4. Ces dernières forment des anneaux hexamériques qui peuvent être reliés entre eux par des ponts composés du complexe activateur CHMP5/LIP5. VPS4 achèvent la réaction de fission de la membrane. Enfin la consommation d'ATP permet de libérer les ESCRT-III de la membrane et de les recycler dans le cytoplasme (Freed, 2015b; Sundquist and Kräusslich, 2012).



**Figure 21. Modèle de l'étape de bourgeonnement et de relargage de la particule virale du VIH-1**

La protéine Gag recrute la protéine TSG101 des ESCRT-I (représentée en rose) et ALIX (représentée en jaune). Ces deux protéines interagissent avec le domaine p6 de Gag et sont capables de recruter d'autres ESCRT-III (représentées en vert). La polymérisation d'ESCRT-III, couplée à l'activité de VPS4 ATPase (représentée en violet) permet la scission de la MP et le relargage viral (inspiré par Freed, 2015b).

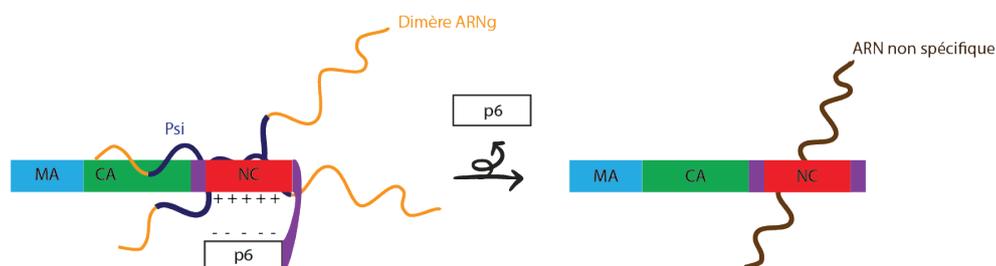


# Objectifs



L'encapsidation de l'ARNg est une étape clé pour la production de virus infectieux. Elle commence par le recrutement de cet ARN par la protéine Gag. Comme décrit précédemment, à ce jour de nombreuses parts d'ombre subsistent sur les mécanismes de reconnaissances spécifiques de l'ARNg. Récemment, des analyses de cartographie (*Mutational interference mapping experiment* ;MIME) de l'ARNg ont permis d'identifier la région minimale d'interaction avec la protéine Gag *in vitro*. Cette région de 111 nucléotides comprend les résidus allant de la position 227 à la position 337 (Smyth et al., 2015). Des études fonctionnelles montrent que l'ARN de 111 nucléotides se lie spécifiquement à Gag et peut être utilisé comme substitut de l'ARNg complet dans les études structurales et fonctionnelles (Smyth et al., 2015). De plus, des analyses de mutagénèses montrent que la tige boucle SL1 jouerait un rôle plus important que SL3 dans le recrutement de la protéine Gag *in vitro* (Bernacchi et al., 2017; El-Wahab et al., 2014; Smyth et al., 2015). Il manque, à ce jour, une structure tridimensionnelle du complexe entre Gag et l'ARNg. Ce modèle permettrait de déterminer les zones de contacts entre la protéine et son ARN et ainsi de mieux comprendre comment l'ARN est recruté spécifiquement dans le cytoplasme. Néanmoins, toutes ces analyses sont réalisées avec la protéine Gag produite en cellules de bactéries *E. Coli*. Or les PTM pourraient être déterminantes pour la fonction de la protéine (Bussienne et al., 2021). Ainsi, le **premier objectif** de ce projet consiste à caractériser les interactions *in vitro* entre l'ARN de 111 nucléotides et le précurseur Gag produit en cellules de mammifères, en combinant des études fonctionnelles avec la résolution de la structure tridimensionnelle de Gag en interaction avec la région Psi.

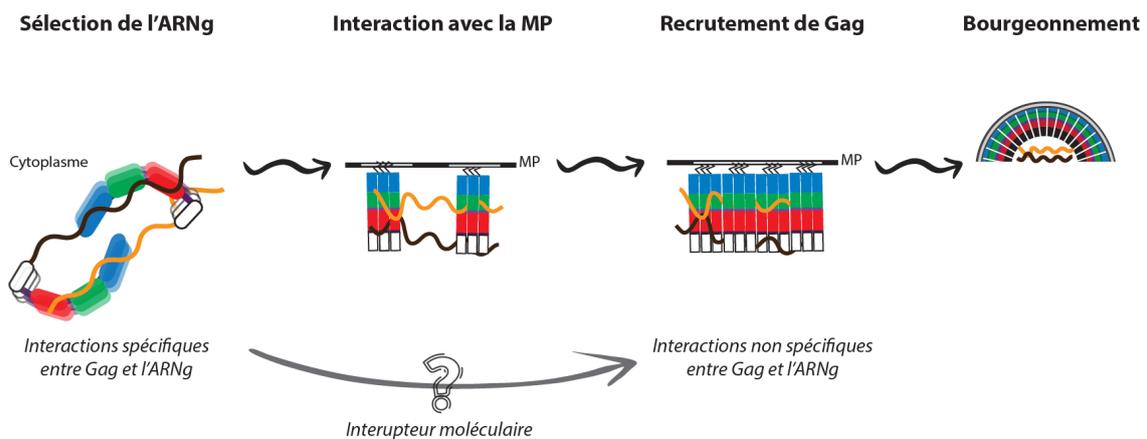
De plus, comme décrit précédemment, le domaine NC de Gag est le domaine majoritaire de fixation à l'ARNg et le déterminant principal de la spécificité de liaison entre les deux partenaires. Or, notre équipe a montré que le domaine C-terminal p6 participe également à la spécificité de liaison de Gag à l'ARN (Dubois et al., 2018a). En effet, la construction de Gag dans laquelle le domaine p6 est délété (Gag $\Delta$ p6) se lie de manière non spécifique à l'ARNg *in vitro*. En d'autres termes, la protéine présente une affinité avec l'ARNg qu'avec l'ensemble des ARN viraux et cellulaires testés (**figure 22**).



**Figure 22. Modèle proposé pour le rôle du domaine p6 dans le recrutement spécifique de l'ARNg**

Le domaine p6 de Gag est chargé négativement et il pourrait interagir avec le domaine NC hautement basique (Wang et al., 2014). Cette interaction favorisait la liaison spécifique de Gag à l'ARNg. En absence du domaine p6, la liaison de Gag $\Delta$ p6 à l'ARN serait à l'inverse non-spécifique. Adaptée de Dubois et al., 2018a.

Cependant, si l'étape de sélection de l'ARNg repose sur des interactions fortement spécifiques entre Gag et l'ARNg, l'étape d'encapsidation à la MP est régulée par des interactions Gag/ARNg non-spécifiques. Il existerait ainsi un interrupteur moléculaire qui régule le passage du mode de fixation spécifique de Gag (stœchiométrie faible), à un mode non-spécifique (stœchiométrie élevée) au moment de l'assemblage (**figure 23**). Le deuxième objectif de ce travail est d'étudier le rôle des partenaires cellulaires et viraux du domaine p6 de Gag (ALIX, Tsg101 et Vpr ; **deuxième objectif**) qui pourraient influencer la spécificité des interactions entre Gag et l'ARNg aux sites d'assemblage.



**Figure 23. Représentation des étapes clés de l'encapsidation de l'ARNg**

Dans le cytoplasme des oligomères de Gag se lient très spécifiquement à l'ARNg dimérique. Ces complexes ribonucléiques impliquent quelques copies de Gag qui sont ensuite redirigés à la MP. L'enchâssement des domaines de MA dans la MP produisent un changement conformationnel de Gag. De ce fait, le domaine MA serait en interaction avec la couche lipidique et le domaine NC serait en association avec l'ARNg. La protéine Gag multimériser jusqu'à atteindre environ 2500 précurseurs permettant le bourgeoisement autour du dimère d'ARNg (adapté de (Rein et al., 2011))

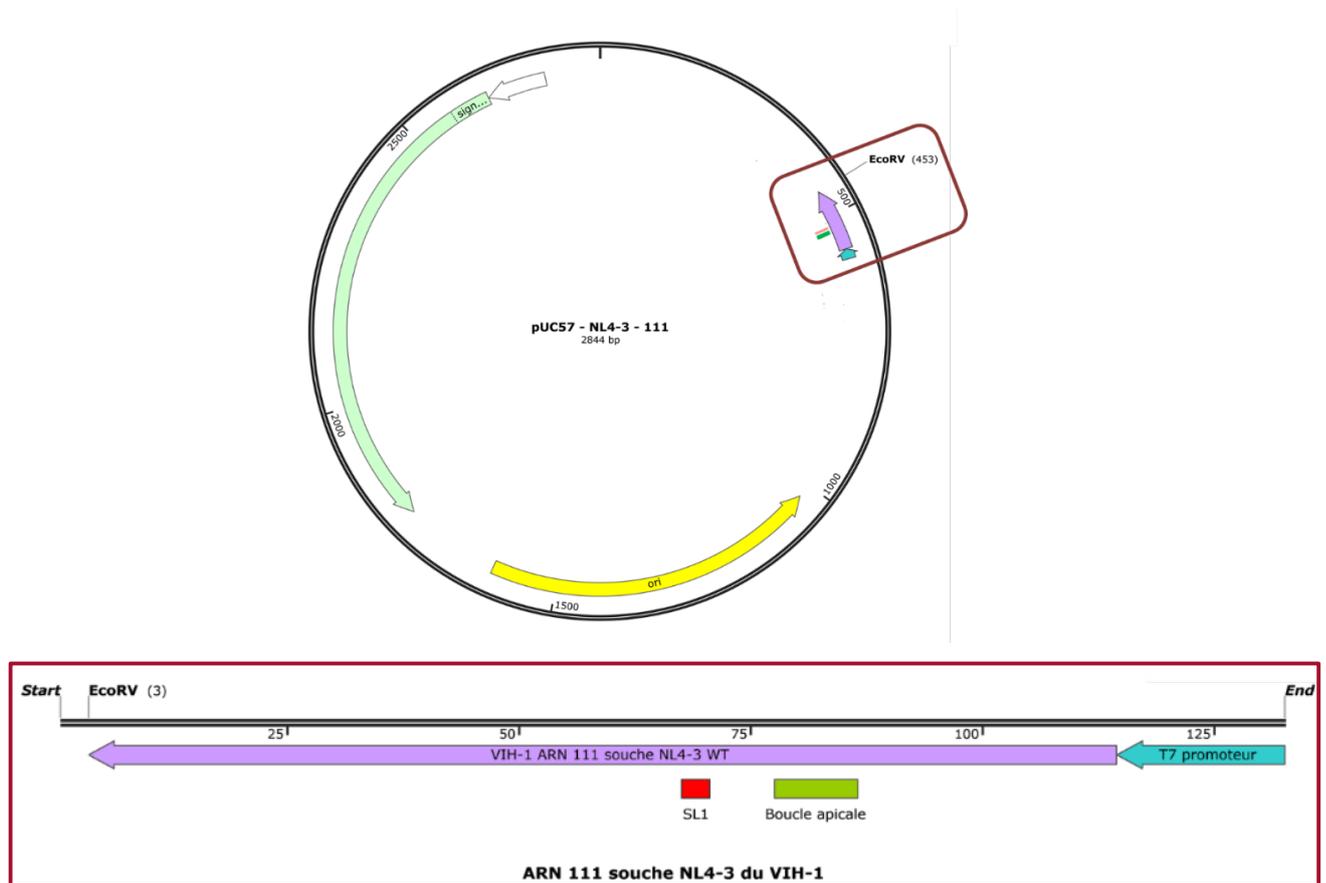
Brièvement, le but de ce travail de thèse est d'apporter une meilleure compréhension des mécanismes moléculaires qui permettent l'encapsidation spécifique de l'ARNg du VIH-1 dans les particules virales. Pour cela des études structurales et fonctionnelles sont combinées afin de (i) **Caractériser, *in vitro*, les interactions entre la protéine Gag et la région Psi de l'ARNg** (ii) **ainsi qu'entre Gag et les partenaires cellulaires (ALIX et TSG101) et viral (Vpr) du domaine p6.**

## **Matériels et Méthodes**



## I. Séquences et cartes des plasmides codants pour les fragments d'ARNg du VIH-1

Les plasmides MaxiPuc18 utilisés pour produire les fragments d'ARNg de la souche de laboratoire NL4-3 codant pour la région 5'UTR de la position 227 à la position 337 (ARN111<sub>WT</sub>) sont produits et décrits précédemment par le laboratoire (Smyth et al., 2015) (**figure 24**).



**Figure 24. Représentation du plasmide MAXIPuc18 utilisé pour produire les fragments d'ARNg**

Représentation du plasmide avec le site de résistance (ampicilline) représenté une flèche verte sous contrôle d'un promoteur représenté par une flèche blanche. L'origine de réplication du plasmide est représentée par une flèche jaune. La séquences codant pour l'ARN de la souche NL4-3 est représentée en violet sous contrôle de promoteur t7, représenté en bleu. Les positions correspondant aux séquences SL1 et à la boucle apicale qui est le site de dimérisation sont respectivement représentées en rouge et en vert. Une représentation linéaire de la séquence de l'ARN111<sub>WT</sub> est représentée dans l'encadré rouge.

Au cours de mes analyses, j'utilise également un fragment d'ARN de la même taille, muté au niveau du site d'initiation de la dimérisation (DIS). Ce site correspond à une séquence palindromique de 6 nucléotides dans la boucle apicale de SL1. La construction correspondante est appelée ARN111<sub>AAA</sub> car les résidus CGC sont mutés en AAA et elle ne peut plus dimériser. De plus, l'ARNSL1<sub>35</sub>, qui correspond à la tige boucle SL1 et qui contient le site minimal d'interaction avec Gag (Bernacchi et al., 2017) est synthétisé chimiquement par la boîte *Integrated DNA Technologies* (IDT). Les informations et les séquences relatives à ces ARNs sont répertoriés dans les **tableaux 1 et 2**.

Nom	Souche	Taille en nucléotides	Poids moléculaire (Da)
ARN111 <sub>WT</sub>	NL4-3	111	36630
	ATCTCTCGACGCAGGACTCGGCTTGCTG <b>AAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGC</b> CAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAG <b>GATATC</b>		
ARN111 <sub>AAA</sub>	NL4-3	111	36630
	ATCTCTCGACGCAGGACTCGGCTTGCTG <b>AAAAACGCA</b> CGGCAAG <b>AGG</b> CGAGGGGCGGCGACTGGTGAGTACGC CAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAG <b>GATATC</b>		

**Tableau 1. Les séquences des plasmides correspondant à la région 5' UTR de l'ARNg de la souche NL4-3 produits au laboratoire**

La boucle apicale de SL1 qui correspond au site de dimérisation de l'ARNg (Paillart et al., 1996) est en vert. La mutation de cette boucle est soulignée et en gras dans la construction ARN111<sub>AAA</sub>. La boucle interne AGG qui correspond au site majeur d'interaction de Gag est représentée en rouge (Bernacchi et al., 2017) et le site de coupure EcoRV utilisé pour linéariser les plasmides est représenté en gras.

Nom	Souche	Taille en nucléotides	Poids moléculaire (Da)
ARNSL1 <sub>35</sub>	NL4-3	35	11550
	CTCGGCTTGCTG <b>AAGCGCGCACGGCAAGAGG</b> CGAG		

**Tableau 2. La séquence de l'ARN correspondant à la tige-boucle SL135, commandé chez IDT**

La boucle apicale qui est site de dimérisation de l'ARNg est représenté en vert et la boucle interne riche en purines, en rouge.

## II. Production des fragments d'ARNg du VIH-1 *in vitro*

À partir d'un plasmide codant la région 5'UTR du VIH-1 les ARN de 111 nucléotides utilisés dans cette étude sont obtenus par transcription *in vitro* en utilisant une ARN polymérase du bactériophage T7, comme décrit précédemment (Dey et al., 2022; Paillart et al., 2002).

### II.1 Amplification des plasmides

Environ 25 µL de bactéries compétentes *E. coli* NEB® 5-*alpha* compétentes, recA1, endA1, lacZΔM15 (Biolabs, ref. C298714) sont mélangées avec 5 ng de plasmides, puis incubées 10 minutes sur de la glace. Les bactéries sont ensuite transformées par un choc thermique à 42°C, de 45 secondes, avec les plasmides pNL4-3 codants pour la région 5' UTR de l'ARNg, l'ARN111<sub>WT</sub> ou l'ARN111<sub>AAA</sub>. Les bactéries sont incubées 2 minutes sur de la glace. Un volume de 500 µL de milieu de culture (*lysogeny broth* ;LB) liquide est ajouté aux bactéries pendant 1 heure à 37°C. Elles sont finalement étalées sur gélose LB-agar supplémentée avec de l'ampicilline (1/1000 v/v). Les colonies capables de pousser sur le milieu sélectif sont ensuite sélectionnées et amplifiées à l'aide d'une pré-culture liquide de 3 mL LB en présence d'ampicilline (1/1000 v/v) pendant une nuit, à 37°C. Cette pré-culture est ensuiteensemencée dans 200 mL de LB supplémenté avec de l'ampicilline (1/1000 v/v) et incubée une nuit à 37°C. L'ADN amplifié est extrait

et purifié selon le protocole du kit Nucleobond Xtra MAXI (Macherey-Nagel, ref. 740416.50). Les séquences sont systématiquement contrôlées en utilisant le service de séquençage d'Eurofin Genomics (<https://eurofinsgenomics.eu>) après avoir mesuré la concentration des ADN purifiés grâce à la mesure d'absorption à 260 nm déterminée par spectrophotométrie en UV-VIS l'aide du NanoDrop (Mettler Toledo).

## **II.2 Linéarisation des plasmides amplifiés**

Les plasmides contenant la matrice ADN pour la transcription des ARN111<sub>WT</sub> et ARN111<sub>AAA</sub> amplifiés, sont ensuite linéarisés à 37°C par digestion enzymatique en utilisant 2 unités d'*EcoRV Hf* (New England Biolabs, ref. R3195SVIAL) par µg de plasmide. Le mélange réactionnel est incubé 1 h à 37 °C. Le plasmide digéré est purifié par extraction au Roti®-Phenol, Chloroform, isoamyl alcohol à pH 8 (v/v, 25 : 24 : 1 ; ThermoFisher, ref. 15593031). Il est ensuite précipité par ajout d'acétate d'ammonium (3 M), pH 5.2 à 4°C (1/10 v/v) en présence de 3 volumes d'éthanol 100%. La précipitation est réalisée sur la nuit à - 80°C. Après une centrifugation de 30 minutes à 20000 g à 4 °C, le culot est lavé avec de l'éthanol 70% froid à l'aide d'une nouvelle étape de centrifugation de 30 minutes à 20000 g à 4 °C. Le culot est alors séché dans une hotte et à température ambiante avant d'être repris dans de l'eau ultrapure. La concentration des ADN est déterminée par spectrophotométrie en mesurant l'absorption à 260 nm comme précédemment indiqué. La digestion du plasmide est finalement contrôlée sur gel d'agarose 1%, Tris Borate EDTA (TBE) 1 x. Le gel est révélé en présence de bromure d'éthidium (BEt) en utilisant le système d'imagerie intégré NuGenius (Syngene). La matrice non digérée est utilisée comme contrôle de digestion.

## **II.3 Transcription des plasmides linéarisés**

Les plasmides ainsi linéarisés sont utilisés comme matrice pour réaliser une transcription *in vitro* avec l'ARN polymérase du bactériophage T7 purifiée par l'équipe du dr. Wexeilbaumer (Dey et al., 2022). Pour cela, 150 µg de plasmides linéarisés sont mélangés à 5 mL de milieu réactionnel pendant 3 h à 37°C. L'ARN nouvellement transcrit est d'abord purifié par extraction au Roti®-Phenol, Chloroform, isoamyl alcohol à pH 5.2 (v/v, 25 : 24 : 1 ; thermoFisher, ref. J62336.AE), puis est précipité par ajout d'1/10<sup>ème</sup> du volume, d'acétate d'ammonium (3 M) pH 5,2 à 4°C en présence de 3 volumes d'éthanol 100%. La précipitation est réalisée sur la nuit à - 80°C. Après une centrifugation de 30 minutes à 20000 g à 4 °C, le culot est lavé avec de l'éthanol 70% froid. Le surnageant est jeté après une nouvelle centrifugation de 30 minutes à 20000 g à 4 °C. Le culot est alors séché dans une hotte à température ambiante avant d'être repris soit dans du bleu de charge (composé d'EDTA (8 mM), d'urée (20 mM), de Tris HCl pH 7,5 (5mM) et de bleu de bromophénol (0.5% w/v)) pour être purifié à l'aide d'un gel SDS-urée

dénaturant 8% bis-acrylamide (ARN111<sub>WT</sub>) ; soit, dans de l'eau ultrapure (ARN111<sub>AAA</sub>) pour être purifié par colonne de séparation par taille TSKgel® G4000 SW (Tosoh bioscience, ref. 0006729).

#### II.4 Purification des fragments d'ARN transcrits

Ainsi, deux méthodes de purification de l'ARN sont utilisées au cours de ce travail de thèse.

**D'une part, la purification par Électroélution**, les ARN111<sub>WT</sub> nouvellement transcrits, sont purifiés par migration sur un gel dénaturant 8%, 1 : 30 bis-acrylamide, urée (8 M) et TBE 1 X (35 cm x 20 cm). Pour cela, la totalité des ARN repris dans le bleu de charge est dénaturée 5 minutes à 95°C, puis déposée dans un puit. Après 6 h de migration à 50 Watts dans du TBE 1 X, la bande correspondant à l'ARN est découpée sous UV, puis est purifiée par électro-élution en utilisant le système Elutrap® (Whatman, ref. WHA10447700). Brièvement, la bande de gel correspondant à l'ARN est découpée en morceaux de petite taille, puis déposée à l'intérieur de la chambre interne, ou d'élution ; qui est fermée par des membranes nommées BT1 et BT2. La chambre est alors soumise à un courant électrique de 150 volts pendant 7 heures. Toutes les heures, les ARN extraits du gel sont récupérés de la chambre de prélèvement dans des microtubes Eppendorfs et leur concentration est déterminée par spectrophotométrie en UV.

**D'autre part, purification par chromatographie de séparation en fonction de la taille**, les ARN111<sub>AAA</sub> transcrits, sont purifiés par chromatographie à exclusion de taille TSKgel® G4000 SW (Tosoh bioscience, ref. 0006729) selon le protocole précédemment établi au laboratoire. Brièvement, une colonne TSKgel® G4000SW (Tosoh Bioscience) est équilibrée avec une solution tamponnée composée d'acétate de sodium à pH 5,0 (0,2 M) et d'1% (v/v) d'éthanol, filtrées sur des membranes de celluloses de 0,22 µm (GE Healthcare). Un débit de 1 mL/minute est appliqué à l'aide d'un système ÄKTA purifier (GE Healthcare). Les échantillons transcrits *in vitro* sont injectés sur la colonne. Les ARN sont purifiés en fonction de leur taille et des fractions de 1 mL sont collectées automatiquement. Celles contenant les acides nucléiques d'intérêts sont identifiées au regard des pics d'absorption à 254 nm sur le chromatogramme puis elles sont rassemblées.

Les ARNs purifiés en utilisant l'une ou l'autre méthode sont extraits et concentrés par extraction au Roti®-Phenol, Chloroform, isoamyl alcohol à pH 5,2 (v/v, 25 : 24 : 1 ; Thermo Fisher, ref. J62336.AE), puis précipité par ajout d'1/10<sup>ème</sup> de volume d'acétate d'ammonium (3 M) pH 5,2 à 4°C en présence de 3 volumes d'éthanol 100%. La précipitation est réalisée sur la nuit à - 80°C. Après une centrifugation de 30 minutes à 20000 g à 4 °C, le culot est lavé avec de l'éthanol 70% froid grâce à une nouvelle centrifugation de 30 minutes à 20000 g à 4 °C. Les culots

sont alors séchés à l'air puis repris dans de l'eau Millipore. Ils sont ensuite systématiquement contrôlés sur un gel dénaturant 8% acrylamide, 1 : 30 bis-acrylamide, 8 M urée et TBE 1 X (8 cm). Le gel est révélé à l'aide d'un bain de BEt comme précédemment décrit. La concentration des ARN est enfin calculée à partir de la mesure d'absorption à 260 nm.

### III. Caractérisation de l'ARN du VIH-1 *in vitro*

#### III.1 Repliement de l'ARN viral et contrôle de la dimérisation

Les ARN111<sub>wt</sub> et ARNSL1<sub>35</sub> sont dénaturés 2 minutes à 95°C, puis refroidis 2 minutes sur glace. La solution tamponnée d'intérêt (tampon A ; **tableau 6**, section Matériels et Méthodes, §IV.2.3) est alors ajoutée à l'ARN dénaturé qui est ensuite incubé à 37°C sur un bloc chauffant pendant 30 minutes.

La proportion des dimères et monomères d'ARN est contrôlée par gel natif d'agarose. Une fois les ARN repliés, 20 pmol d'ARN sont supplémentés avec du bleu glycérol puis chargés dans un gel d'agarose 0,8%, Tris acide borique et MgCl<sub>2</sub> (TBM ; 0.5 x), supplémenté avec du BEt (1 : 25000 v/v). La migration est réalisée à 4°C, pendant 3h à 150 volts. Le contrôle négatif de dimérisation correspond à un ARN muté dans la boucle de SL1 incapable de dimériser, l'ARN111<sub>AAA</sub>. Le gel est révélé comme précédemment décrit. La proportion des dimères/monomères d'ARN sont également analysée par AUC (§IX.4).

#### III.2 Marquage des ARN viraux

Certaines méthodes biophysiques telles que l'anisotropie de fluorescence et la FIDAbio, utilisées lors de ce travail pour caractériser l'ARN et la formation de complexes avec les protéines Gag, nécessitent le marquage préalable des ARN par une sonde fluorescente (ou éventuellement de la protéine comme indiqué dans la §VII.4).

Ainsi, les ARN de 111 nucléotides (ARN111<sub>wt</sub> et ARN111<sub>AAA</sub>) et le ARN de 35 nucléotides (ARNSL1<sub>35</sub>) sont marqués en utilisant le protocole du kit fluorescéine 5-thiosemicarbazide (FTSC ; Invitrogen, ref. #F121). Brièvement, les sucres présents dans la région 3' du ribose des ARN sont oxydés par du périodate de sodium, formant des aldéhydes réactifs. Le sucre oxydé est alors conjugué au marqueur fluorescent FTSC. Les ARN marqués, sont ensuite purifiés en utilisant le protocole des colonnes micro Bio-spin p6® (Biorad ; ref. 7326227) préalablement équilibrées avec l'une des solutions tamponnées de formation des complexes (**tableau 6**, section Matériels et Méthodes, §IV.2.3). L'efficacité de marquage est évaluée en utilisant les formules suivantes :

Correction de la concentration molaire de l'ARN en fonction de l'absorbance de la sonde,

$$C^{\text{mol}} = \frac{\text{Abs } 260\text{nm} - (\text{Abs } 294\text{nm} * 0.11)}{\epsilon \text{ ARN111wt}} \quad (1)$$

Calcul du nombre de fluorophores par ARN,

$$\frac{Abs\ 294\ nm}{\epsilon\ Dye\ (71000) \times C\ molaire\ ARN111wt} \quad (2)$$

La concentration des ARN est calculée à partir de la mesure d'absorption à 260 nm. L'efficacité de marquage est aussi contrôlée en réalisant un gel dénaturant 8% acrylamide, 1 : 30 bis-acrylamide, 8 M urée et TBE 1 X (8 cm) qui est révélé d'une part, au Typhoon (GE Healthcare) en utilisant le filtre Cy2 et d'autre part, à l'aide d'un bain de BET comme précédemment décrit.

#### IV. Séquence en acides aminés des protéines recombinantes Gag et des partenaires de Gag

##### IV.1 Protéines Gag produites en système bactérien

Les protéines Gag de la souche NL4-3, présentent une étiquette composée de 6 résidus histidines (6His) accrochée à l'extrémité C-terminale et elles sont produites et purifiées par le laboratoire du docteur Johnson Mak (Université de Griffith, Queensland, Australie), selon le protocole décrit dans l'article suivant (McKinstry et al., 2014). Le poids moléculaire (PM), l'étiquette et la séquence de la protéine Gag-6His, sont répertoriés dans le **tableau 3**. Le PM et le coefficient d'extinction molaire à 280 nm sont calculés à partir de la séquence en acides aminés des protéines par le site internet Protparam (<https://web.expasy.org/protparam/>).

Nom	PM (Da)	Coefficient d'extinction molaire à 280 nm (M <sup>-1</sup> .cm <sup>-1</sup> )	Étiquette
Gag-6His	56884	63535	6His
<p>MGARASVLSGGELDKWEKIRLRPGGKKQYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQQTGSEELRSLYNTIAVLVYCVHQRID            VKDTKEALDKIEEEQNKSKKKAQAAAADTGNNSQVSQNYPIVQNLQGMVHQAI SPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQD            LNTMLNTVGGHQAAMQMLKETINEEAAEWDRHLHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKRWIILGLNKI            VRMYSPTSILDIRQGPKPEFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQANANPDCKTILKALGPGATLEEMMTACQGVGGPGHKARVL            AEAMSQVTNPATIMIQKGNFRNQRKTVKFCNCGKEGHIAKNCRAPRKKGCWKCCKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQS            RPEPTAPPEESFRFGEETTPSQKQEPIDKELYPLASLRSLFGSDPSSQLEHHHHHHH</p>			

**Tableau 3. Informations sur les protéines Gag produites en système bactérien**

La séquence reprend le code couleur présenté figure 5 (dans l'introduction §III.2). Avec les domaines de MA représenté en bleu, de CA représenté en vert, de NC représenté en rouge et encadré par les deux espaceurs sp1 et sp2 représentés en violet. Enfin le domaine p6 est représenté en noir. L'étiquette composée de 6 résidus histidines est surligné en bleu.

#### IV.2 Protéines Vpr produites par synthèse chimique

Les protéines Vpr de groupe M de sous-type B (souche 89.6) sont produites par synthèse chimique et purifiées par le laboratoire du docteur Serge Bouaziz (Université Paris Descartes), selon le protocole décrit dans l'article de Cornille et al., 1999. Les informations et la séquence sont répertoriées dans le **tableau 4**. Le PM et le coefficient d'extinction molaire à 280 nm sont calculés à partir de la séquence en acides aminés des protéines comme décrit précédemment.

Nom	PM (Da)	Coefficient d'extinction molaire à 280 nm ( $M^{-1}.cm^{-1}$ )	Étiquette
Vpr <sub>(synthétique)</sub>	11395	20970	Absente
MEQAPEDQGPQREPYNDWTLELLEELKNEAVRHFPRIWLHSLGQHIYETYGDTWTGVEALIRILQQLLFHFHFRIGCRHSRIGIIQQRTRNGASKS			

**Tableau 4. Informations sur les protéines Vpr produites par synthèse chimique**

#### IV.3 Séquences des protéines produites en cellules de mammifères

Les séquences des gènes utilisés pour produire les protéines en cellules embryonnaires de reins de hamsters (*baby hamster kidney 21*, BHK21) sont synthétisés par Twist Bioscience (<https://www.twistbioscience.com/>) en utilisant les séquences du VIH-1 de la souche NL4-3 pour les protéines virales (Gag et Vpr) et *homo sapiens* pour les protéines cellulaires (ALIX et UEV). Elles sont ensuite clonées dans le vecteur de transfert pVote2 (Hebben et al., 2007). Toutes les protéines produites en BHK21 possèdent soit une étiquette composée de 6 résidus Histidine (6His), soit une étiquette *tween Strep-tag*<sup>®</sup> composé des résidus suivants WSHPPQFEK-GGGSGGGSGGS-SAWSHPPQFEK (R) dans leur région N-terminale. Les informations, les étiquettes et les séquences sont répertoriées dans le **tableau 5**. Le PM et le coefficient d'extinction molaire à 280 nm sont calculés comme décrit pour la protéine Gag produite en cellules bactériennes.

Nom	PM (Da)	Coefficient d'extinction molaire à 280 nm (M <sup>-1</sup> .cm <sup>-1</sup> )	Étiquette
6His-Gag	59 252	65025	6-His
<p>MGSSHHHHHHTPITSLYKKAGF<b>LEVLFQGP</b>MGARASVLSGGELDKWEKIRLRPGGKKQYKLKHIVWASRELERFAVNPGLLETSEGCRQI  LGQLQPSLQTGSEELRSLYNTIAVLYCVHQRIDVKDTKEALDKIEEQNKSKKKAQQAADTGNNSQVSQNYPIVQNLQGGQMVHQAI<b>SPRTL</b>  NAWVKVVEEKAFSPVPIPMFSALSEGATPQDLNLTMLNTVGGHQAAMQMLKETINEEAAEWDRLLHPVHAGPIAPGQMREPRGSDIAGTT  STLQEQIGWMTNHPPIPVGEIYKRWIILGNKIVRMYSPSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQANANPDCKTI  LKALGPGATLEEMMTACQGVGGPGHKARVLA<b>EAMSQVTNPATIMIQKGNFRNRKTKVCFNCGKEGHIAKNCRAPRKKGCWKCGKEGH</b>  <b>QMKDCTERQANFLGKIWPSHKRPGNF</b><b>LQSRPEPTAPPEESFRFGEETTPSQKQEPIDKELYPLASLRSLFGSDPSSQ</b></p>			
R-ALIX	99776	71795	R
<p><b>WSHPQFEKGGGSGGGSGGSAWSHPQFEK</b><b>LEVLFQGP</b>MATFISVQLKKTSEVDLAKPLVKFIQQTYPSPGEEQAQYCRAAEELSKLRRAAV  GRPLDKHEGALETLRRYYDQICSIEPKFPFSENQICLFTTWKDAFDKGSFLGGSVKLALASLGYEKSCVLFNCAALASQIAAEQNLNDNDEGLKI  AAKHQFASGAFHiketVLSALSREPTVDISPDVTGTLISLIMLAQAQEVFLKATRDKMKDAIIAKLANQAADYFGDAFKQCQYKDTLPKE  VFPVLAAKHCIMQANAAYHQSIKAKQKQKFGEEIARLQHAELIKTVASRYDEYVNVKDFSDKINRALAAAKDNDFIYHDRVDPDLKDLDP  GKATLVKSTPVNVPISQKFTDLFEKMVPVSVQSLAAYNQRKADLVNRSIAQMREATTLANGVLAASLNLPAIADVSGDTPVQSILTKSRV  IEGGIQTVDQLIKELPELLQRNREILDESLRLDEEEATDNDLRAKFKERWQRTPSNELYKPLRAEGTNFRTVLDKAVQADGQVKECYQSH  RDTIVLLCKPEPELNAIIPSANPAKTMQGEVNVNLSLLSNLDEVKEREGLNDLKSVMFDMTSKFLTALAQDGVINEEALSVELDRVY  GGLTTKVQESLKKQEGLLKNIQVSHQEFKMKQSNNEANLREEVLKNLATAYDNFVELVANLKEGTFYNELTEILVRFQNKCSDIVFARKT  ERDELLKDLQSIAREPSAPSIPTPAYQSSPAGGHAPTPPTAPRTMPPTKQPPARPPPPVLPANRAPSATAPSPVGAGTAAPAPSQTP  GSAPPQAQGGPPYTPGYPGYCCMPMPMGYNPYAYQYNMPPYVYHQSPGQAPYGPQPQPSYFPQPPQSSYYPQQ</p>			
R-UEV	20367	37025	R
<p><b>WSHPQFEKGGGSGGGSGGSAWSHPQFEK</b><b>LEVLFQGP</b>MAVSESQLKMMVSKYKYRDLTVRETVNVITLYKDLKPVLDVSVFNDGSSRELM  NLTGTIPVYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFSRP</p>			
R-Vpr	15132	31970	R
<p><b>WSHPQFEKGGGSGGGSGGSAWSHPQFEK</b><b>LEVLFQGP</b>MEQAPEDQGPQREPYNEWTLELLEELKSEAVRHFPRHWLNLGQHIYETYGD  TWAGVEAIIRILQQLLFIHFRIGCRHSRIGVTRRRARNGASRS</p>			

**Tableau 5. Informations sur les protéines 6His-Gag, R-Alix, R-UEV et R-Vpr produites en cellules de mammifères**

Dans ces constructions les étiquettes 6His et R sont surlignées en bleu dans la séquence des protéines. Elles sont suivies par la séquence consensus reconnue par les protéases des virus Human Rhinovirus 3C (P3C) ou Tobacco Etch Virus (TEV), représentée en jaune. La séquence de la protéine Gag reprend le code couleur présenté figure 5 (dans l'introduction). Avec le domaine de MA représenté en bleu, la CA représentée en vert, la NC représentée en rouge, encadrée par les deux espaceurs sp1 et sp2 représenté en violet. Finalement, le domaine p6 est représenté en noir. Les résidus reliants (linker) l'étiquette 6His à la séquence de la protéine sont soulignés.

#### V. Production des vecteurs viraux recombinants pour l'expression des protéines en système eucaryote (BHK21).

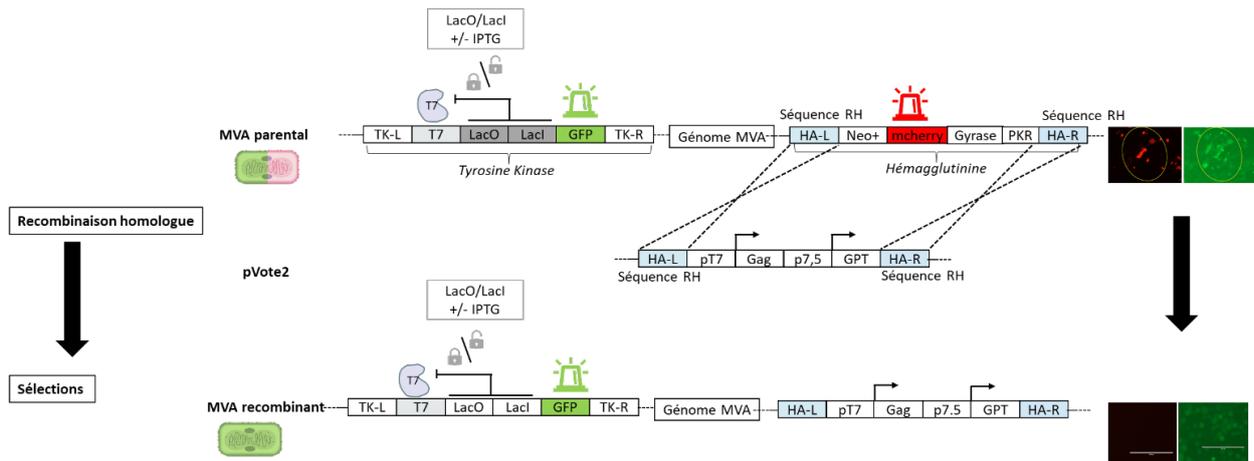
L'utilisation d'un système en cellules de mammifères pour la production de protéines présente l'avantage d'apporter les modifications post-traductionnelles (*post translational modification, PTM*). Ces modifications sont nécessaires au fonctionnement des protéines et favorisent les interactions avec les partenaires (Bussienne et al., 2021) Pour ce faire mon laboratoire d'accueil a mis au point un système d'expression basé sur l'utilisation du virus de la vaccine modifié (souche Ankara ; MVA) comme vecteur d'expression (Hebben et al., 2007) permettant d'obtenir de bons rendements de protéines d'intérêts en infectant des cellules BHK21 en suspension.

### V.1 Culture cellulaire

Pour produire les protéines utilisées dans cette étude, des cellules BHK21 adhérentes et en suspension sont cultivées dans des incubateurs à atmosphère humide, enrichie avec 5% de CO<sub>2</sub> et à température contrôlée à 37°C, sous agitation continue pour les cellules en suspension. Elles sont produites dans un milieu de croissance (*Glasgow minimum essential medium* ; GMEM, de Fischer scientifique, ref. 15188319) supplémenté avec du Bacto tryptone phosphate (BTP ; 1,5 mg/mL), de la gentamicine (200 ng/mL) et du sérum de veau fœtal (*fetal calf serum*, FBS ; 10%).

### V.2 Production et sélection des vecteurs recombinants

Un virus MVA dit « virus parental » est modifié au niveau des gènes codants pour la tyrosine kinase (TK) et l'hémagglutinine (HA ; **figure 25**). Le gène de la tyrosine kinase est modifié pour exprimer le gène qui code pour l'ARN polymérase du bactériophage T7 sous contrôle de l'opéron lactose et le gène rapporteur exprimant la protéine fluorescente verte (*green fluorescent protein*, GFP). Le gène HA est modifié pour exprimer un gène de résistance à la néomycine, le gène rapporteur qui code pour la protéine fluorescence mCherry et les gènes codants pour le système de sélection gyrase/protéine kinase R (PKR). Cette région présente aussi des séquences de recombinaisons homologues compatibles avec celles présentes sur le plasmide pVote2. Ce plasmide code pour la protéine d'intérêt sous contrôle du promoteur T7 et pour la *guanine phosphoribosyl transferase* (GPT) sous contrôle d'un promoteur viral constitutif, p7.5.



### Figure 25. Représentation des étapes de production des virus recombinants

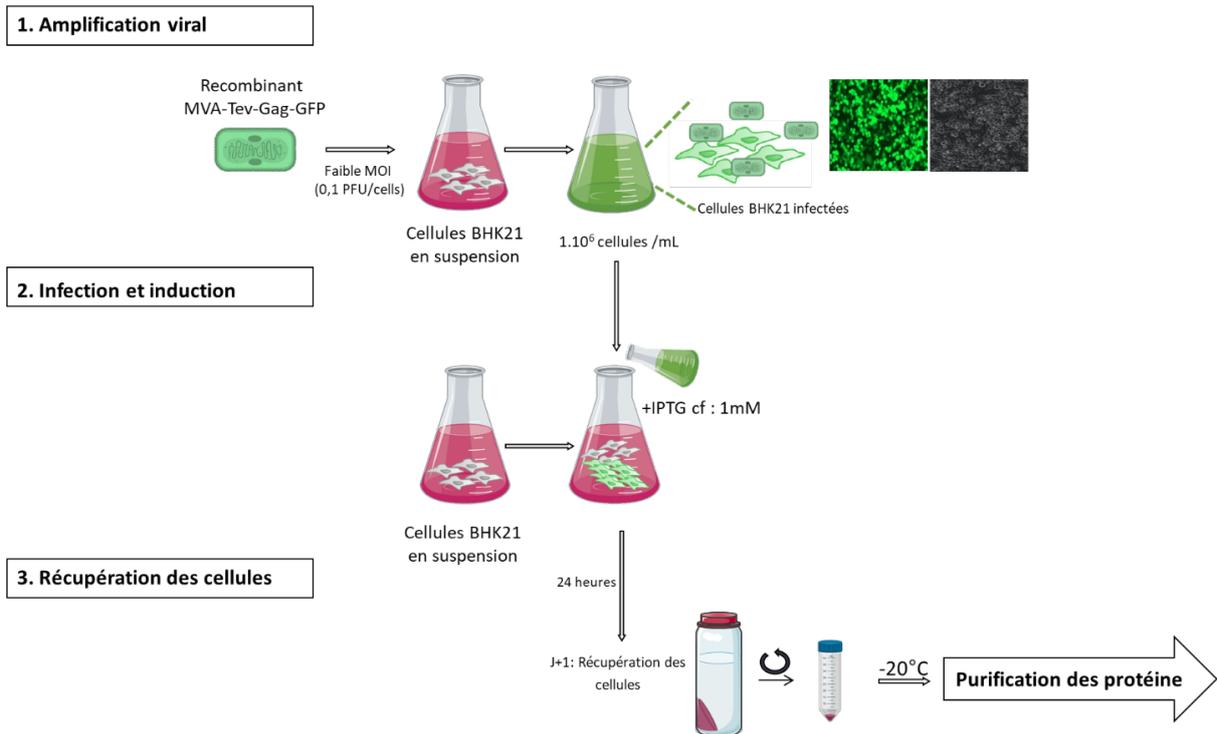
Représentation du virus parental modifié au niveau des gènes codants pour la tyrosine kinase (TK) et l'hémagglutinine (HA). TK exprime les gènes qui codent pour l'ARN polymérase du bactériophage T7 sous contrôle de l'opéron lactose et le gène rapporteur exprimant la protéine fluorescente verte (green fluorescent protein, GFP). HA pour exprime un gène de résistance à la néomycine, le gène rapporteur qui code pour la protéine fluorescence mCherry et les gènes codants pour le système de sélection gyrase/protéine kinase R (PKR). Cette région présente aussi des séquences de recombinaisons homologues compatibles avec celles synthétisées sur le plasmide pVote2. Ce plasmide code pour la protéine d'intérêt sous contrôle du promoteur T7 et pour la guanine phosphoribosyl transférase (GPT) sous contrôle d'un promoteur viral constitutif (p7.5). La sélection des virus recombinants est suivie par microscopie à épifluorescence. L'expression de la RFP diminue alors que celle de la GFP reste stable.

Pour produire le virus recombinant des cellules BHK21 adhérentes sont infectées par le virus parental à 0.1 unités formant des plages de lyse par cellule. Elles sont en parallèle transfectées avec le plasmide pVote2 codant pour la protéine d'intérêt (0,5 µg/µL) en utilisant de la lipofectamine 3000 (Invitrogen, ref. 15282465) comme agent de transfection. Plusieurs cycles de sélections vont permettre de sélectionner les virus recombinants. Un cycle de sélection consiste à supplémenter le milieu de culture des cellules infectées et transfectées avec du FBS (2%), de l'acide mycophénolique (10 mg/mL), de la xanthine (10 mg/mL), de l'hypoxanthine (1,5 mg/mL) et de la coumermycine (50 µg/mL). L'acide mycophénolique est un inhibiteur non compétitif et réversible de l'inosine monophosphate déshydrogénase. Sa présence dans le milieu de culture inhibe la synthèse des nucléotides guanines. Le virus recombinant, grâce à l'expression du gène de la GPT et en présence de xanthine et d'hypoxanthine est capable d'activer une voie de biosynthèse secondaire lui permettant de synthétiser de la guanine et donc de se multiplier. De plus, le virus parental est sensible à la coumermycine qui inhibe la traduction *via* une réaction en cascade qui commence par la dimérisation de la gyrase, induisant la dimérisation de la PKR qui va alors phosphoryler le facteur d'initiation de la traduction eucaryote 2α et inhiber la traduction. Finalement, trois cycles de sélection sont nécessaires pour obtenir une majorité de virus recombinants. L'efficacité de la sélection est caractérisée à l'aide d'un microscope à épifluorescence (EVOS, Thermo-Fischer) en suivant la diminution de l'expression de la mCherry tandis que l'expression de la GFP reste stable.

### V.3 Contrôle de l'expression des nouveaux vecteurs recombinants et production des protéines recombinantes

L'expression des protéines d'intérêts est contrôlée en réalisant une production de protéines à l'aide de cellules BHK21 adhérentes. Pour cela, 30 mL de milieu de culture GMEM, contenant  $1.10^6$  cellules adhérentes, supplémentés avec du BTP (1,5 mg/mL), de la gentamicine (200 ng/mL) et du FBS (10%) sont cultivés et infectés à l'aide des MVA recombinants. L'expression des protéines est induite par ajout d'isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG ; 1 mM) dans le milieu de culture. A l'issue de 24 h d'infection et d'induction, les cellules sont décrochées, rincées avec une solution tamponnées phosphate saline et conservées à  $-20^{\circ}\text{C}$  jusqu'à purification des protéines.

Afin de produire une grande quantité de protéines d'intérêts, de 6 L à 12 L de culture de cellules BHK21 en suspension sont cultivés (**figure 26**). L'infection des cellules est réalisée en ajoutant 40 mL de virus pour 6 L de culture contenant une concentration de  $1.10^6$  de cellules par mL dans un milieu GMEM supplémenté avec du BTP (1,5 mg/mL), de la gentamicine (200 ng/mL) et du FBS (10%). Au terme de 24h d'infection et d'induction, réalisées dans un incubateur à  $37^{\circ}\text{C}$  sous atmosphère contrôlée, les cellules sont récupérées, culottées à l'aide d'une étape de centrifugation, rincées avec une solution tamponnées phosphate saline et conservées à  $-20^{\circ}\text{C}$  jusqu'à purification des protéines.



**Figure 26. Représentation des grandes étapes de production des protéines dans les cellules BHK21**

**1. Amplification virale** : Des cellules BHK21 sont amplifiées dans le milieu GMEM supplémenté, jusqu'à atteindre une concentration de  $1.10^6$  cellules/mL. Elles sont alors infectées avec le virus MVA recombinant avec une faible MOI. Cette étape permet d'amplifier les virus avant d'infecter 6 à 12L de cellules. **2. Infection et induction** : les virus amplifiés sont mélangés avec des cellules amplifiées. L'expression de la protéine d'intérêt est induite par ajout d'IPTG (1 mM). Après 24h les cellules sont récupérées. **3. Récupération des cellules** : les cellules sont alors culottées et lavées avec du PBS avant d'être conservées à  $-20^{\circ}\text{C}$  jusqu'à la purification des protéines.

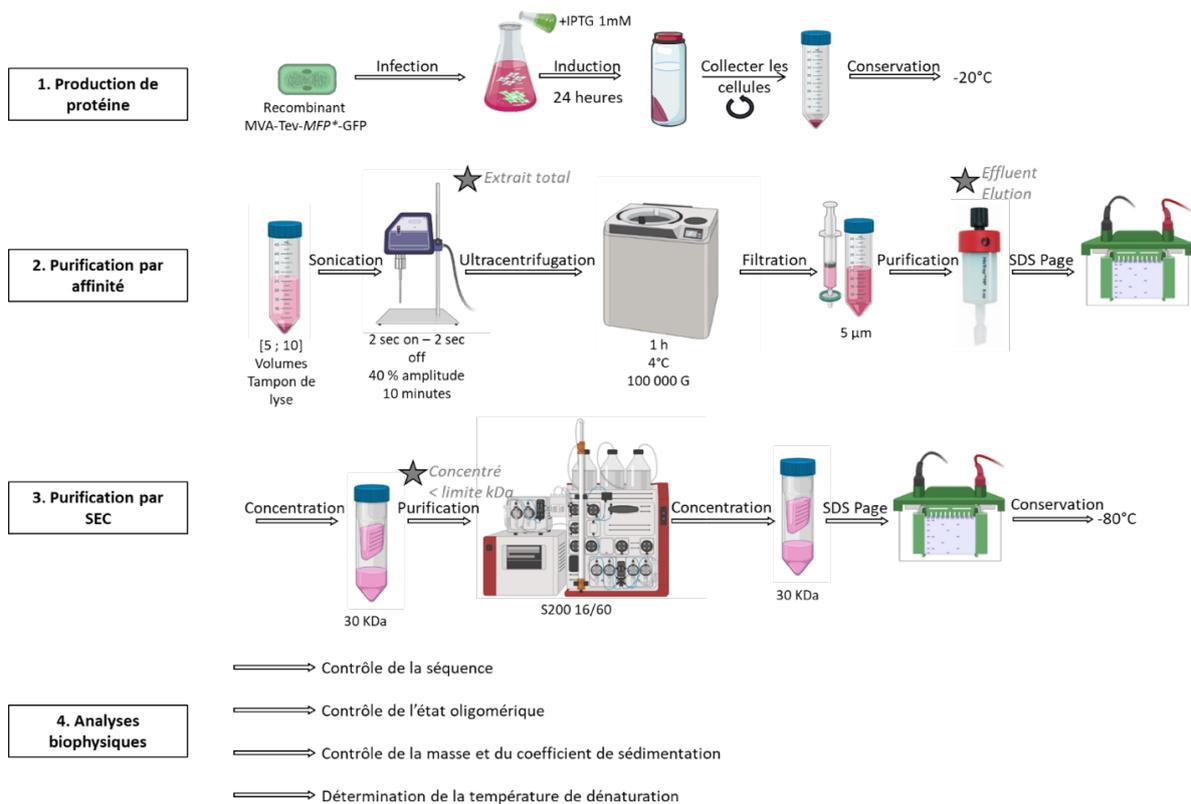
## VI. Purification des protéines recombinantes en cellules de mammifères

### VI.1 Procédures communes à la purification des protéines produites en cellules de mammifères

Les culots cellulaires sont décongelés et repris dans le tampon de lyse afin de casser les cellules par sonication pendant 5 minutes, avec une alternance de 2 secondes à 40% d'amplitude et 2 secondes de pause, par le sonicateur VibraCell (Bioblock Scientific, ref. 432-0133). Le tampon de lyse dépend de la protéine purifiée (**tableau 6**). L'étiquette présente dans la partie N-terminale des protéines recombinantes est utilisée pour la première étape de purification par colonne d'affinité (voir §suivant). Les fractions contenant la protéine d'intérêt sont identifiées en réalisant une électrophorèse sur SDS-PAGE bis-acrylamide de 10 à 15% en fonction du PM des protéines, révélée avec de l'InstantBlue® Coomassie (ABCAM, ref. ISB1L). Les fractions sont rassemblées pour être concentrées à l'aide d'un concentrateur Amicon Ultra-15 avec un seuil de passage des molécules de 10 (R-Vpr) ou 30 kDa (6His-Gag, R-ALIX et R-VPR ; Merck Millipore, ref. UFC903008). Les protéines sont ensuite séparées en fonction de leur taille par chromatographie à exclusion de taille en utilisant une colonne Superdex S200 (Cytiva) dans une solution de conservation (détaillée **tableau 6**). De nouveau, les fractions contenant les protéines sont identifiées en réalisant une électrophorèse sur gel dénaturant SDS bis-acrylamide de 10 à 15% révélée avec de l'InstantBlue® Coomassie (ABCAM, ref. ISB1L) et sont rassemblées pour être concentrées. La concentration des protéines est estimée en mesurant l'absorbance de la protéine purifiée à 280 nm à l'aide du Nanodrop (Mettler Toledo) et en appliquant la loi de *Beer Lambert* :

$$c^{\text{mol}} = \text{Abs} / (\epsilon_{\text{théorique}} \cdot l) \quad (3)$$

Avec, Abs, l'absorbance à 280 nm ;  $\epsilon_{280\text{nm}}$ , le coefficient d'absorption molaire à 280 nm en L.mol<sup>-1</sup>.cm<sup>-1</sup> ; l : chemin optique en cm (Nanodrop = 1 cm (valeur corrigée)) et  $c^{\text{mol}}$  : la concentration de la solution en mol/L. Les coefficients d'extinction molaire de chaque protéine étudiée sont récapitulés dans les **tableaux 5 et 6** (paragraphe IV.2 & IV.3). Les protéines purifiées sont ensuite séparées en aliquotes de 25 µL pour être congelées rapidement par immersion dans de l'azote liquide et conservées à - 80°C jusqu'à utilisation.



**Figure 27. Représentation des étapes de production et de purification des protéines**

**1. Production de la protéine :** le virus de la vaccine Ankara recombinant est utilisé pour infecter des cellules BHK21 en suspension. L'expression de Gag est induite par l'ajout d'1 mM d'IPTG dans le milieu de culture. L'infection et l'induction des cellules durent 24h à l'issue desquelles les cellules sont culotées, récupérées et conservées à -20°C jusqu'à l'étape de purification. **2. Purification par affinité :** les culots sont repris dans la solution tamponnée de lyse pour être détruits par sonication. Les débris de cellules sont séparés des protéines et des acides nucléiques grâce à une étape d'ultracentrifugation et de filtration. Les protéines sont purifiées grâce à l'étiquette présente dans leur région N-terminale, grâce à une colonne d'affinité (ici représentée par une HisTrap™ Excel). À l'issue de la purification par affinité un SDS page est réalisé pour contrôler les fractions qui contiennent la protéine d'intérêt, les regrouper et les concentrer avant de réaliser la purification en fonction de leur taille. **3. Purification en fonction de la taille (SEC) :** les protéines ainsi concentrées sont ensuite purifiées lors d'une seconde étape de purification en fonction de la taille. Pour cela une colonne S200 16/60 est utilisée. Les fractions contenant la protéine sont de nouveau identifiées grâce à un SDS-Page avant d'être regroupées et concentrées. Les protéines nouvellement purifiées sont ensuite analysées à l'aide de différentes méthodes biophysiques. **4. Analyses biophysiques :** la séquence, les modifications post traductionnelles, le PM, le rayon hydrodynamique, le coefficient de sédimentation et la température de dénaturation sont déterminés, comparés et discutés à l'aide d'une batterie d'analyses biophysiques

## VI.2 Purification des protéines recombinantes en fonction de leur étiquette

### VI.2.1 Protéines 6His-Gag

Les culots de cellules BHK21 exprimant la protéine 6His-Gag sont re-suspendues dans une solution tamponnée de lyse. Elle est constituée d'un mélange entre les solutions tamponnées de conservation (tampons A, **tableau 6**) et 2% de solution tamponnée d'éluion (tampon B, **tableau 6**). Cette dernière, contient les mêmes éléments que le tampon A mais est supplémenté avec de l'imidazole (500 mM). Différentes solutions tamponnées (A/B) ont été mises au point pour purifier 6His-Gag (**tableau 6**). La première solution (ici appelée A1<sub>(solubilisation)</sub>) favorise la solubilisation de Gag grâce à une forte concentration ionique et de détergent. Les protéines 6His-Gag sont alors

purifiées comme décrit ci-dessus par affinité en utilisant une colonne HisTrap™ Excel (Cytiva, ref. 17-3712-06). La deuxième solution (ici appelée A2<sub>(pH acide)</sub>) est compatible avec la solubilisation de la protéine Vpr (voir §VI.2.3). Enfin, une dernière solution (ici appelée A3<sub>(500mM NaCl)</sub>) favorise la formation d'interactions en réduisant la concentration de chlorure de sodium (NaCl).

#### VI.2.2 Protéines R-ALIX

Les culots de cellules exprimant R-ALIX sont re-suspendues dans le tampon ALIX décrit (**tableau 6**). Suite à la sonication les protéines ALIX sont purifiées par affinité en utilisant une colonne Strep-Tactin® XT Sepharose™ (Cytiva, ref. 16486149). De la desthiobiotine (2,5 mM) est utilisée pour l'étape d'élution (B1<sub>(ALIX-solubilisation)</sub>).

#### VI.2.3 Protéines R-Vpr

Les culots de cellules exprimant R-VPR sont re-suspendues dans la solution tamponnée de conservation favorisant la solubilisation de VPR, tampon A2<sub>(pH Acide)</sub> (**tableau 6**). Une fois re-suspendues, les cellules sont lysées par sonication puis purifiées par affinité en utilisant une colonne Strep-Tactin® XT Sepharose™ (Cytiva, ref. 16486149). De la biotine (50 mM) est utilisée pour l'étape d'élution (B2<sub>(VPR pH Acide)</sub>).

Colonne d'affinité		Tampon de conservation (A)	Tampon d'élution (B)
<b>6His-Gag</b>	HisTrap™ Excel	<b>A1<sub>(solubilisation)</sub></b> : 7 mM CHAPS 500 mM NaCl 50 mM HEPES pH 7,5 2 mM MgCl <sub>2</sub> 2 mM 2-betamercaptoethanol	<b>B1<sub>(solubilisation)</sub></b> : 7 mM CHAPS 500 mM NaCl 50 mM HEPES pH 7,5 5 mM MgCl <sub>2</sub> , 2 mM 2-betamercaptoethanol <b>500 mM imidazole (pH 7,5)</b>
<b>R-ALIX</b>	Strep-Tactin® XT Sephacrose™	<b>ALIX</b> : 50 mM Tris HCl pH 7,5 500 mM NaCl	<b>ALIX</b> : 50 mM TRISHCL pH 7,5 500 mM NaCl <b>2,5 mM Desthiobiotine</b>
<b>6His-Gag</b>	HisTrap™ Excel	<b>A2<sub>(pH Acide)</sub></b> : 7 mM CHAPS 500 mM NaCl 50 mM TRIS HCl pH 6,8 2 mM MgCl <sub>2</sub> 2 mM 2-betamercaptoethanol	<b>B2<sub>(Gag-pH Acide)</sub></b> : 7 mM CHAPS 500 mM NaCl 50 mM TRIS HCl pH 6,8 5 mM MgCl <sub>2</sub> , 2 mM 2-betamercaptoethanol <b>500 mM imidazole (pH 7,5)</b>
<b>R-Vpr</b>	Strep-Tactin® XT Sephacrose™	<b>A2<sub>(pH Acide)</sub></b> : 7 mM CHAPS 500 mM NaCl 50 mM TRIS HCl pH 6,8 2 mM MgCl <sub>2</sub> 2 mM 2-betamercaptoethanol	<b>B2<sub>(Vpr-pH Acide)</sub></b> : 7 mM CHAPS 500 mM NaCl 50 mM TRIS HCl pH 6,8 5 mM MgCl <sub>2</sub> , 2 mM 2-betamercaptoethanol <b>50 mM Biotine (pH 6,8)</b>
<b>6His-Gag</b>	HisTrap™ Excel	<b>A3<sub>(500 mM NaCl)</sub></b> : 500 mM NaCl 50 mM HEPES pH 7,5 5 mM MgCl <sub>2</sub> 2 mM 2-betamercaptoethanol	<b>B3<sub>(500 mM NaCl)</sub></b> : 500 mM NaCl 50 mM HEPES pH 7,5 5 mM MgCl <sub>2</sub> 2 mM 2-betamercaptoethanol <b>500 mM imidazole (pH 7,5)</b>

**Tableau 6. Solutions tamponnées utilisées au cours des purifications des protéines produites en cellules de mammifères**  
Les éléments qui permettent l'étape d'élution lors de la purification d'affinité sont représentés en bleu et en gras.

## VII. Caractérisation des protéines purifiées

### VII.1 Détection des protéines recombinantes par Western Blot (WB)

Les protéines produites par des cellules de mammifères sont identifiées par WB. Dans ce but, 1 µg de protéine 6His-Gag est mélangé avec du bleu de charge LDS NuPAGE™ et de l'agent réducteur NuPAGE™ (Invitrogen, ref. NP0009). Le mélange est déposé sur un SDS-PAGE bis-acrylamide de 12,5% dans un tampon de Tris glycine SDS. À

l'issue de la migration, le gel est transféré sur une membrane en polyfluorure de vinylidène (PVDF) Midi, 22 µm (Bio-RAD, ref. 1704157) en utilisant la technologie de transfert du système Trans-Blot Turbo (Bio-RAD, ref. 10039859). La membrane est bloquée avec une solution tamponnée de Tris HCl pH 7,5, NaCl et Triton X100 supplémentée avec 5% de lait en poudre pendant 1 heure à 4°C. Les protéines sont détectées grâce à l'ajout d'un anticorps monoclonal de souris dirigé contre l'étiquette 6His (Santa Cruz Biotechnology, ref. sc-8036). La réaction dure toute la nuit à 4°C. Cet anticorps primaire est détecté à l'aide d'anticorps secondaires de chèvre dirigés contre les immunoglobulines G (IgG) de souris et conjugués à de la peroxydase de raifort (HRP ; Bio-Rad, ref1706516 ; 1/3000 v/v).

Pour les partenaires de Gag (R-Vpr, R-ALIX et R-UEV), un anticorps polyclonal dirigé contre le l'étiquette Strep-tag® conjugué à de la HRP (IBA, ref. 2-1507-001) est utilisé pour détecter les protéines (1 : 1000, v/v).

Enfin, l'ensemble des protéines est visualisé par chimioluminescence à l'aide du réactif de détection ECL *Prime Western blotting detection* (Cytiva, ref. RPN2236) et sont analysées en utilisant l'appareil de lecture de gels Chemi-Doc Touch Imaging system (Bio-RAD).

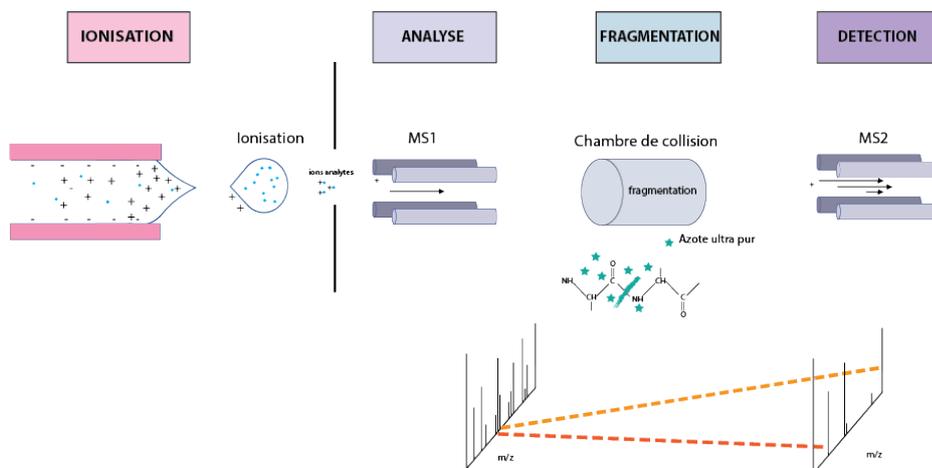
#### **VII.2 Contrôle de la séquence et des PTM des protéines par spectrométrie de masse en tandem (MS/MS)**

La spectrométrie de masse en tandem (MS/MS) est utilisée pour contrôler les séquences des protéines nouvellement produites, ainsi que pour localiser les PTM en se concentrant notamment sur les ubiquitinations et les phosphorylations. Ces analyses sont réalisées grâce à la « Plateforme Protéomique Strasbourg-Esplanade » (IBMC, Strasbourg) en collaboration avec Philippe Hammam, Lauriane Kuhn et Johana Chicher.

Pour les analyses, 10 µg de protéines sont préparés, soit en solution dans leur solution tamponnée, soit en découpant la bande d'intérêt d'un SDS-PAGE. La préparation d'échantillon en solution permet, d'obtenir des informations sur la pureté de la protéine, sur l'identité des contaminants et sur la présence de PTM. Les analyses réalisées sur la bande de SDS-PAGE d'intérêt, permettent d'obtenir des informations plus exhaustives sur la séquence de la protéine d'intérêt et notamment sur la présence d'éventuelles PTM. Les échantillons en solution sont précipités avec de l'acétate ammonium (0,1 M final) et dénaturés avec du méthanol (MeOH ; 100%). Les bandes de gel révélés en utilisant une solution de bleu de Coomassie sont découpées et décolorées à l'aide de plusieurs bains. Le premier bain contient du bicarbonate d'ammonium (50 mM), le second contient du bicarbonate d'ammonium (50 mM) supplémenté avec de l'acétonitrile (50% final). Les morceaux de gel sont immergés cycliquement dans ces deux solutions jusqu'à leur décoloration totale (gel translucide).

Les protéines issues de l'une ou l'autre méthode (en gel ou en solution), sont re-suspendues dans une solution de dithiothreitol (5 mM) et iodoacétamide (10 mM) pour pouvoir réduire et alkyler les protéines. Elles sont ensuite digérées toute la nuit à 37°C dans une étuve avec de la trypsine (Promega, ref. V5280) pour hydrolyser les liaisons peptidiques après les acides aminés basiques (lysine, arginine) ou de la chymotrypsine (Promega, ref. V1061) pour couper après les résidus aromatiques (phénylalanine, tyrosine, tryptophane) et les résidus méthionine et leucine. Les enzymes sont utilisées avec un ratio 1 : 20 v/v. Les peptides digérés sont ensuite séparés à l'aide d'une chromatographie en phase stationnaire chromatographie en phase liquide haute performance (HPLC) *easy nano* C100, de type C18 (ThermoFischer, ref. 164569). Cette HPLC en phase inverse permet de séparer les peptides par hydrophobicité. Les protéines sont ensuite ionisées par électro-spray (Smith et al., 1990). Le nébulisât rentre dans le spectromètre de masse QExactive-Plus Orbitrap (ThermoFisher scientific, ref. F019). Les ions sont triés en fonction de leurs valeurs de masse-sur-charge ( $m/z$ ). Pour pouvoir identifier les peptides, une étape supplémentaire de fragmentation est réalisée. Pour cela, les peptides sont bloqués dans une chambre de collision et des molécules de d'azote ultra pur sont ajoutées. La collision entre le gaz et les peptides va induire des coupures au niveau des liaisons peptidiques entre les acides aminés. Ces ions fragmentés sont ensuite accélérés dans un champ électrique, puis dirigés et piégés entre les pôles d'un aimant selon une trajectoire circulaire qui dépend de leur rapport  $m/z$ . L'ensemble des mesures constitue un spectre de fragmentation (**figure 28**). Par convention, les ions fragmentés contenant l'extrémité N-terminale du peptide sont appelés ions b et les fragments contenant l'extrémité C-terminale sont appelés ions y (Roepstorff and Fohlman, 1984). Tous les ions mesurés, constituent les spectres expérimentaux. En parallèle, la séquence FASTA *Homo sapiens* et *VIH* contenues dans la base de données publique *UniProtKB* et la séquence théorique de la protéine d'intérêt sont découpées en peptides après les acides aminés R et K, en utilisant l'algorithme Mascot version 2.6 (Matrix Science). Chaque séquence en acides aminés découpée virtuellement, constitue un spectre MS2 théoriques. Les spectres expérimentaux qui ont des pics avec des hauteurs de signal différentes sont alors comparés aux spectres théoriques. Afin d'assigner le plus de spectres expérimentaux à des séquences peptidiques, l'algorithme tolère la présence de certaines modifications lors des comparaisons des spectres expérimentaux et théoriques. En effet, l'algorithme Mascot cherche systématiquement la présence de carbamidomethylation sur les Cystéines (modification fixe), la présence éventuelle d'acétylation sur la région N-terminale des protéines ainsi que l'oxydation sur les Méthionines (modifications variables). Toutes les autres modifications sont indiquées comme variables. Ainsi, l'algorithme réalise des recherches de phosphorylation des Serines, Thréonine et Tyrosine, d'ubiquitination des Lysines et des Cystéines, d'oxydation des Méthionines et des SUMOylations par la protéine SUMO1 qui induit un incrément de 2135 Da et la SUMOylation les

protéines SUMO2/3 induisant un incrément de 3549 Da, sur les Lysines. La configuration de l'instrument est spécifiée comme « ESI-Trap » et les tolérances de masse en MS et MS/MS sont fixées à 10 ppm et 0,02 Da respectivement. Les fichiers Mascot.dat résultants sont importés dans le logiciel *Proline* v2.0 (<http://proline.profiroteomics.fr>) pour le traitement des résultats. Les protéines sont validées avec Mascot en utilisant les rangs égaux à 1, un taux de faux-positifs (*False Discovery Rate*, FDR) inférieur à 1% au niveau des correspondances spectre-peptide (PSM ou *Peptide Spectrum Match*) mais aussi au niveau de l'ensemble de protéines. Enfin, les localisations des PTM sont revérifiées manuellement.



**Figure 28. Représentation de l'analyse des macromolécules par MS/MS**

Les protéines sont séparées à l'aide d'une HPLC en phase inverse à l'issue de laquelle les protéines sont ionisées par la méthode ESI. Les ions entrent dans le spectromètre ou ils sont triés en fonction de leur valeur de rapport masse-sur-charge ( $m/z$ ). Pour pouvoir identifier les peptides, une étape supplémentaire de fragmentation est réalisée dans la chambre de collision. Ces ions fragmentés sont ensuite accélérés dans un champ électrique, puis dirigés et piégés entre les pôles d'un aimant selon une trajectoire circulaire qui dépend de leur  $m/z$ . L'ensemble des mesures constitue un spectre de fragmentation.

### VII.3 Echange de solutions tamponnées par dialyse des protéines 6His-Gag purifiées

Dans le but de caractériser l'état oligomérique et la solubilité de la protéine Gag dans les différentes solutions tamponnées A (**tableau 6**), la protéine recombinante 6His-Gag est soit purifiée directement dans la solution d'intérêt, soit elle est dialysée. Notamment pour le tampon A3. L'utilisation d'un tel tampon ne permet pas d'avoir de forts rendements de purification car il présente une limite de solubilisation autour de 0,7 mg/mL, tandis que le tampon A1 permet d'obtenir des concentrations de 4 mg/mL. Ainsi, des dialyses sont réalisées pour réduire les pertes de protéines lors des étapes de purifications. Pour cela, la protéine Gag purifiée dans le tampon A1 est décongelée sur glace. Un volume minimum de 50  $\mu$ L est déposé dans des D-Tubes™ Dialyzer Midi ayant un seuil d'exclusion de 3,5 kDa (Merck Millipore, ref. 71506) pour être dialysé, pendant la nuit à 4°C, contre la solution tamponnée d'intérêt. La protéine dialysée est ensuite centrifugée 30 minutes à 14000 g à 4°C. Un aliquote du

surnageant est contrôlée par DLS (détaillé section Matériels et méthodes, §IX.1). Le reste de l'échantillon est transféré dans des tubes Eppendorf à faible adhésion de surface (ThermoFisher Scientific, 90410) et conservé sur glace jusqu'à utilisation.

#### VII.4 Marquage fluorescent des protéines Gag purifiées

Certaines méthodes biophysiques telles que l'anisotropie de fluorescence et FIDABio utilisées lors de la caractérisation des protéines Gag (produites en système bactérien et mammifère) et des complexes entre ces dernières et leurs partenaires, nécessitent le marquage préalable de Gag par une sonde fluorescente. Pour cela, les protéines Gag sont marquées en suivant le protocole du kit Fluorospin 490 (Dyomics, ref. PMKNAHC). Brièvement, un ester succinimidyle réactif de DY-490 (Dyomics) forme une liaison covalente stable avec les lysines libres des protéines Gag. Le volume de marqueur fluorescent à ajouter dans le milieu dépend directement du volume initial et de la concentration initiale de la protéine comme décrit dans la formule ci-dessous :

$$\frac{C_{massique\,Gag} * Volume_{Gag} * 1000}{C_{fluorophores} * MW} * MR (= 10) \quad (4)$$

Pour améliorer le marquage, 0,05% de Tween 1% sont ajoutés à la préparation. La réaction de marquage est réalisée à température ambiante [22; 25°C] ; à l'obscurité, pendant 1h. La protéine marquée est purifiée en suivant le protocole des colonnes Sephadex™ – G-25 DNA (Ge Healthcare, ref. 17-0853-02) préalablement équilibrées avec la solution tamponnée expérimentale (A1, A2 ou A3 ; **tableau 6**). La protéine est éluée par ajout de 50 µL de tampon A. Cette étape est réalisée 9 fois. Cela permet de récupérer de plus fortes concentrations de protéines plutôt qu'une concentration diluée par les volumes moins concentrés. La protéine marquée est conservée dans des tubes Eppendorf à faible adhésion de surface (ThermoFisher Scientific, 90410) à l'obscurité à 4°C jusqu'à analyse. L'efficacité de marquage est évaluée en utilisant les formules suivantes,

i. Correction de la concentration molaire de la protéine en fonction de l'absorbance de la sonde :

$$C^{mol} = \frac{Abs\,280nm - (Abs\,294nm * 0.11)}{\varepsilon\,280\,nm\,Gag} \quad (5)$$

ii. Calcul du nombre de fluorophores par protéine :

$$\frac{Abs\,294\,nm}{\varepsilon\,fluorophore\,(71000) * C_{molaire\,Gag}} \quad (6)$$

### VIII. Formation des complexes entre Gag et ses partenaires protéiques et nucléiques

#### VIII.1 Formation des complexes entre la protéine 6His-Gag et l'ARN111<sub>WT</sub>

Pour former les complexes entre Gag et l'ARNg viral produit *in vitro*, nous avons suivi deux procédures. D'une part, les deux partenaires sont mélangés dans un Eppendorf 1,5 mL et immédiatement caractérisés par l'ensemble des

méthodes utilisées au sein de ce travail de thèse (DLS, anisotropie de fluorescence, CryoEM, paragraphes IX et X). D'autre part, le mélange est incubé 30 minutes à température ambiante dans des tubes Eppendorf à faible adhésion de surface (ThermoFisher, ref. 90410) avant d'être analysée par DLS, AUC, FIDABio. Les contrôles négatifs sont réalisés en utilisant l'ARN111<sub>AAA</sub> (**tableau 2**) et l'ARN de transfert (ARNt) de levure Ambion™ (ThermoFisher ; ref. AM7119).

### **VIII.2 Formation des complexes entre 6His-Gag et ses partenaires protéiques (ALIX, Vpr)**

Pour former les complexes entre Gag et ses partenaires (ALIX et Vpr) différentes solutions tamponnées sont testées. D'une part, le tampon ALIX est utilisé pour former les complexes entre 6His-Gag et R-ALIX. D'autre part, le tampon A3 est utilisé pour former les complexes entre 6His-Gag et les protéines Vpr synthétisée chimiquement et produites en cellules BHK21. Vpr est décrit dans la littérature comme étant soluble à pH acide (Cornille et al., 1999). Les complexes sont formés par mélange et incubation 30 minutes à température ambiante dans des tubes Eppendorf à faible adhésion de surface (ThermoFisher, ref. 90410). Ces complexes sont caractérisés en utilisant la méthode biophysique FIDABio (Section Matériels et Méthodes, §IX). Un contrôle est réalisé avec la protéine Gag-6His produites en cellules bactériennes et la protéine R-Vpr.

## **IX Méthodes biophysiques d'analyse de l'ARNg, de Gag et des complexes**

### **IX.1 Diffusion dynamique de la lumière (DLS)**

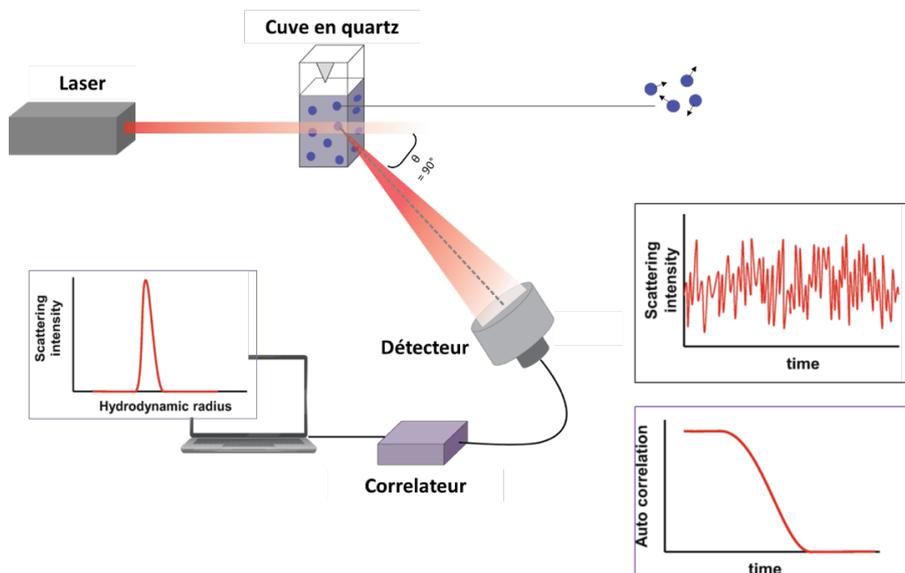
La méthode DLS (*Dynamic Light Scattering*) est utilisée en routine dans cette étude pour déterminer l'état oligomérique et le taux d'agrégation des macromolécules en solution. Cette méthode permet de caractériser l'effet des conditions expérimentales sur la solubilité, la stabilité et l'état oligomérique des macromolécules. C'est un outil crucial dans la mise au point des solutions tamponnées de solubilisation des protéines produites en cellules de mammifères. De manière générale, elle permet de déterminer le rayon hydrodynamique (Rh) des protéines, des ARN, ainsi que des complexes protéines-ARN en solution.

Pour contrôler l'état multimérique et le Rh de l'ARN111<sub>WT</sub> ou des protéines 6His-Gag, les échantillons sont préparés avec une concentration minimale de 0,3 mg/mL, correspondant à limite expérimentale de détection de Gag et de l'ARN. En parallèle les complexes entre Gag et l'ARN sont formés comme décrit section Matériels et Méthodes ; §VIII.1 et les échantillons sont préparés avec une concentration minimale de 0,4 mg/mL de Gag (équivalent à 6,8 µM). Pour réaliser un point de mesure, 5 µL d'échantillon sont déposés dans une cuvette en quartz (Wyatt Technology, ref. JC-006, PIN16269702). Les intensités de lumière diffusées sont mesurées en utilisant l'appareil

DynaPro Nanostar™ (Wyatt Technologies). Un laser hélium/ néon (HeNe) de 100 mW d'une longueur d'onde de 633 nm (Wyatt Technologies) éclaire des particules en suspension dans la solution d'analyse (**figure 29**). Ces particules sont sujettes aux mouvements browniens et diffusent la lumière dans toutes les directions (diffusion de Rayleigh). Les photons diffusés sont mesurés par un détecteur qui se trouve à un angle donné (90°), au cours du temps. Les mouvements, induisent la fluctuation de l'intensité de lumière diffusée. Le signal est analysé par un auto-corrélateur qui prend en compte les signaux pris à des intervalles de temps variables (de la nanoseconde à la microseconde). La dérivée de la fonction d'autocorrélation permet de déterminer le coefficient de diffusion (D). En mettant l'hypothèse que les protéines sont des sphères pleines, il est possible de déterminer le Rh de la molécule en solution à l'aide de l'équation de Stokes-Einstein décrite ci-dessous.

$$D = \frac{k_b \cdot T}{6\pi \cdot \eta_s \cdot Rh} \tag{7}$$

Avec D le coefficient de diffusion,  $k_b$  la constante de Boltzmann  $1,380\ 649 \times 10^{-23} \text{ J}^{\circ}\text{K}^{-1}$ , T la température (°K),  $\eta_s$  la viscosité du solvant et Rh rayon hydrodynamique. Les mesures sont réalisées à 20°C et en moyenne 10 acquisitions sont acquises pour l'analyse des résultats après un temps d'équilibration de deux minutes pour permettre à la température et aux particules de se stabiliser. Les distributions en intensité de lumière diffusée et en masse sont utilisées pour l'analyse des résultats, bien que la dernière suppose que toutes les molécules en solutions ont la même forme et les mêmes propriétés optiques.



**Figure 29. Représentation d'un instrument de diffusion dynamique de la lumière (DLS ; wyatt)**

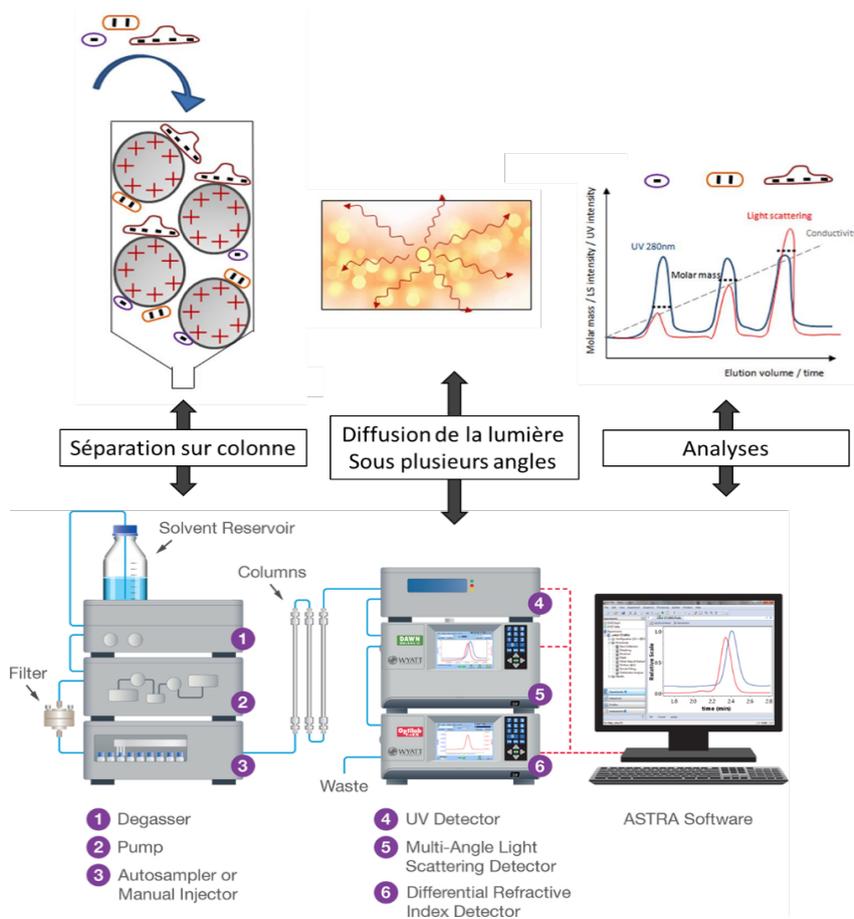
Une cuve en quartz est éclairée par un laser hélium néon. Les particules composant l'échantillon réalisent des mouvements browniens en solution. Elles sont alors capables de diffuser la lumière dans toutes les directions (diffusion de Rayleigh). Les photons diffusés sont mesurés par un détecteur qui se trouve à un angle de 90°. Il mesure l'intensité de lumière diffusée au cours du temps. Le signal est analysé par un auto-corrélateur qui prend en compte les signaux à des intervalles de temps variables. La dérivée de la fonction d'autocorrélation permet de déterminer le coefficient de diffusion (D). Figure adaptée de Misono, 2019.

## IX.2 Diffusion de lumière multi-angle couplée à une chromatographie à exclusion de taille (SEC-MALS)

La diffusion de lumière multi-angle couplée à une chromatographie à exclusion de taille (SEC-MALS) est une méthode biophysique combinant la diffusion de la lumière sous plusieurs angles avec une chromatographie à exclusion de taille. Dans cette étude, elle est utilisée pour caractériser le PM expérimental et l'état multimérique de 6His-Gag en solution dans différentes solutions tamponnées (A1 et A3, **tableau 6**). Elle permet aussi de déterminer le pourcentage de micelles de 3-[(3-cholamidopropyl)diméthylammonio]-1-propanesulfonate (CHAPS) entourant la protéine purifiée dans le tampon A1. En effet, la concentration de formation de micelle critique (CMC) du CHAPS est de 6 mM. Ces analyses sont réalisées au sein de la « Plateforme Biologie structurale intégrée, plateau biophysique moléculaire » (CBI, IGBMC, Illkirch), en collaboration avec Pierre Poussin.

Afin de réaliser l'analyse la protéine doit être purifiée et la plus concentrée possible. Ainsi, pour déterminer l'état oligomérique de 6His-Gag purifiée dans le tampon A1, 100 µL de protéines sont décongelés sur glace et centrifugés 30 minutes à 14000 g à 4°C. Le surnageant est prélevé et la concentration est estimée en mesurant l'absorbance de la protéine purifiée à 280 nm comme décrit précédemment. Dans le but de contrôler l'état multimérique et le PM de 6His-Gag dans le tampon A3, la protéine purifiée dans le tampon A1 est décongelée et dialysée contre la solution tamponnée A3.

Une fois les échantillons préparés, entre 50 et 100 µL de protéines sont injectés sur le système de chromatographie en phase liquide à haute pression Superdex S200, 10/300 (de Cytiva, ref. 17-5175-01), à un débit de 0,5 mL/minute. Une fois séparées en fonction de leur taille, les particules passent devant le détecteur à diffusion de la lumière miniDAWN TREOS (Wyatt technology). La diffusion de la lumière des particules en solution est mesurée selon trois angles, 43,6°, 90° et 136,4°. Le miniDAWN TREOS possède un laser de longueur d'onde de 658,9 nm avec un module de diffusion dynamique de la lumière (*Quasi Elastic Light Scattering*, QELS) pour déterminer le Rh, une cellule UV pour mesurer l'absorbance de la protéine à 280 nm et un réfractomètre Optilab T-rEX (Wyatt Technology) pour mesurer l'indice de réfraction de l'échantillon (**figure 30**).



**Figure 30. Représentation du système de chromatographie d'exclusion de taille couplé à de la diffusion de lumière multi-angle (SEC-MALS ; Wyatt)**

L'échantillon est séparé sur une colonne de chromatographie à exclusion de taille grâce à un appareil AKTA purifier. Une fois les protéines séparées en fonction de leur taille, elles passent à travers des systèmes de détections aux UV, de diffusion de lumière multi-angle (MALS) et de différence d'indice de réfraction (RI). Les signaux de sortie des détecteurs sont combinés et analysés par le logiciel ASTRA (d'après Malverne Panalytical et Amartely et al., 2018).

L'indice de réfraction de 6His-Gag ( $dn/dc$ ) est défini comme étant de 0,185 mL/g et celui du CHAPS comme étant de 0,1323 mL/g. Toutes les analyses sont réalisées à 25°C. La collecte des données et l'analyse SEC-MALS sont réalisées avec le logiciel ASTRA 7 (Wyatt Technology ; Wyatt, 1993 ; Lorber, 2018; Velours et al., 2022).

Les PM sont déterminés grâce à la formule suivante :

$$I_{\theta} \propto PM * C * \left(\frac{dn^2}{dc}\right) \tag{8}$$

Avec  $I_{\theta}$  correspond à l'intensité de signal mesurée par le MALS, PM correspond au PM expérimentale de la protéine, C la concentration et le  $dn/dc$  est mesuré par le DRI.

Pour la protéine 6His-Gag purifiée dans le tampon A1 (**tableau 6**) le modèle *Protein Conjugate Analysis* est utilisé.

Il est défini par l'équation suivante

$$\frac{\Delta n}{\left(\frac{dn}{dc}\right)_{\text{molécule1}} * \left(\frac{dn}{dc}\right)_{\text{molécule2}} * (1 - x)} = \frac{Abs/L}{\epsilon_1 * \epsilon_2 * (1 - x)} \quad (9)$$

Avec x la fraction massique du composant 1 (ici Gag) et donc (1 - x) est la fraction massique du composant 2 (ici CHAPS). Les dn/dc pour chaque composant et le coefficient d'extinction molaire pour Gag ( $\epsilon_1$ ). Abs correspond à l'absorbance mesuré au détecteur UV et L la longueur du chemin parcouru (Wyatt, 1993).

### IX.3 Analyse de la dispersion induite par flux (FIDAbio)

La technologie FIDAbio consiste à calculer la dispersion d'une macromolécule marquée en solution dans un capillaire micro fluidique à partir de son profil d'écoulement laminaire. Cette méthode permet de caractériser l'état oligomérique et le Rh des protéines marquées. Elle permet aussi de suivre la formation de complexes entre la molécule marquée, appelée Indicateur et son partenaire dans la solution tamponnée, appelé Analyte (**figure 31**). Les analyses FIDA sont réalisées à l'aide de l'instrument FIDalyzer (Fida Biosystems, ApS). Au cours de ces analyses nous avons choisi de marquer la protéine 6His-Gag, comme décrit dans le §VII.4.

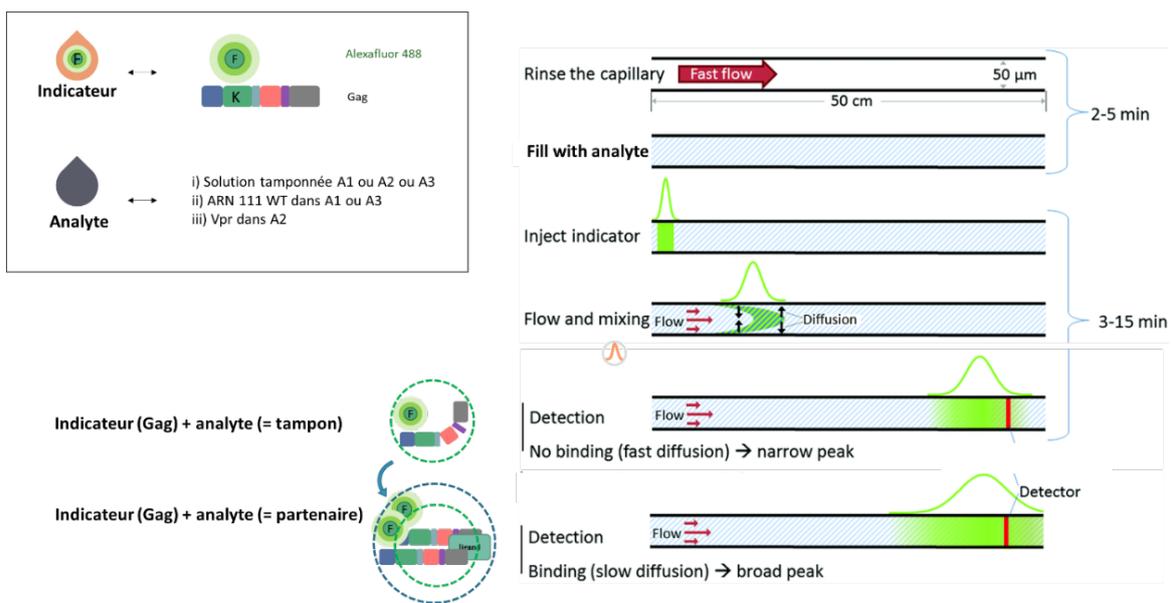
Une fois marquée, 12  $\mu$ L de protéines sont ajoutés dans des inserts en verre (Fida Biosystems, ref. 202-200). Le reste de l'analyse est réalisé automatiquement. Ainsi, l'Indicateur est aspiré dans un capillaire en silice (de 75  $\mu$ m de diamètre intérieur, 375  $\mu$ m de diamètre extérieur, d'une longueur totale de 100 cm ; Fida Biosystems, ref. 110-001V) et passe devant un laser diode d'une longueur d'onde d'excitation de 488 nm. La dispersion de la protéine marquée dépend des dimensions du capillaire en silice (fixes), du débit utilisé (fixe de 3 mm/s) et de son coefficient de diffusion dans le flux laminaire qui est calculé au cours de l'analyse.

Ainsi, la première étape consiste à déterminer la concentration à laquelle le signal d'intensité est suffisant pour réaliser le reste des tests. L'intensité du signal dépend de l'efficacité de marquage et de la concentration de la macromolécule. Ainsi des étapes de titrations sont nécessaires avant chaque analyse pour déterminer la concentration d'indicateur à utiliser pour réaliser les titrations (généralement comprise entre 100 et 500 nM). À la suite de chaque point de mesure, le capillaire est nettoyé à 1500 mbar pendant 3 minutes avec la solution tamponnée de conservation (A2 ou A3), puis il est équilibré avec la solution tamponnée pendant 45 secondes à 1500 mbar. L'indicateur est alors injecté à 50 mbar pendant 10 secondes (cela correspond à un volume de 39 nL). Une fois injecté, l'échantillon est poussé vers le point de détection par la solution tamponnée à 100 mbar pendant 600 secondes. Tous les échantillons sont analysés en duplicata. FIDalyzer réalise un calcul de la viscosité de l'échantillon, pour cela la température est contrôlée à 20°C à l'intérieur de l'instrument et du capillaire. Les

Taylorgrammes représentant l'intensité de fluorescence (RFU) en fonction du temps (en minutes) sont traités à l'aide du logiciel d'analyse de données FIDA (Fida Biosystems ApS) qui permet de calculer le Rh à partir des coefficients de diffusion mesuré (grâce à l'équation de Stokes-Einstein), compensé par les fluctuations de viscosité, comme décrit dans (Poulsen et al., 2015).

Ainsi ; le Rh (en nm) de 6His-Gag marquée dans les tampons A2 ou A3 est mesuré et est le contrôle dans l'étude de la formation des complexes entre Gag et ses partenaires protéiques ou nucléiques

Les complexes formé par 6His-Gag et l'ARN111<sub>WT</sub> sont étudiés dans le tampon A3. Pour cela, des concentrations croissantes d'ARN (150 pM à 300 nM) sont ajoutées à une concentration stable de Gag allant de 150 nM à 300 nM en fonction des analyses. Finalement, concernant les analyses sur les complexes entre 6His-Gag et Vpr produites synthétiquement ou en cellules BHK21, le Rh des complexes est évalué en rajoutant des quantités croissantes de Vpr (de 200 nM à 6000 nM) à une quantité fixe de Gag (de 300 mM) dans le tampon A2.



**Figure 31. Représentation du principe de la méthode d'analyse de la dispersion induite par un flux (FIDABio)**

Dans l'encadrée, association des mots de vocabulaire « indicateur » et « analyte » aux molécules utilisées dans l'étude. Et présentation du flux de travail dans une analyse FIDABio. Après le rinçage, le capillaire est rempli avec l'analyte (tampon ou tampon + partenaires). Un volume de 10 nL d'indicateur est injecté et déplacé vers le détecteur par un flux sous pression. Pendant le trajet l'indicateur est mélangé à l'échantillon. La taille du pic de l'indicateur est liée à la diffusion apparente. Plus le degré de liaison est élevé plus la valeur de diffusion l'est (refait d'après Poulsen et al., 2015).

#### IX.4 Ultracentrifugation analytique (AUC).

L'ultracentrifugation analytique (*Analytical UltraCentrifugation*, AUC) permet de déterminer le coefficient de sédimentation (c(s)) en Svedberg, le coefficient de friction (f/f<sub>0</sub>), la distribution des espèces, le PM expérimental

et le Rh de macromolécules en solution (**figure 32**). Brièvement, l'AUC se base sur la théorie de Svedberg ou le coefficient de sédimentation équivaut à  $10^{-13}s^{-1}$  et peut être défini par la formule suivante :

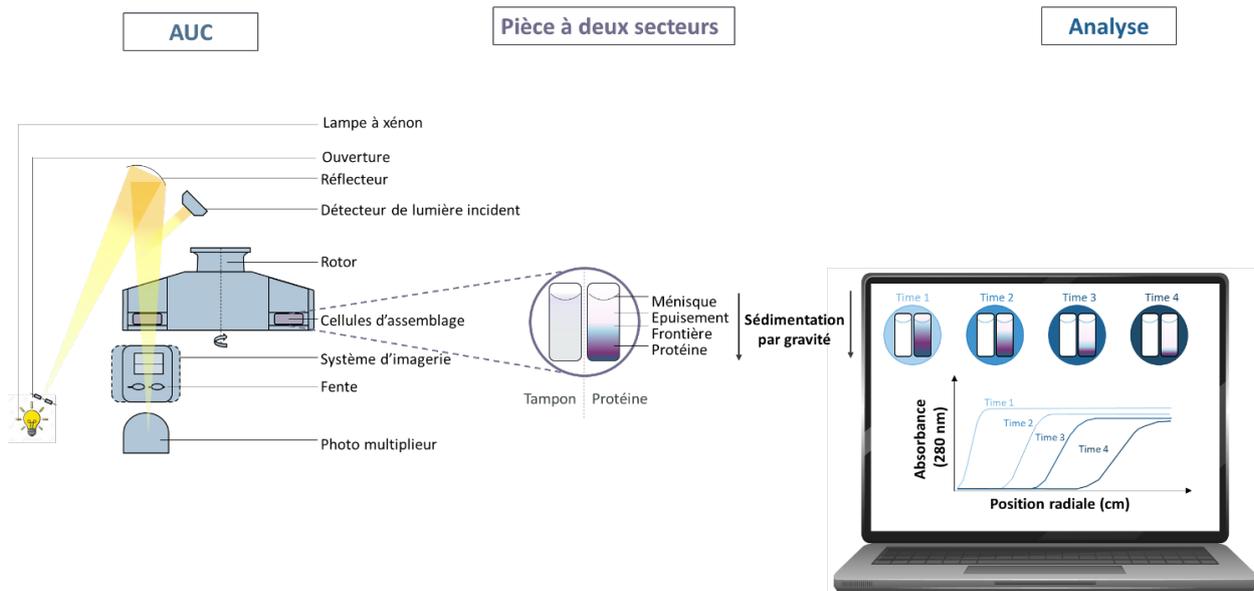
$$\frac{M(1 - \bar{v} \cdot \rho_s)}{N_A \cdot f} = \frac{u}{\omega^2 r} \equiv s \quad (10)$$

Il est alors possible de déterminer la masse molaire en utilisant l'équation de Svedberg et le coefficient de friction (f) qui dépend à la fois de la taille et de la forme de la particule. Ici dans un régime laminaire. Ainsi en utilisant l'équation de Stokes-Einstein, (équation 7, section Matériels et méthodes ; §IX.1), il est possible de déterminer :

$$\frac{s \cdot RT}{D \cdot (1 - \bar{v} \cdot \rho_s)} = M \quad (11)$$

Cette technique est utilisée pour caractériser la dimérisation de l'ARN111<sub>WT</sub> et la formation de complexes entre la protéine 6His-Gag et l'ARN111<sub>WT</sub> dans les solutions tamponnées A1 et A3. Ces analyses sont réalisées au sein de la « Plateforme Biologie structurale intégrée, plateau biophysique moléculaire » (CBI, IGBMC, Illkirch) en collaboration avec [Catherine Birck](#).

Afin de réaliser des mesures de la vitesse de sédimentation, 400 µL d'ARN111<sub>WT</sub> purifiés et repliés ayant une absorbance à 260 nm comprise entre [0,3 et 1,2] sont déposés dans des cellules d'analyse AUC à double secteur (Beckam Coulter, ref. 334602). Finalement, exactement le même volume de solution tamponnée correspondante est déposé dans le deuxième secteur des cellules d'analyse AUC à double secteur (Beckam Coulter, ref. 334602). Les cellules sont alors placées dans le rotor AN-50 TI Analytical 8-Place Titanium Rotor (Beckam Coulter, ref. 363782). Pendant l'analyse le rotor est placé sous vide à 20°C. Les mesures de vitesses de sédimentation sont réalisées à 45000 rpm (155358 g ; avec r = 6,5 cm). L'AUC ProteomeLab XL-I® (Beckam Coulter) permet également de réaliser des mesures d'absorbance. Elles utilisent une lampe au Xénon pouvant aller de 190 nm à 800 nm, permettant de calculer la concentration de l'échantillon dans la cellule en appliquant la loi de Beer-Lambert. Elle mesure 1 point pendant 1 minute tout le long de l'analyse. Les résultats sont ensuite analysés en utilisant le logiciel *Sedfit*, en appliquant le modèle de distribution continue avec une variable discrète. Les paramètres appliqués sont une résolution de 100, un intervalle de valeur de sédimentation allant de 0.5 à 20 et un intervalle de confiance de 0.68. Le modèle permet de déterminer le coefficient de friction dépendant de la taille et de la forme des molécules et de calculer la distribution différentielle des coefficients de sédimentation c(s). La déviation de la racine de la moyenne des carrés (*Root Mean Square Deviation* ; RMSD) est utilisée pour valider ou infirmer les modèles.

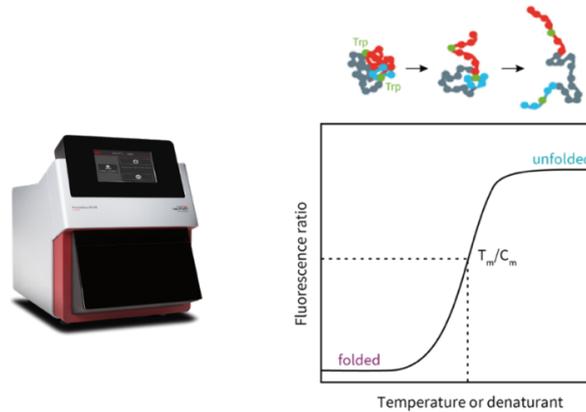


**Figure 32. Représentation du système d'ultracentrifugation analytique (AUC)**

Représentation de l'arrangement du système optique et des rotors pour une ultracentrifugeuse Beckman XL-1. Représentation d'une pièce à deux secteurs et de l'effet de la gravité sur les secteurs. Finalement représentation des scans optiques obtenus à l'issue d'une analyse d'AUC dans lesquelles le fond de migration descend vers le fond de la cellule au cours du temps d'après (Mitra, 2009)

### IX.5 Nano-fluorimétrie à balayage différentiel (NanoDSF)

Le Prometheus NT.48 fluorimeter (NanoTemper) est un appareil de NanoDSF utilisé pour déterminer la stabilité, la dénaturation thermique et l'oligomérisation de 6His-Gag sur la base de la fluorescence intrinsèque des acides aminés aromatiques tryptophane, tyrosine et phénylalanine (**figure 33**). La protéine 6His-Gag contient 9 tryptophanes, 31 tyrosines et 14 phénylalanines. La stabilité thermique d'une protéine est généralement décrite par la température de fusion (*Melting Temperature*,  $T_m$ ) et correspond à la température à laquelle 50% des protéines sont dénaturées. La dénaturation de la protéine entraîne l'accessibilité des acides aminés aromatiques provoquant une augmentation du signal d'intensité de fluorescence à 350 nm et 330 nm après excitation aux UV à 280 nm. La protéine 6His-Gag purifiée et concentrée dans les tampons A1 ou A3 (**tableau 6**) est centrifugée suite à sa décongélation comme précédemment décrit afin de perdre les éventuels agrégats présents en solutions comme vérifié par DLS. À l'issue de la centrifugation, 10  $\mu$ L de protéines sont aspirés dans les nano-capillaires NanoDSF-grade standard (Nanotemper). Les capillaires sont placés dans le porte échantillon du Prometheus NT.48 Fluorimeter (Nanotemper). Le logiciel *PR.ThermControl* (version 2.1.2) permet de contrôler la fluorimétrie. Un gradient de température linéaire est alors appliqué aux protéines de 20°C à 95°C avec un intervalle de température d'1°C par minute. La fluorescence à 330 nm et 350 nm émise par la protéine est mesurée en continue.



**Figure 33. Représentation du principe de la nano-fluorimétrie à balayage différentiel (NanoDSF ; Prometheus)**

Lorsqu'avec l'augmentation de la température la protéine est dénaturée, la fluorescence augmente d'après <https://2bind.com/nanodsf/>.

### IX.6 Caractérisation des complexes entre 6His-Gag et les ARN viraux par anisotropie de fluorescence

La polarisation de fluorescence (FP) est utilisée pour déterminer l'affinité de liaison entre la protéine 6His-Gag et les ARN111<sub>WT</sub> et ARNSL1<sub>35</sub> dans la solution tamponnée A3. Les complexes sont formés comme précédemment décrit section Matériels et Méthodes ; §VIII à concentration fixe d'ARN de 5 nM préalablement marqué par le fluorophore FTSC (comme décrit dans la section Matériels et Méthodes ; §III.2). Le mutant ARN111<sub>AAA</sub> sert de contrôle négatif. L'anisotropie est évaluée en rajoutant des quantités croissantes de Gag (de 5 nM à 500 nM) à l'ARNg, dans un volume final de 100 µL. L'ajout de quantité croissante de protéines permet d'obtenir une courbe de liaison entre l'ARN et la protéine dans le but d'estimer le  $K_D$  apparent du complexe. L'anisotropie est mesurée à 20°C dans les plaques 96 puits Greiner 96 F-Bottom (Merck, ref. M8185-100EA) par le PHERAstar FSX (BMG Labtech) en utilisant le module optique FP485520520. Une lampe à Xénon émet une lumière de longueur d'onde monochromatique de 485 nm qui est polarisée par un filtre de polarisation. Cette lumière polarisée excite les ARN viraux marqués en association avec les protéines en solution (**figure 34**). Le degré de polarisation de fluorescence est alors défini par l'anisotropie (A) :

$$A = \frac{P_{\parallel} - P_{\perp}}{P_{\parallel} + P_{\perp}} \quad (12)$$

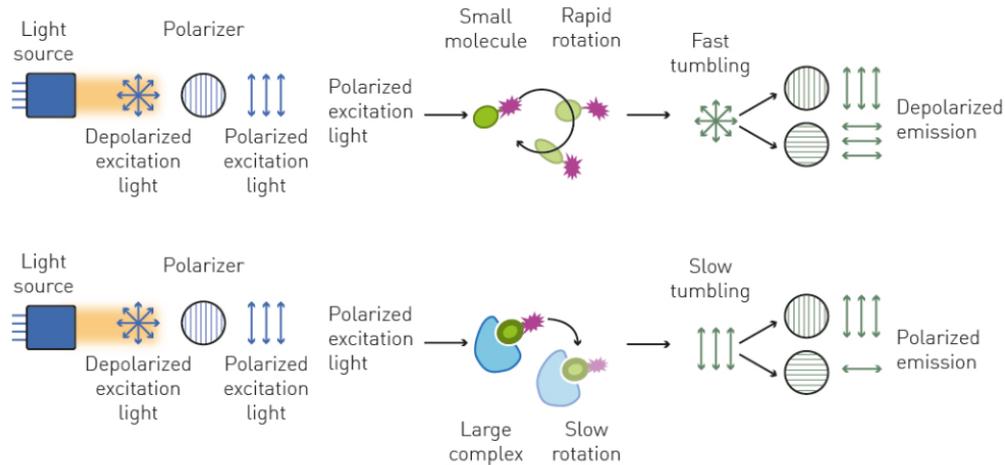
Où  $P_{\parallel}$  représente l'intensité de fluorescence émise parallèlement au rayon incident et  $P_{\perp}$  représente l'intensité de fluorescence émise perpendiculairement au rayon incident. L'émission fluorescente est évaluée à la longueur d'onde 520 nm.

L'équation de Scatchard sur la moyenne des courbes est utilisée pour calculer l'affinité et le nombre de sites de liaison entre la protéine et l'ARN (Scatchard, 1949).

$$\frac{K_{obs} \cdot P_f}{1 + K_{obs} \cdot P_f}$$

(13)

Avec  $K_{obs}$  la constante d'affinité observée,  $P_f$  la concentration de protéine libre.



**Figure 34. Représentation de la méthode d'anisotropie de fluorescence**

D'une part, l'excitation par une source de lumière polarisée de petites molécules fluorescentes marquées induit une rotation rapide des fluorophores. La lumière émise est alors dépolarisée. D'autre part, l'excitation par une source de lumière polarisée de ces mêmes petites molécules fluorescentes en complexe induit une rotation plus lente des fluorophores. La lumière émise est alors majoritairement polarisée d'après BmgLabtech (<https://www.bmglabtech.com/en/fluorescence-polarization/>).

## X. Etudes Structurales

### X.1 Résolution de la structure de 6His-Gag par microscopie électronique à transmission

L'ensemble des analyses de microscopie sont réalisées au sein de la plateforme « Biologie structurale intégrée, plateau microscopie électronique » (CBI, IGBMC, Illkirch) en collaboration étroite avec Alexandre Durant, Corinne Crucifix et Nils Marechal.

#### X.1.1 Par coloration négative

La première étape dans le processus de résolution de structure par microscopie électronique est la caractérisation de l'échantillon par coloration négative. Elle permet de contrôler le comportement de 6His-Gag sur les grilles et de tester l'homogénéité de la préparation de la protéine pour déterminer les conditions optimales de stabilité et de monodispersité. Ainsi, pour réaliser les grilles de coloration négative la protéine 6His-Gag purifiée dans le tampon A3 est centrifugée 40 minutes à 20000 g à 4°C. En parallèle la surface des grilles est rendue hydrophile (chargée négativement) par décharges lumineuses (*Glow Discharge*, GD). Brièvement, cette étape consiste à former un plasma grâce au passage d'un courant électrique de 100 v à travers des gaz (80% argon et 20% CO<sub>2</sub> plasmatique) pendant 30 secondes en utilisant le nanoclean 1070 (Fischione instruments). Une fois activée les grilles de carbone (*Cu continuous carbon 300 MESH*, Euromedex, ref CF-300-Cu-50) sont recouvertes par une goutte de 5 µL d'échantillon à une concentration comprise entre 10 ng/µL et 30 ng/µL et fixée avec 0,1% de glutaraldéhyde

pendant 1 minute. La grille est ensuite nettoyée et colorée avec des bains contenant de l'uranyle acétate (2%). C'est un colorant négatif capable de diffuser fortement les électrons et de s'adsorber rapidement sur les protéines. Les grilles sont observées grâce au microscope TECNAI F20 possédant une caméra Gatan CCD 2K et un canon à électrons *Field Emission Gun* (FEG) de 200 kV. Le microscope est piloté par les logiciels *serial EM 4.0* et *digital micrograph*<sup>®</sup>.

### X.2.2 Par Cryo-microscopie électronique (CryoEM)

La CryoEM est utilisée pour obtenir un modèle 3D à haute résolution de la protéine 6His-Gag purifiée. La protéine ayant un PM d'environ 59 kDa représente un vrai défi, notamment en termes de contraste. Les grilles portant les échantillons biologiques sont vitrifiées par refroidissement rapide dans un bain d'éthane liquide. Pour ce faire deux systèmes disponibles à la plateforme sont utilisés, le Vitrobot (ThermoFisher) et le Chameleon (Spt Labtech). D'une part, le Vitrobot (ThermoFisher) permet de tester des grilles de carbone et d'or. Ces grilles sont utilisées pour cribler différentes conditions de concentrations de 6His-Gag, allant de 0,18 mg/mL à 0,30 mg/mL et de concentrations de NaCl allant de 500 mM à 300 mM. Toutes les grilles vitrifiées à l'aide du Vitrobot sont hydrophilisées par GD avec 80% argon et 20 % CO<sub>2</sub> plasmatique pendant 1 minute et 30 secondes en utilisant le Nanoclean 1070 (Fischione instruments). Un volume de 5 µL est chargé manuellement sur les grilles. La température de préparation des échantillons dans le Vitrobot est contrôlée à 15°C, l'essuyage (*blot*) est appliqué avec une force de 5 pendant 2 secondes. Une fois préparées, les grilles sont conservées dans de l'azote liquide jusqu'à leur observation par le microscope électronique le Glacios (ThermoFisher).

Les premières grilles testées sont les grilles de carbone Quantifoil (Quantifoil, Ref. N1-C14nCR30-01), avec des trous de 1,2 µm de diamètre et une périodicité de 1,3 µm. Dans le but d'améliorer l'homogénéité de la glace, des grilles de carbone C-flat (*C-flat de 20 nm R 2/2 copper 300 MESH*, EDEN instruments, Ref. CF-2/2-3Cu-50), avec des trous de 2 µm de diamètre, une périodicité de 2 µm et une épaisseur de carbone de 20 nm sont également testées. En parallèle, pour améliorer l'homogénéité de la glace obtenue par vitrification, une autre surface est utilisée. Pour cela, des grilles d'or (*Au-flat R1,2/1,3 45nm on AU/Pd 300 MESH*, EDEN instrument, Ref. AUFT313-50) avec des trous de 1,2 µm de diamètre et une périodicité de 1,3 µm sont préparées. Finalement, la plateforme a mis au point une méthode de PEGylation de ces mêmes grilles d'or pour repousser et concentrer les protéines à l'intérieur des trous plutôt que sur la surface. L'avantage théorique est de pouvoir réduire la concentration de protéines. Ainsi, de nouveaux cribles de concentration de 6His-Gag allant de 0,1 mg/mL à 0,20 mg/mL sont réalisées. Afin d'améliorer la qualité des grilles d'or ou de carbone, la protéine 6His-Gag est fraîchement purifiée dans le tampon A3 avant chaque préparation de grille. Pour éviter la formation de précipités ou d'agrégats, la

protéine est prélevée en cours de concentration, lorsqu'elle est concentrée entre 0,15 et 0,20 mg/mL. Toutes les grilles sont préparées à partir de 5  $\mu$ L de protéines purifiées qui sont fixés avec du glutaraldéhyde (0,5%) et supplémenté avec du Dodécyl- $\beta$ -D-maltoside (DDM ; 0,0025%). Le temps de fixation de la protéine est d'une minute. Les acquisitions peuvent se dérouler sur un à trois jours.

D'autre part, le Chameleon nouvellement acquis par la plateforme, est utilisé pour réaliser des grilles de carbone quantifoil® active, de 1,2  $\mu$ m de diamètre, d'une périodicité de 0,8  $\mu$ m, recouvertes de nano-fils de cuivre favorisant la formation de films protéiques minces (SPT Labtech). Les grilles sont chargées négativement en appliquant entre 30 secondes et 1 minute de GD directement dans le Chameleon. Différents temps de plongeon allant de 100 à 500 millisecondes sont réalisés à partir d'un échantillon de 100  $\mu$ L de 6His-Gag purifiées dans le tampon A3 et ayant une concentration de 0,7 mg/mL. Une fois préparées, les grilles sont conservées dans de l'azote liquide jusqu'à observation au microscope électronique.

La technique de CryoEM consiste à utiliser un microscope électronique à transmission, pour acquérir de nombreuses images de protéines vitrifiées. Les données de CryoEM sont enregistrées sur le microscope électronique Glacios (ThermoFisher) qui est composé d'un optique X-FEG à 200 kV, de détecteurs directs d'électron *K2 Summit* et d'une caméra CETA-D, optimisée pour l'enregistrement de données de diffraction électronique. Le microscope est contrôlé par les logiciels *Serial EM* et *Digital Micrograph*. Des acquisitions sont réalisées en utilisant 8 secondes d'acquisition pour une dose totale de 55,97  $e^-/\text{\AA}^2$  sur les grilles C-flat ; 56,17  $e^-/\text{\AA}^2$  sur les grilles AU-PEG et 56,25  $e^-/\text{\AA}^2$  sur les grilles quantifoil® active.

Pour l'analyse des données et la résolution des structures, l'équipe est équipée d'un ordinateur rapide avec 4 processeurs GPU (SUPER-LVX-256GB-4-RTX-2080Ti-XEON, Linuxvixion) et des logiciels de dernière génération installés (*Gctf v1.06*, *Motioncorr v2.1.1*, *Relion 3.0*, *EMAN v2.22*, *CryoSparc*). Une fois les acquisitions réalisées, la résolution des structures est réalisée en utilisant principalement le logiciel *Cryo-SPARC*. Pour éliminer les déformations et les mouvements de l'échantillon pendant l'enregistrement d'une image, celle-ci est composée de 40 images qui sont alignés (*motion correction*). Ensuite, l'estimation de la fonction de transfert de contraste (CTF) est réalisée. Après nettoyage du jeu de données entre 300 et 500 particules sont sélectionnées manuellement. Elles sont utilisées pour réaliser un modèle d'apprentissage profond (*deep learning*). Ce modèle est ensuite utilisé pour la sélection et l'extraction des particules sur tout le jeu de données. Après élimination des mauvaises particules par plusieurs cycle de classifications 2D, les particules restantes sont utilisées pour générer un modèle initial. Plusieurs cycles d'affinement 3D permettront de générer le modèle final.

### **X.3 Modélisation de Gag par AlphaFold**

Les structures AlphaFold de Gag sont générées à partir d'AlphaFold v2.1.0 (ColabFold, <https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>). Pour la prédiction de la protéine, les séquences sont saisies une à une.

AlphaFold est exécuté une fois et génère 5 modèles par séquence indiquée. Afin d'indiquer la précision d'une prédiction, AlphaFold calcule le score d'estimation de confiance par résidu (pLDDT), représenté par un nombre sur une échelle entre 0 à 100 (Mirdita et al., 2022). La valeur de pLDDT est utilisée pour classer les modèles entre eux, ainsi plus le pLDDT est élevé, meilleure est la prédiction. Indépendamment de la structure 3D, AlphaFold donne une valeur d'erreur d'alignement prédit (*Predicted Aligned Error* ; PAE) et génère une carte d'erreur d'alignement entre paires de résidus pour chaque prédiction. Cette carte et cette valeur donnent une indication de confiance dans le positionnement des domaines, plus la valeur est faible entre les paires de résidus, plus les positions et les orientations relatives sont bien définies. Au contraire une valeur PAE élevée indique que les positions et/ou orientations relatives de ces domaines dans la structure 3D sont incertaines et ne peuvent pas être dans ce cas, interprétées.

### **X.4 Tests de cristallisation de 6His-Gag**

La résolution de structures protéiques par cristallographie aux rayons X repose sur la capacité intrinsèque des cristaux de protéine à diffracter ce type de rayonnement. Cette méthode permet donc d'obtenir des structures 3D à résolution atomique à partir de cristaux de protéines. Les analyses de cristallographie ici présentées sont réalisées au sein de la plateforme « Biologie structurale intégrée, plateau biocristallographie » (CBI, IGBMC, Illkirch) en collaboration étroite avec Alastair McEwen, Pierre Poussin et Nicolas Lévy.

Différentes conditions sont testées pour former des cristaux de 6His-Gag dans le tampon A3. La première étape est la dialyse de la protéine (purifiée dans le tampon A1) contre le tampon A3. Ensuite, des gouttes de cristallisation sont réalisées en mélangeant 200 nL de protéines à 100 ou 200 nL d'agents précipitants issus des kits commerciaux (PACT/PEG ion pH, Wizard I & II, Classics et JCSG+) dans les boîtes de cristallisation MRC à deux zones de cristallisation par condition, au format 96 puits. Le criblage est réalisé à 20°C grâce à un robot Mosquito Crystal nano-volumes (SPT Labtech). Les gouttes ainsi réalisées sont qualifiées de gouttes assises et la technique d'équilibration est ici la diffusion de vapeur. Les boîtes de cristallisation sont alors suivies à intervalle défini à l'aide d'un système automatique d'imagerie des cristaux (Formulatrix). La qualité des cristaux est finalement testée sur le générateur de rayons X de l'IGBMC.

# Résultats et discussion

## objectif n°1



## I. Résultats : caractérisation des interactions *in vitro* entre l'ARNg et Gag

Pour comprendre comment l'ARNg est recruté spécifiquement par Gag dans le cytoplasme, l'analyse structurale tridimensionnelle de ce complexe ribonucléoprotéique apporterait des informations précieuses. Afin de réaliser l'étude de la structure de la protéine Gag en association avec la région Psi de l'ARNg du VIH-1 de nombreuses mises au point sont réalisées. La première étape consiste à optimiser les conditions de purifications du précurseur Gag produit en cellules de mammifères, de manière à obtenir une protéine soluble, ayant toutes les modifications post-traductionnelles (PTM) nécessaires à son fonctionnement (**annexe 1**). Pour cela, deux solutions tamponnées sont utilisées et discutées dans la partie ci-dessous (section Matériels et Méthodes ; §I.1). La première est une solution de solubilisation (appelée dans le manuscrit tampon A1, dont la composition est indiquée dans le **tableau 6** (partie Matériels et Méthodes, §IV.2) et la deuxième est sans détergent et avec une plus faible force ionique (nommée tampon A3, **tableau 6**). Les protéines Gag purifiées dans l'une ou l'autre solution sont caractérisées par des méthodes biophysiques (récapitulées **tableau 8**). Ensuite, les conditions de formation des complexes entre la protéine Gag et le fragment d'ARNg précédemment identifié au laboratoire, l'ARN 111<sub>WT</sub> (Smyth et al., 2015) sont analysées en partie I.2, des résultats du manuscrit. Finalement cette analyse permet l'acquisition des premières images montrant la structure tri-dimensionnelle de Gag seule et en interaction avec l'ARN 111<sub>WT</sub> (voir partie I.3).

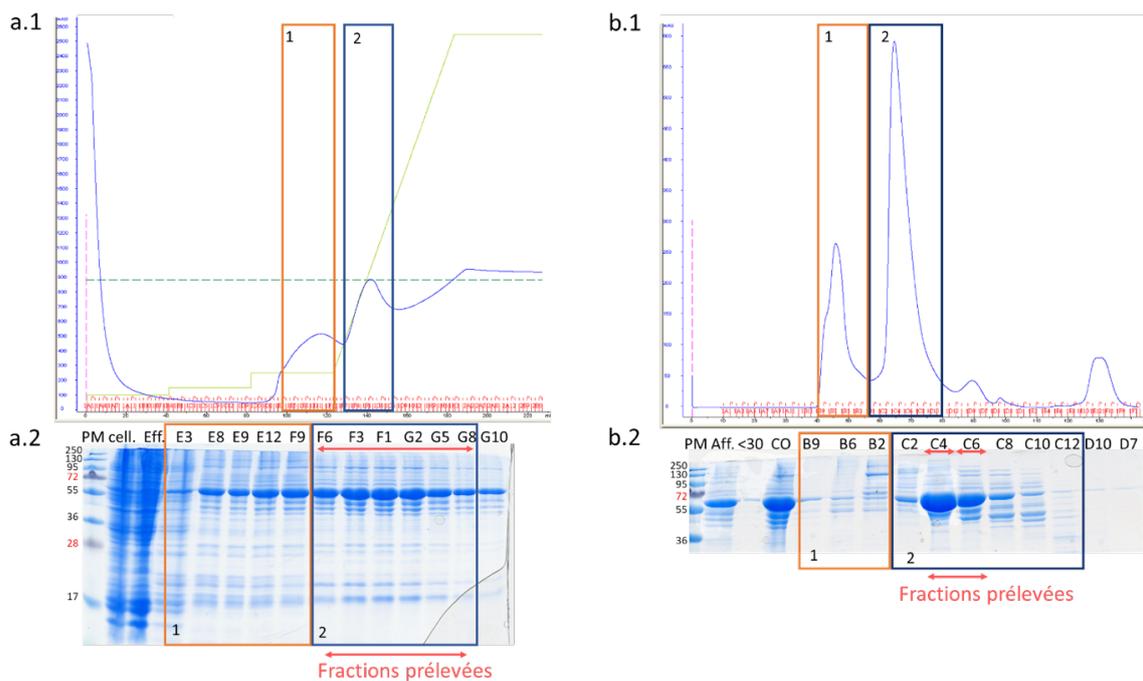
### I.1. Purification de 6His-Gag produite et sa caractérisation par des méthodes biophysiques

La protéine 6His-Gag est purifiée et caractérisée dans les solutions tamponnées A1 et A3, comme décrit dans la section Matériels et Méthodes. La stratégie expérimentale suivie est récapitulée **figure 27**, §VI.1.

#### I.1.1 Purification de la protéine 6His-Gag

La solution tamponnée A1 composée d'une forte concentration en NaCl (1 M) et de détergent (7 mM CHAPS), est utilisée pour solubiliser la protéine Gag. Ainsi, les culots cellulaires correspondant à 12 L de culture sont décongelés et re-suspendus dans 420 mL de solution tamponnée de lyse A1/B1 composée de 2% de tampon B (**tableau 6**, section Matériels et Méthodes ; §IV.2). Après une étape de lyse et d'ultracentrifugation, le surnageant est chargé sur la colonne HisTrap™ Excel Crude. En suivant l'absorbance à 280 nm, deux pics apparaissent sur le chromatogramme (**figure 35, a.1**). Le premier, est compris entre la fraction E3 et la fraction F9 et le second, entre la fraction F6 et la fraction G10 suivant le gradient d'imidazole. La pureté des fractions est contrôlée à l'aide d'un SDS-PAGE 12% (**figure 35, a.2**). Des aliquotes contenant les cellules lysées (indiquées comme « cell. » dans la **figure 35**) et d'effluent (« Eff. », **figure 35**) sont déposées dans les puits du gel pour contrôler la présence de la protéine dans les échantillons de départ. Les fractions présentant une bande de PM de 59 kDa et le moins de protéines contaminantes, capables de se lier à la résine de nickel qui sont représentées par des bandes de PM

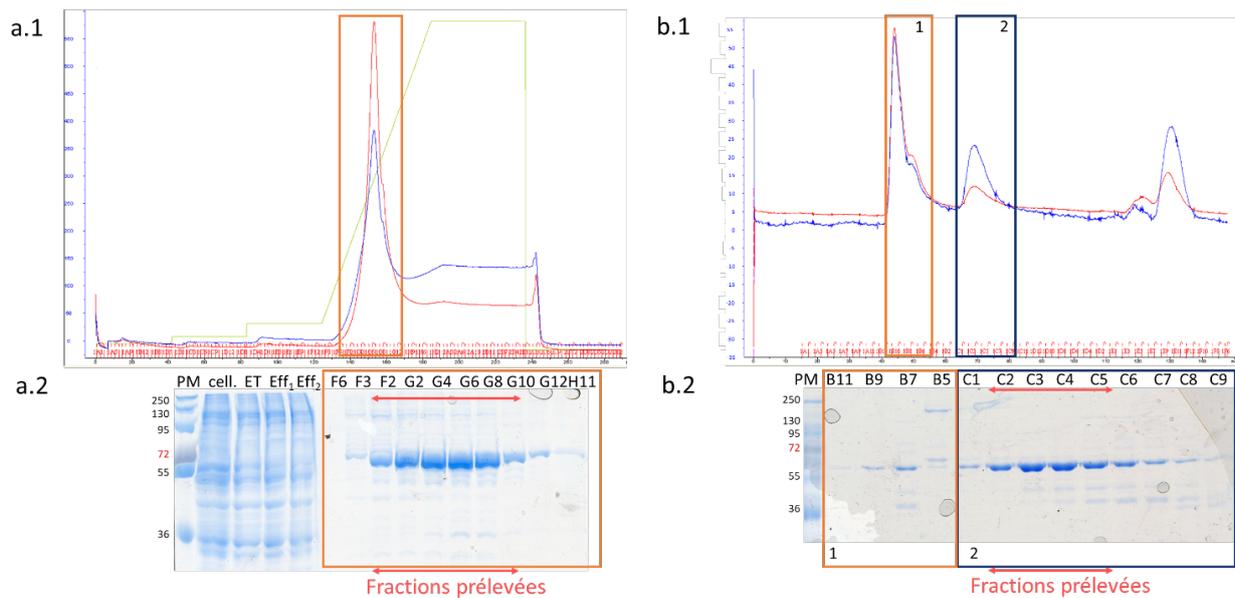
supérieur ou inférieur à 59 kDa, sont regroupées pour être concentrées. Ainsi les fractions allant de F6 à G8 sont regroupées. Une deuxième étape de purification est effectuée en utilisant la colonne S200 16/60 équilibrée dans le tampon A1 (**figure 35, b.1**). Au regard de l'absorbance à 280 nm, deux pics apparaissent sur le chromatogramme. Le premier pic est compris entre les fractions B9 à B2 et le second s'étend de C2 à D7. De nouveau la pureté de la protéine 6His-Gag sont contrôlées à l'aide d'un SDS-PAGE 12% (**figure 35, b.2**). Certains puits sont utilisés pour contrôler l'enrichissement de la pureté de l'échantillon par rapport à la sortie de l'affinité (cette piste du gel est indiquée comme « Aff. ») et pour contrôler les pertes lors de l'étape de concentration (piste indiquée comme « <30 »). Finalement, un puit sert à contrôler la pureté de la protéine après concentration (nommé « CO ») avant la charge sur la colonne d'exclusion de taille. A l'issue du gel, d'une part, la fraction C4 et d'autre part, les fractions C5 et C6 sont sélectionnées pour être. Finalement, 1,3 mL de la fraction « pure » correspondant à la fraction C4 et 1,3 mL correspondant aux fractions C5 et C6 sont récupérés ayant une concentration de 3,6 mg/mL. L'un et l'autre échantillon sont ensuite utilisés pour la caractérisation biophysique des protéines 6His-Gag.



**Figure 35. Purification de 6His-Gag dans le tampon A1**

**a.1** Chromatogramme : les encadrés orange (1) et bleu (2) représentent les différents pics observés. **a.2** SDS-Page 12% : à l'issue de la purification par affinité à l'aide d'une colonne HisTrap™ Excel. Le marqueur de poids moléculaire (PM), un échantillon de cellules (cell) et d'effluent (Eff) sont déposés dans les trois premiers puits du gel. Les fractions prélevées et utilisées pour être concentrées et purifiées par chromatographie à exclusion de taille sont soulignées par une flèche rouge. **b.1** Chromatogramme : les encadrés orange (1) et bleu (2) représentent les différents pics observés sur le chromatogramme. **b.2** SDS-Page 12% suivant la purification chromatographie à exclusion de taille à l'aide d'une colonne S200 16/60. Le poids moléculaire (PM), un échantillon de la solution de protéines regroupées après la purification d'affinité (Aff.), du liquide passé à travers la membrane du concentrateur ayant une limite de 30 kDa (<30) et des fractions concentrées (CO) sont déposés dans les quatre premiers puits. Les fractions prélevées et utilisées pour être conservées sont soulignées par une flèche rouge.

En utilisant le même protocole de purification, les culots cellulaires provenant d'une production de 2 L sont décongelés et re-suspendus dans 100 mL de solution tamponnée de lyse A3/B3 (**tableau 6**, section Matériels et Méthodes ; §IV.2). Sans détergent et présentant une faible concentration de NaCl, cette solution tamponnée pourrait ne pas favoriser la solubilisation de la protéine. C'est pourquoi, seuls des petits culots de 2 L sont utilisés pour purifier la protéine dans ce tampon. De plus, dans le but d'augmenter les rendements et notamment la fixation de Gag sur la résine, l'effluent est passé deux fois sur la colonne. Sur le chromatogramme obtenu suite à la purification sur colonne d'affinité, un seul pic allant de la fraction F6 à la fraction H11, suivant le gradient d'imidazole est observable (**figure 36, a.1**). La pureté de l'échantillon après purification par colonne d'affinité est ensuite contrôlée à l'aide d'un SDS-PAGE 10% (**figure 36, a.2**). Ce gel couplé à l'analyse du chromatogramme, permet de sélectionner les fractions s'étendant de F2 à G10. Elles sont regroupées et concentrées. De manière comparable à ce qui est précédemment décrit, une deuxième étape de purification est effectuée en utilisant la colonne S200 16/60 dans le tampon A3 (**figure 36, b.1**). Sur le chromatogramme deux pics sont visibles, le premier allant des fractions B11 à B5 et le second s'étendant des fractions C1 à C9. La pureté de la protéine est contrôlée sur un gel dénaturant 10% (**figure 36, b.2**). Ainsi les fractions C2 à C5 sont regroupées avant d'être concentrées. Finalement, 1 mL de protéines purifiées est récupéré pour une concentration finale de 0,62 mg/mL. Dans une telle solution tamponnée, la solubilité de la protéine est relativement faible. En effet, la protéine précipite plutôt que de se concentrer et ceci est observé au-delà d'environ 0,7 mg/mL (mesure réalisée au Nanodrop).



**Figure 36. Purification de 6His-Gag dans le tampon A3**

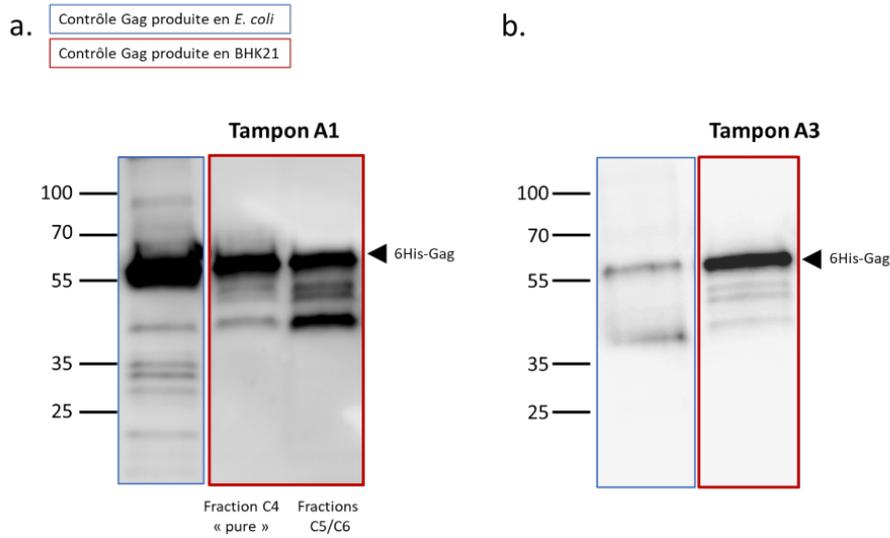
**a.1** Chromatogramme : l'encadré orange représente le pic observé. **a.2** Gel SDS de 10% : Le marqueur de poids moléculaire (PM), un échantillon de cellules (cell) et de cellules après sonication (ET) sont déposés dans les trois premiers puits. Ils sont suivis d'un aliquote après l'étape d'ultracentrifugation et des deux d'effluent (Eff<sub>n</sub>). Les fractions prélevées et utilisées pour être concentrées et purifiées par chromatographie à exclusion de taille sont soulignées par une flèche rouge. **b.1** Chromatogramme : les encadrés orange (1) et bleu (2) représentent les différents pics observés sur le chromatogramme. **b.2** Gel SDS-Page de 10% : suivant la purification chromatographique à exclusion de taille à l'aide d'une colonne S200 16/60. Les fractions prélevées, concentrées et conservées sont soulignées par une flèche rouge.

### 1.1.2 Contrôle de la séquence et des modifications post-traductionnelle de Gag

Une fois la protéine purifiée dans l'une ou l'autre solution tamponnées (A1 ou A3), des analyse par western blot (WB) et par spectrométrie de masse en tandem (MS/MS) sont réalisées pour contrôler la séquence des protéines et l'ensemble des PTM.

Tout d'abord, nous avons contrôlé la séquence de la protéine. Pour cela, une aliquote de protéines 6His-Gag concentrées dans le tampon A1 ou A3 est déposée sur un SDS-PAGE 12%. Après migration le gel peut être utilisé pour réaliser un transfert sur une membrane de polyvinylidene fluoride (PVDF) et servir à réaliser un WB ou être découpé dans le but de réaliser des analyses de MS/MS. Pour les WB, un contrôle est réalisé avec 1 µg de protéines recombinantes Gag produites en cellules de bactériennes (*E. Coli*) déjà caractérisées au laboratoire (Dubois et al., 2018b). Comme les deux types de protéines possèdent la même étiquette 6His, elles sont détectées à l'aide d'un anticorps dirigé contre celle-ci. Une bande largement majoritaire est visible légèrement au-dessus de 55 kDa pour les pistes contenant les protéines Gag recombinantes produites en cellules de mammifères (pistes 2 et 3, **figure 37, a**), ainsi que pour la protéine recombinante produite en cellules bactériennes (piste 1, **figure 37**). D'autres bandes apparaissent à un PM inférieur à 55 kDa pour l'une et l'autre protéine recombinante. Ces bandes sont

largement amplifiées dans les échantillons moins purs des protéines recombinantes produites en BHK21 purifiées, correspondant ici aux fractions C5/C6 après purification dans le tampon A1 (**figure 35,b2**, piste 3 pour le SDS-page et **figure 37, a.** pour le WB). Un résultat tout à fait similaire est obtenu pour la protéine 6His-Gag purifiée dans le tampon A3 (**figure 37, b**). Ces bandes correspondent probablement à des formes tronquées ou dégradées du précurseur.



**Figure 37. Analyse par Western blot (WB) des protéines 6His-Gag purifiées dans les tampons A1 et A3**

La protéine recombinante Gag-6His produite en cellules bactériennes est utilisée comme contrôle et déposée dans le premier puit (encadré bleu, gauche). La protéine recombinante 6His-Gag purifiée dans les tampons A1 ou A3 est déposée dans un SDS PAGE 12% et analysée par WB grâce à un anticorps dirigé contre l'étiquette 6His **a.** Deux fractions (C4 et C5/6) avec un degré de pureté différent sont comparés (encadré rouge, droite). **b.** 6His-Gag purifiée (encadré rouge, droite).

En parallèle des WB, la séquence en acides aminés est déterminée par MS/MS. Pour chaque séquençage les protéines sont analysées en parallèle à partir d'une préparation en solution et en gel. La dernière est aussi utilisée pour l'analyse de la bande de contaminants observée sur le WB. Dans un premier temps, la séquence de Gag est confirmée avec un recouvrement de 94% pour les préparations en solution et 98% pour la préparation en gel avec un nombre de spectres supérieur à 3200 dans les deux cas (**figure 38**). Deux régions ne sont pas détectées, ceci pourrait être dû à l'absence des sites de digestion par la trypsine et la chymotrypsine (séquence en noir sur la **figure 38**). Néanmoins, la présence de région N-terminale et notamment de l'étiquette 6His révélée par WB, ainsi que l'analyse de la taille de Gag sur SDS-PAGE montrent que dans ces fractions les protéines ne sont pas clivées.

En solution

3265 spectra on GAG-protein

Protein sequence coverage: 94%

Matched peptides shown in **bold red**.

1 MGARASVLSG **GELDKWEKIR** LRPGGKKQYK LKHIVWASRE LERFAVNFGL  
 51 LETSEGCQRQI LGQLQPSLQT GSEELRSLYN TIAVLYCVHQ RIDVKDTKEA  
 101 LDKIEEBQNK SKKGAQQAAA DTGNSQVSQ NYPIVQNLQG QMVHQAI SPR  
 151 TLNAWVKVVE EKAFSPEVIP MFSALSEGAT PQDLNMLNT VGGHQAAQM  
 201 LKETINEEAA EWDRLHPVHA GPIAPGQMRP PRGSDIAGTT STLQEQIGWM  
 251 THNPPIPVGE IYKRWIILGL NKIVMYSPT SILDIRQGEK EPFRDYVDRF  
 301 YKTLRAEQAS QEVQNMVET LLVQNPDC KTIKALGPG ATLEEMTAC  
 351 QGVGGPGHKA RVLAEAMSQV TNPATIMIQK GNFRNQRKT V KCFNCGKEGH  
 401 IAKNCRAPRK KGCWKCGKEG HQMDCTERQ ANFLGKIWPS HKGRPGNFLQ  
 451 SRPEPTAPPE ESFRFGETT TPSQKQEPID KELYPLASLR SLFGSDPSSQ

En gel

3361 spectra on GAG-protein

Protein sequence coverage: 98%

Matched peptides shown in **bold red**.

1 MGARASVLSG **GELDKWEKIR** LRPGGKKQYK LKHIVWASRE LERFAVNFGL  
 51 LETSEGCQRQI LGQLQPSLQT GSEELRSLYN TIAVLYCVHQ RIDVKDTKEA  
 101 LDKIEEBQNK SKKGAQQAAA DTGNSQVSQ NYPIVQNLQG QMVHQAI SPR  
 151 TLNAWVKVVE EKAFSPEVIP MFSALSEGAT PQDLNMLNT VGGHQAAQM  
 201 LKETINEEAA EWDRLHPVHA GPIAPGQMRP PRGSDIAGTT STLQEQIGWM  
 251 THNPPIPVGE IYKRWIILGL NKIVMYSPT SILDIRQGEK EPFRDYVDRF  
 301 YKTLRAEQAS QEVQNMVET LLVQNPDC KTIKALGPG ATLEEMTAC  
 351 QGVGGPGHKA RVLAEAMSQV TNPATIMIQK GNFRNQRKT V KCFNCGKEGH  
 401 IAKNCRAPRK KGCWKCGKEG HQMDCTERQ ANFLGKIWPS HKGRPGNFLQ  
 451 SRPEPTAPPE ESFRFGETT TPSQKQEPID KELYPLASLR SLFGSDPSSQ

Figure 38. Couverture de la séquence de 6His-Gag par des spectres MS/MS pour les préparations en solution et en gel

Ensuite, l'analyse de MS/MS est réalisée sur les bandes qui présente un PM inférieur à 59 kDa, qui pourrait correspondre à des formes tronquées de la protéine, appelés ici les « contaminants ». Pour cela, les fractions moins pures (C5 et C6) sont déposées sur un gel SDS 12%. A l'issue de sa migration, la bande de plus petite taille (bande basse ; en orange **figure 39**) est découpée pour être analysée en parallèle de celle correspondant à 6His-Gag complète (bande haute ; en vert **figure 39**). Afin d'éviter les contaminations croisées d'un échantillon à l'autre, l'échantillon « contaminants » est injecté en premier. La couverture des séquences par résidu, représentée par le nombre de spectres MS/MS, est superposée entre les échantillons de Gag intacte et les contaminants (comme indiqué dans la **figure 39**). L'absence de spectres observée au niveau des derniers résidus (à partir de la position 289) de la séquence protéique pour la bande contaminante confirme qu'il s'agit bien des formes tronquées au niveau de l'extrémité C-terminale de Gag.

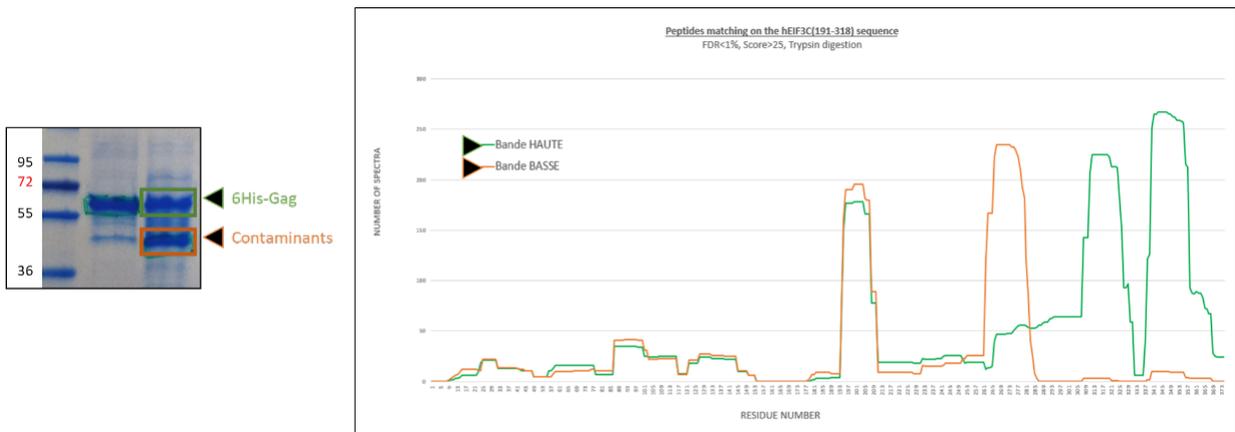


Figure 39. Couverture de séquences des bandes protéines (bande haute dans le gel, encadré en vert) et contaminants (bande basse dans le gel, encadré en orange) pour chaque résidu

Finalement, la MS/MS sert également à identifier les PTM de Gag à partir des échantillons en solution et en gel SDS, issus de quatre productions et/ou purification de protéines en cellules de mammifères indépendantes. Pour cette étude, une intention particulière est portée vers les phosphorylations et les ubiquitinations qui sont décrites comme ayant un rôle crucial pour la fonctionnalité de Gag (**annexe 1**). Toutes les modifications obtenues en fonction du type d'échantillon (en gel ou en solution) et de l'enzyme de digestion utilisées (trypsine ou chymotrypsine) sont répertoriées dans le **tableau 7** où les modifications retrouvées dans au moins deux analyses sont représentées.

Enzymes		Trypsine				Chymotrypsine
En gel	<b>Recouvrement de séquences</b>	> 90%	98%	93%	98%	> 90%
	<b>Phosphorylations</b>	S278, S495	S148, T151, S278	S77, S148, T151, S278, S495, S498	S148 T151 S278, T318 T342 S495	S495, S498
	<b>Ubiquitinations</b>	C350, K359	K18, K32, C57, C87, K331, C350, K359, K380, K481	K15, K18, C57, C87, K272, C330, K331, C350, K359, K481	K359	K331, T348
En solution	<b>Recouvrement de Séquence</b>	> 90%	95%	93%	94%	> 90%
	<b>Phosphorylations</b>	T122, S129, S148, S278, T470, S473, S498	S148, S498	S148, T151, S278, T318, T342, S498	S77, S148, T151, S278, S495, S498	T122, S129, S148, S278, T470, S473, S498
	<b>Ubiquitinations</b>	K331, K359	/	K359	K15, K18, C57, C87, K272, C330, K331, C350, K359, S368, K481	K331, K359

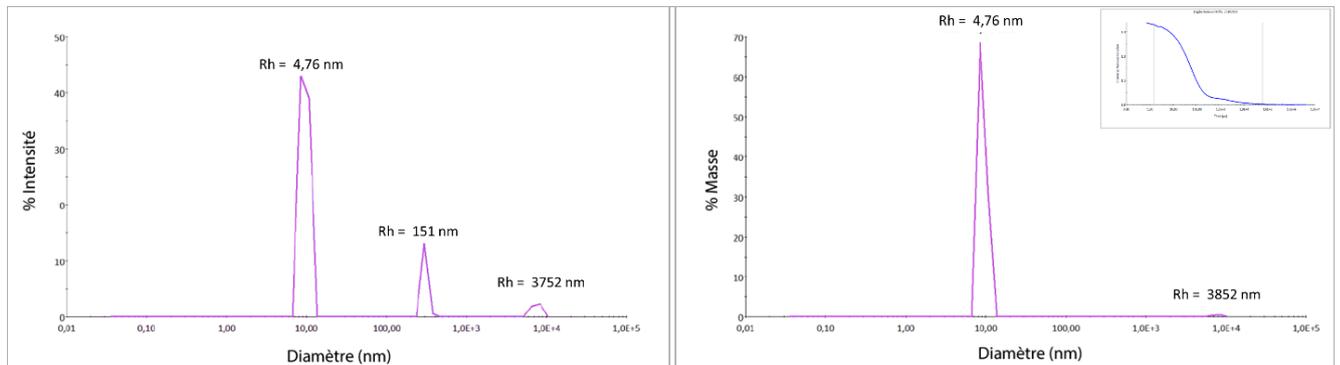
**Tableau 7. Les PTM identifiées par MS/MS à partir de quatre différentes production et/ou purification de protéines 6His-Gag purifiée dans le tampon A1**

### 1.1.3 Analyse par des méthodes de diffusion de la lumière des protéines 6His-Gag produites en cellules de mammifères dans les solutions tamponnées A1 et A3

#### 1.1.3.a Analyse par DLS des protéines Gag produites en cellules de mammifères dans les tampons A1 et A3

Ces analyses permettent de caractériser la solubilité et la capacité à multimériser des protéines Gag purifiées dans les deux tampons de purifications (A1 et A3) dans le but de déterminer la qualité des protéines avant leur utilisation. Ainsi, 6His-Gag produites en cellules de mammifères et purifiées dans les tampons A1 et A3 et produites en cellules bactériennes dans la solution tamponnée contenant 1 M NaCl, 50 mM Tris pH 8,5% glycérol et 2,5 mM DTT, sont analysées et comparées en suivant le protocole décrit dans la partie Matériels et Méthodes (**figure 27**, §VI.1). Les analyses des trois protéines présentent un chromatogramme DLS relativement différent. La distribution en intensité est bimodale pour les protéines Gag produites en *E. coli* (**figure 40**) et en BHK21 purifiées dans le tampon A3 (**figure 41**). Elle est multimodale pour la protéine produite en BHK21 et purifiée dans le tampon A1 (**figure 42**). Les pics sont relativement mono-disperses avec un indice de polydispersité de 16,5 et 12,1 pour les protéines recombinantes Gag produites en cellules de mammifères et purifiées dans les tampons A1 et A3 respectivement, alors qu'on observe un indice de 28,6 pour la protéine recombinante produite en cellules de bactéries.

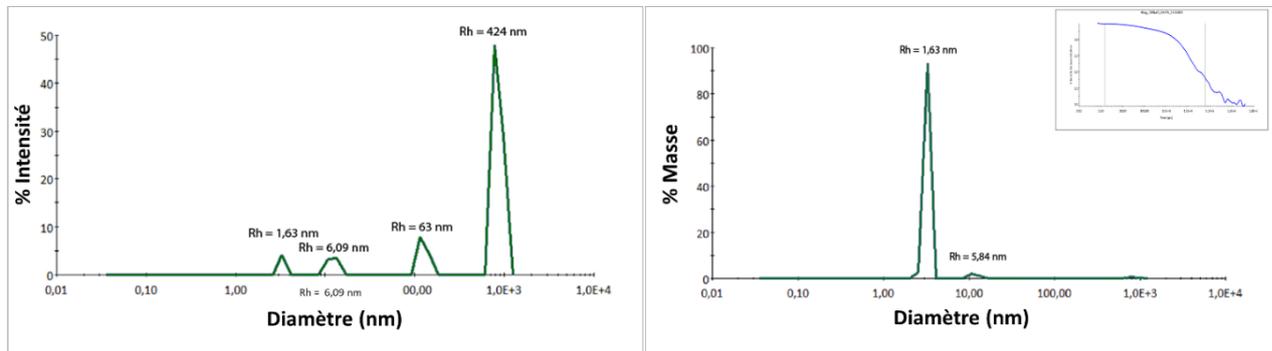
Pour commencer, l'analyse DLS des protéines Gag produites en cellules bactériennes, permet d'obtenir un chromatogramme composé d'un pic largement majoritaire correspondant à des molécules ayant un Rh de 4,76 nm dans les deux représentations (**figure 40**). Des pics correspondant à des molécules ayant un Rh de plus grande taille sont présents en proportions minoritaires dans la distribution en intensité (151 nm et 3752 nm, **figure 40**). En appliquant la loi de Stokes-Einstein, le PM associé aux molécules représentée par le premier pic, serait de 126 kDa, soit un dimère de Gag, alors que les autres pic correspondent à des multimères de grande taille.



**Figure 40. Analyse DLS des protéines 6His-Gag produites en cellules bactériennes**

Représentation des distributions en intensité (à gauche) et en masse (à droite) en fonction des diamètres des protéines en solution. Le Rh est indiqué au-dessus du pic correspondant. Chaque graphique correspond à la moyenne de dix mesures. Le corrélogramme est représenté en bleu dans l'encadré en haut à droite.

Ensuite, les protéines recombinantes Gag produites en BHK21 et purifiées dans le tampon A1 analysées par DLS ont un profil à 4 pics, dans la représentation en intensité (**figure 41**). La majorité du signal est due à la présence de macromolécules de grandes taille (424 nm). Alors que dans la représentation en masse, la majorité du signal correspond à des molécules ayant un Rh de 1,63 nm. Contrairement aux autres protéines de l'étude, celle-ci est purifiée dans une solution tamponnée contenant du détergent, le CHAPS, au-dessus de sa CMC qui est de 6 mM (Rodi, Bocco Gianello et Gennaro 2014). A l'issue de l'analyse DLS, un deuxième pic est présent dans les représentations du signal en masse et en intensité et il correspond à une molécule d'un Rh de 6,09 nm et correspondrait à un tétramère de Gag.

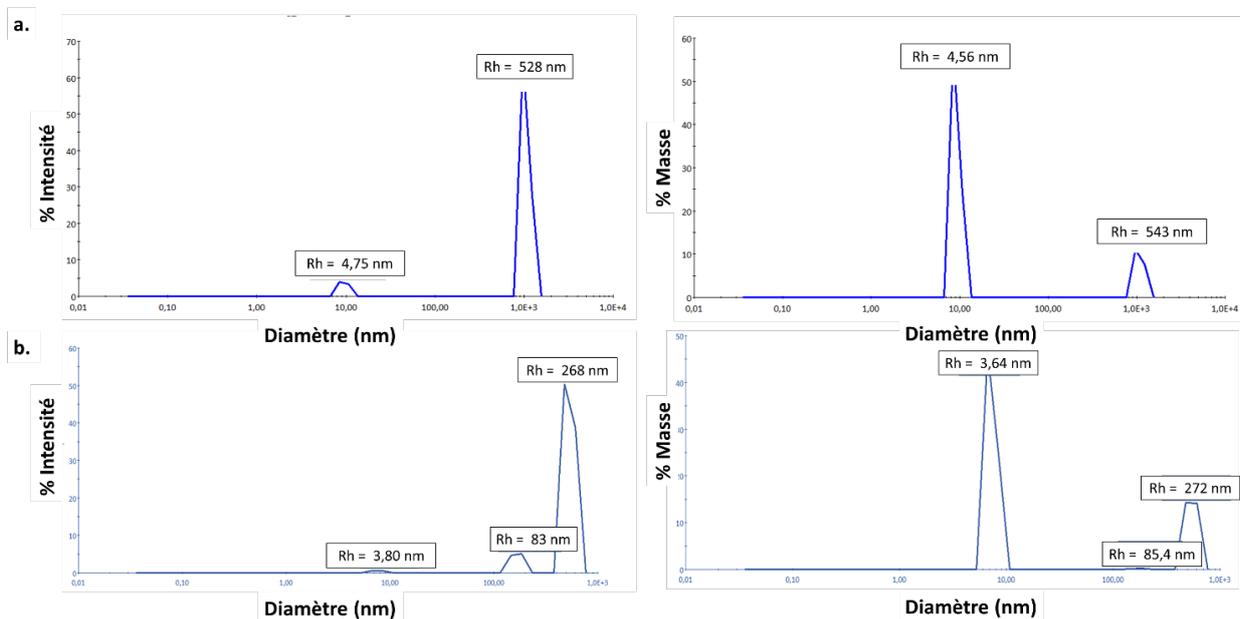


**Figure 41. Analyse DLS des protéines 6His-Gag produites en BHK21 et purifiées dans le tampon A1**

Représentation des distributions en intensité (à gauche) et en masse (à droite) en fonction des diamètres des protéines en solution. Le Rh est indiqué au-dessus du pic correspondant. Chaque graphique correspond à la moyenne de dix mesures. Le corrélogramme est représenté en bleu dans l'encadré en haut à droite.

Finalement, les analyses DLS réalisées sur la protéine 6His-Gag purifiée dans le tampon A3 donnent deux types de profils différents. Aucune corrélation avec le temps de décongélation, la température de conservation ou les tubes utilisés pouvant expliquer une différence de profils n'est relevée. D'une part, les distributions en intensité et en masse représentent un signal avec deux pics correspondant à des molécules ayant un Rh de 4,79 nm et 528 nm. Le premier pic est minoritaire en intensité (7% du signal) mais majoritaire en masse (81% du signal ; **figure 42, a**). En appliquant la loi de Stokes-Einstein les PM estimés pour ces pics correspondent respectivement à un dimère de Gag (130 kDa) et à un très grand multimère voir de l'agrégation (7901460 kDa). De plus, cette tendance à l'agrégation est visible au niveau du chromatogramme (**figure 42, a**). D'autre part, le second profil est composé de 3 pics correspondant à des molécules ayant des Rh de 3,8 nm, 83,3 nm et 268 nm (**figure 42, b**). Le premier pic est minoritaire en signal d'intensité (1,2% du signal) mais majoritaire en signal de masse (71% du signal ; **figure 42, b**). En appliquant la loi de Stokes-Einstein le PM estimé pour chacun des pics correspond à un monomère de Gag (76 kDa) et à des très grands multimères voir à de l'agrégation (104940 kDa - 1612490 kDa). De manière

parfaitement équivalente à ce qui est observé précédemment, la tendance à l'agrégation est visible au niveau du chromatogramme.



**Figure 42. Analyse DLS des protéines 6His-Gag produites en BHK21 et purifiées dans le tampon A3**

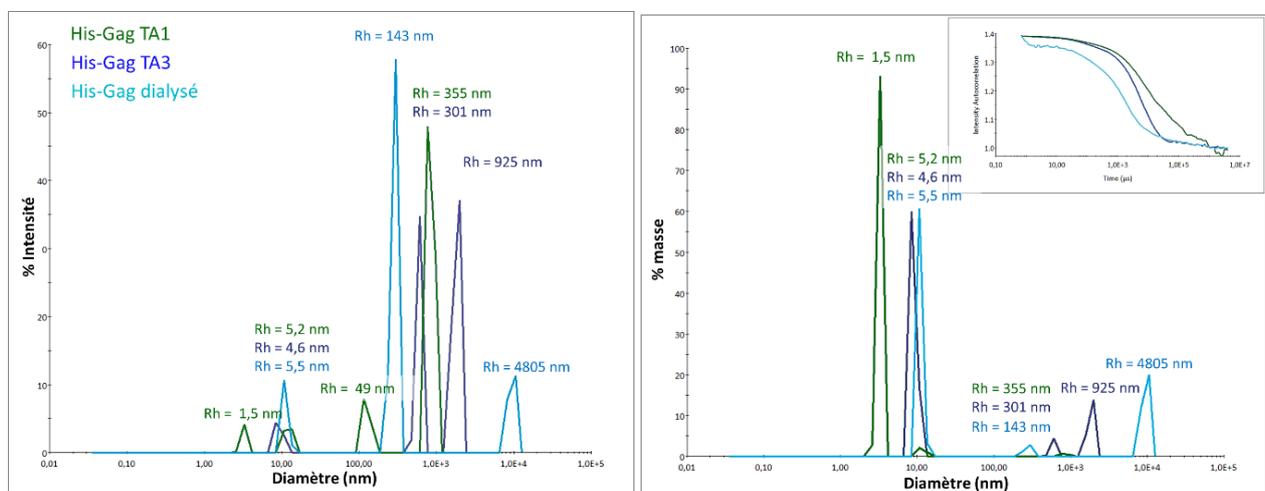
Pour ces échantillons deux profils sont observés **a.** et **b.** Représentation des distributions en intensité (à gauche) et en masse (à droite) en fonction des diamètres des protéines en solution. Le Rh est indiqué au-dessus du pic correspondant. Chaque graphique correspond à la moyenne de dix mesures. Les corrélogrammes sont représentés en bleu dans l'encadré en haut à droite.

Ainsi, peu importe la méthode ou le tampon de purification utilisés lors de la production de la protéine Gag, cette dernière forme des agrégats et présente des Rh correspondant à des monomères ou à des petits oligomères. Les valeurs de Rh déterminées expérimentalement par DLS sont récapitulées **tableau 8** (section Résultats ; §I.1.5).

### I.1.3.b Analyse par SEC-MALS des protéines Gag produites en cellules de mammifères dans les solutions tamponnées A1 et A3

Pour contrôler la qualité des protéines recombinantes produites en cellules de mammifères et conservées dans les solutions tamponnées A1 et A3, des analyses de SEC-MALS sont également réalisées. Contrairement aux résultats de DLS présentés ci-dessus, la protéine 6His-Gag conservée dans le tampon A3 n'est pas purifiée directement dans ce tampon, mais elle est dialysée (comme décrit dans la section Matériels et Méthodes ; §VII.3) à partir d'un aliquote purifiée dans le tampon A1. Cette étape nécessite des mises au point permettant l'optimisation des conditions de dialyse afin d'empêcher l'agrégation de la protéine. Pour cela, de nouvelles analyses de DLS sont réalisées. Les profils DLS en intensité et en masse sont superposés à ceux de la protéine purifiée dans le tampon A1 et ceux de la protéine purifiée dans le tampon A3 (**figure 43**). Ainsi, l'analyse DLS de la protéine dialysée présente un profil multimodal, comparable à celui des protéines purifiées dans le tampon A1 ou A3. En effet le chromatogramme

issue de l'analyse de la protéine dialysée présente 3 pics en intensité et en masse. Comme pour la protéine purifiée dans le tampon A3, le pic majoritaire en masse correspond à une molécule ayant un Rh de 5,5 nm identifié comme étant un dimère de Gag. De plus, par rapport à la protéine purifiée dans le tampon A1, le premier pic représentant des molécules ayant un Rh de 1,5 nm n'apparaît plus sur le chromatogramme de la protéine dialysée, suggérant fortement que la concentration de CHAPS diminue efficacement lors de l'étape de dialyse. En outre, les pics relatifs aux protéines présentant de grands Rh ne se superposent pas et présentent une faible reproductibilité. Ils pourraient ainsi correspondre à des oligomères de grande taille ou probablement à des agrégats comme observés dans le corrélogramme. La protéine Gag purifiée dans le tampon A1 et dialysée dans le tampon A3 sera utilisée pour réaliser son analyse biophysique.

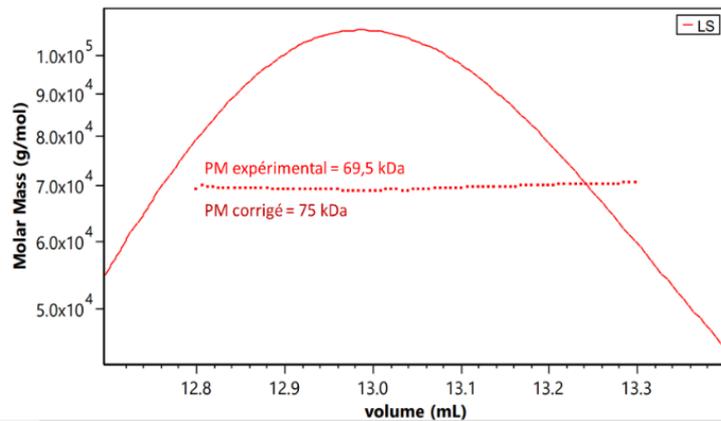


**Figure 43. Analyse DLS des protéines 6His-Gag purifiées dans le tampon A1 et A3 en comparaison avec le profil d'analyse DLS des protéines 6His-Gag purifiées dans le tampon A1 et dialysées contre le tampon A3**

La protéine purifiée dans le tampon A1 est représentée en vert, la protéine purifiée dans le tampon A3 est représentée en bleu foncé et la protéine purifiée dans le tampon A1 et dialysée contre le tampon A3 est représentée en bleu clair. Les distributions en intensité (à gauche) et en masse (à droite) en fonction des diamètres des particules en solution sont représentées pour chacune des protéines testées. Le Rh de chaque protéine est indiqué au-dessus du pic correspondant. Chaque graphique correspond à la moyenne de dix mesures.

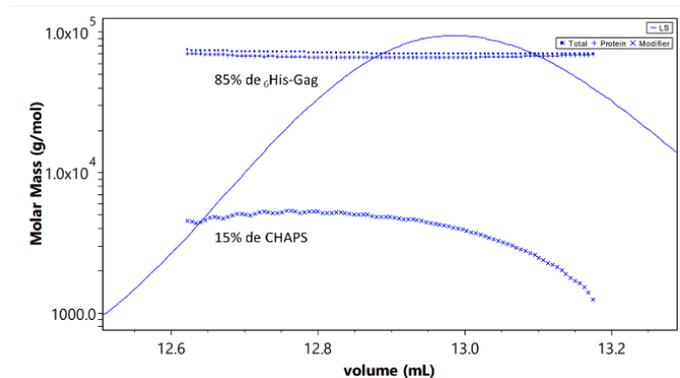
Pour confirmer l'état oligomérique et le PM de l'espèce majoritaire en solution après la purification de Gag dans le tampon A1, des analyses de SEC-MALS sont donc réalisées. Le pic obtenu sur le chromatogramme de l'analyse SEC-MALS est symétrique, la protéine est majoritairement monodisperse. Dans un premier temps, le modèle classique est appliqué. Il permet de déterminer le PM expérimental à partir des mesures d'intensité,  $dn/dc$  et concentration (section Matériels et méthodes ; §IX.2). Le PM expérimental de 6His-Gag dans le tampon A1 serait de 69,5 kDa (**figure 44**). Cependant, ce PM est déterminé en utilisant le  $dn/dc$  caractéristique d'une protéine (0,185). Or, la protéine recombinante 6His-Gag est purifiée dans une solution tamponnée composée de CHAPS dont la concentration est supérieure à la CMC. Ainsi, en prenant en compte la masse théorique de la protéine 6His-Gag (59252 Da), il est possible de supposer que l'échantillon est composé d'environ 85% de protéine et de 15% de

CHAPS. En corrigeant la valeur de  $dn/dc$  en prenant en compte le  $dn/dc$  du CHAPS (0,1323) et la distribution du détergent et de la protéine dans l'échantillon, le PM apparent correspondant serait, dans ce cas, de 75 kDa.



**Figure 44. Analyse SEC-MALS de 6His-Gag purifiée dans le tampon A1 après traitement avec le modèle classique**  
 Sans appliquer de correction le PM expérimental est de 69,5 kDa. En corrigeant avec la valeur  $dn/dc$  avec celle correspondant à 85% de protéine Gag et 15% de CHAPS (0,177), le PM expérimental est obtenu de 75 kDa.

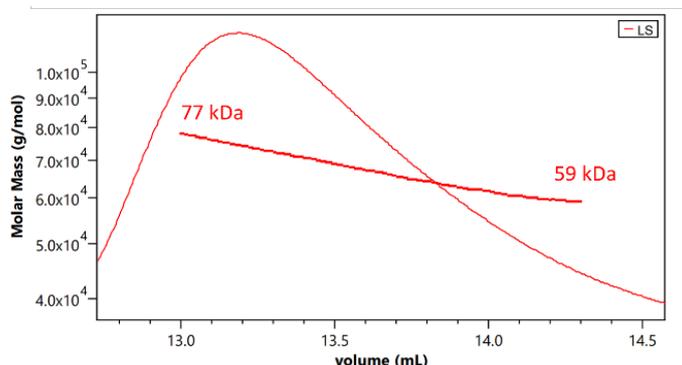
Pour contrôler ces mesures, un second modèle est appliqué le *Protein Conjugate Analysis* (section Matériel et méthodes ; §IX.2). Ce dernier modèle permet de calculer le PM expérimental en prenant en compte la valeur d'extinction molaire ( $1,087 \text{ mL}/[\text{mg}\cdot\text{cm}^{-1}]$ ) et les  $dn/dc$  de la protéine (0,185) et du CHAPS (0,1323 ; **figure 45**). Dans ce cas, le PM du conjugué est de  $74,53 \pm 0,017 \text{ kDa}$ , comprenant  $70,02 \pm 0,016 \text{ kDa}$  pour la protéine et  $4,51 \pm 0,17 \text{ kDa}$  pour le CHAPS. Les deux analyses donnent des résultats comparables.



**Figure 45. Analyse SEC-MALS de 6His-Gag purifiée dans le tampon A1 après traitement avec le model Protein Conjugate Analysis**

Ensuite, pour confirmer l'état oligomérique et le PM de l'espèce majoritaire en solution après la purification de Gag dans le tampon A1 et dialyse dans le tampon A3 de nouvelles analyses de SEC-MALS sont réalisées. Le profil de la protéine Gag purifiée dans le tampon A1 et dialysée avec le tampon A3 correspond à un pic asymétrique, ce qui

est synonyme de polydispersité. En conséquence la masse molaire calculée relative aux espèces en solution, varie de 77 kDa à 59 kDa (**figure 46**).

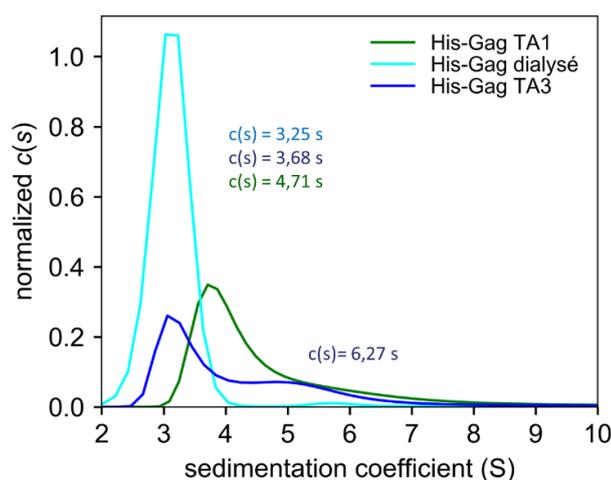


**Figure 46. Analyse SEC-MALS de 6His-Gag purifiée dans le tampon A1 et dialysée contre le tampon A3**

Ainsi la protéine Gag purifiée dans le tampon A1 serait recouverte de CHAPS à hauteur de 85%. Lorsqu'elle est dialysée contre le tampon A3, la protéine présenterait différents états oligomériques. Les PM calculée par SEC-MALS sont récapitulés **tableau 8** (section Résultats ; §1.1.5).

#### 1.1.4 Analyse de la protéine Gag par AUC

Pour réaliser une caractérisation de la protéine par une autre méthode biophysique n'utilisant pas la diffusion de la lumière, des analyses d'AUC de vitesse de sédimentation sont réalisées sur la protéine 6His-Gag purifiée dans les tampons A1 et A3, ainsi que sur la protéine Gag purifiée dans le tampon A1 et dialysée dans le tampon A3. Les coefficients de sédimentation (C(s)) déterminées pour les solutions de 6His-Gag dans les différentes solutions tamponnées présentent des similitudes (**figure 47**). L'analyse AUC des protéines 6His-Gag purifiées dans les tampons A1 et A3 présentent deux limites de sédimentation qui se résolvent en deux pics (**figure 47**). Elles correspondent au coefficient de sédimentation c(s) de différents états oligomériques de la protéine. Les valeurs de PM expérimentaux calculés et de Rh correspondant sont récapitulées dans le **tableau 8** et correspondraient à un monomère et un dimère de la protéine. De plus, le pic correspondant au dimère n'est plus retrouvé suite à l'étape de dialyse. Le c(s) de la protéine dans le tampon A1 est plus grand que celui de la protéine purifiée dans le tampon A3 et de la protéine dialysée contre le tampon A3. Le CHAPS présent dans la solution tamponnée pourrait expliquer cette différence. Les détails des analyses et des valeurs se trouvent en **annexes 2**.

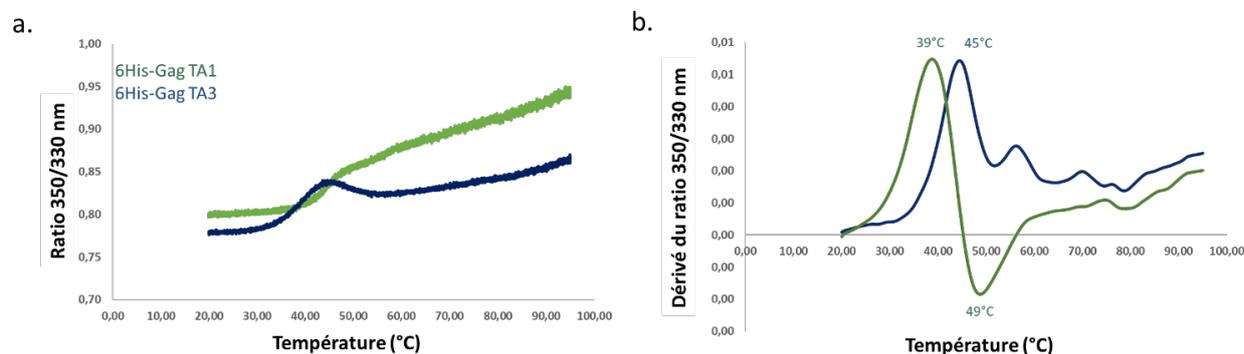


**Figure 47. Analyse des vitesses de sédimentation de la protéine recombinante 6His-Gag dans les différentes solutions tamponnées**

L'analyse de la distribution du coefficient de sédimentation  $C_s$  en continu est réalisée à partir des données d'absorbance. La courbe bleu foncé représente la protéine purifiée dans le tampon A3, la courbe verte, la protéine purifiée dans le tampon A1 et finalement la courbe bleu clair, la protéine purifiée dans le tampon A1 et dialysée contre le tampon A3. Les graphiques sont réalisés en utilisant la plateforme Gussi (Brautigam, 2015).

#### 1.1.5 Analyse de la protéine 6His-Gag par NanoDSF

Le signal fluorescent intrinsèque des protéines Gag, qui contiennent 9 résidus Trp repartis dans ses différents domaines, est exploité afin de déterminer les températures de dénaturation (TM) des protéines par NanoDSF (**figure 48**). Cette analyse est réalisée à partir de différentes concentrations de Gag dans les tampons A1 et A3. Les profils obtenus sont différents entre les deux solutions tamponnées. Ainsi, la moyenne de deux analyses permet d'estimer des TM dans le tampon A1 égale à  $36,5 \pm 3^\circ\text{C}$  et une deuxième étape de dénaturation aurait lieu à  $46,5 \pm 3,2^\circ\text{C}$ , tandis que pour la protéine purifiée dans le tampon A3 la TM estimée est de  $44,1 \pm 0,6^\circ\text{C}$ . Ces valeurs sont répertoriées dans le tableau ci-dessous.



**Figure 48. Détermination des températures de dénaturation (TM) de la protéine recombinante 6His-Gag les solutions tamponnées A1 et A3**

L'analyse de la distribution des ratios 350 nm/330 nm et de leur dérivé pour la protéine 6His-Gag purifiée dans le tampon A1 (courbe verte) et dans le tampon A3 (courbe bleue) en fonction de la température.

Protéine	Caractéristiques théoriques		DLS	SEC-MALS	AUC			NanoDSF	
	Tampon	PM (kDa)	Rh (nm)	PM (kDa)	S	Rh (nm)	PM (kDa)	f/f0	TM (°C)
6His-Gag	A1	59252	6,09	75	4,71	3,18	72227	1,10	36,5 et 46,5
6His-Gag	A3	59252	3,80 4,75	x	3,68	3,18	63769	1,20	44,1
6His-Gag	Dialysé A1 → A3	59252	5,50	[59 ; 77]	3,25	3,58	48774	1,48	

**Tableau 8. Données biophysiques de la protéine recombinante 6His-Gag produite en BHK21 purifiée dans les tampons A1 et A3**

Pour conclure, l'objectif de ces résultats est de déterminer les conditions idéales de solubilisation de la protéine Gag compatibles avec les études structurales par CryoEM. Ainsi, dans ce but et pour caractériser des complexes protéines/acide nucléiques, il est préférable de diminuer la concentration de NaCl ainsi que celle en détergent. Pour ces raisons, malgré les problèmes rencontrés et décrits dans la session précédente, la solution tamponnée A3 qui présente une plus faible force ionique et ne contient pas de détergent, est privilégiée pour la suite de ce travail. Ainsi les protéines qui sont sélectionnées pour l'analyse des complexes seront soit directement purifiées dans le tampon A3, soit purifiées dans le tampon A1 puis dialysée dans le tampon A3.

**Résumé :** Les conditions de purification, les séquences et les PTM sont caractérisées pour la protéine 6His-Gag produites en cellules de mammifères. L'état oligomérique, le coefficient de sédimentation et la température de dénaturation du précurseur sont déterminés pour les deux solutions tamponnée A1 et A3.



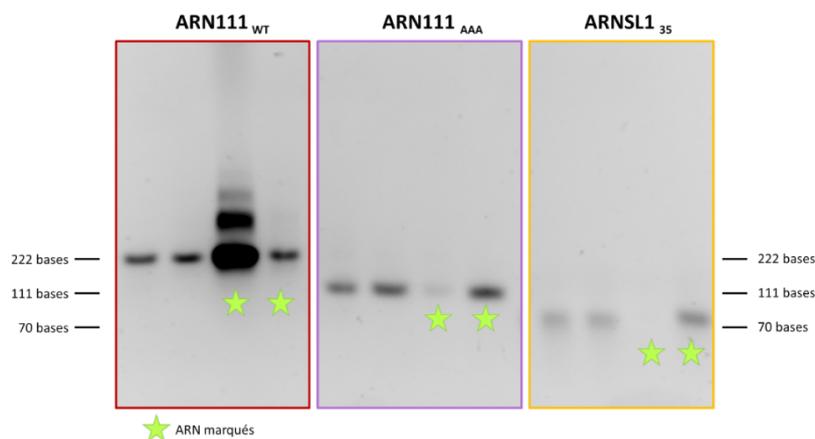
## 1.2. Analyse de l'interaction de 6His-Gag avec la région 5'UTR de l'ARNg viral *in vitro*

La deuxième étape de ce travail consiste à déterminer les conditions optimales pour la formation *in vitro* des complexes entre la protéine Gag et les fragments d'ARNg, ARN111<sub>WT</sub>, ou ARNSL1<sub>35</sub>. Ces interactions sont analysées par différentes méthodes biophysiques.

### 1.2.1 Caractérisation de la dimérisation de l'ARNg

L'une des étapes cruciales dans la formation des complexes entre la protéine Gag et l'ARNg est la dimérisation de l'ARN (Bernacchi et al., 2017). Ainsi des analyses de migration en gel d'agarose et d'AUC sont réalisées pour contrôler la dimérisation des ARN111<sub>WT</sub> ou ARNSL1<sub>35</sub> utilisés dans cette étude. Le contrôle est l'ARN111<sub>AAA</sub> incapable de dimériser car il muté au niveau du site de dimérisation de l'ARNg.

D'une part, un gel TBM agarose 1% est utilisé pour contrôler la dimérisation des ARNg non marqués et marqués par un chromophore de type fluorescéine à leur extrémité 3' (section Matériels et Méthodes ; §III.2). Les ARN sont repliés (section Matériel et Méthodes ; §IV.2) et repris dans une solution tamponnée composée de 50 mM HEPES pH 7,5, 400 mM NaCl et 5 mM MgCl<sub>2</sub> (tampon A3b) A l'issue de la dimérisation, une aliquote de chacun des ARN, l'ARN111<sub>WT</sub>, l'ARN111<sub>AAA</sub> et l'ARNSL1<sub>35</sub> est mélangée avec du bleu glycérol et chargés dans les puits du gel pour une migration de 3 heures à 4°C (**figure 49**). Les l'ARN111<sub>WT</sub> et l'ARNSL1<sub>35</sub> sont dimériques qu'ils soient marqués ou non marqués. L'ARN111<sub>AAA</sub> est quant à lui, monomérique.

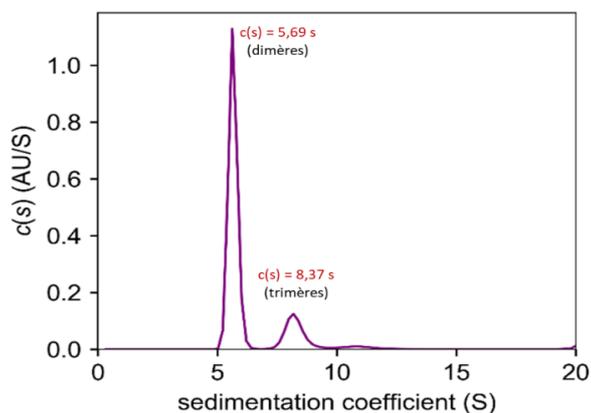


**Figure 49. Analyse par gel agarose Tris Borate magnésium (TBM) 1% de la dimérisation des ARNg marqués et non marqués**  
Les puits correspondants à la migration d'ARN marqués sont représentés par une étoile verte. Les ARN111<sub>WT</sub> sont encadrés en rouge (gauche), les ARN111<sub>AAA</sub> sont encadrés en mauve (milieu) et finalement les ARN SL1<sub>35</sub> sont encadrés en jaune (droite).

D'autre part, l'AUC est utilisée pour déterminer le coefficient de sédimentation permettant de déterminer les proportions de dimères d'ARN par rapport aux monomères (**tableau 9**). Pour cela l'ARN111<sub>WT</sub> est dilué pour obtenir une concentration finale de dimère de 1  $\mu$ M dans le tampon A1 (**figure 50**) ou dans le tampon A3 (**figure 51**). Il est

ensuite replié dans l'une ou l'autre solution tamponnée (**tableau 6**, section Matériel et Méthodes ; §IV.2) avant d'être déposé dans les cellules de mesures. Le PM théorique de l'ARN111<sub>WT</sub> dimérique est de 73260 Da.

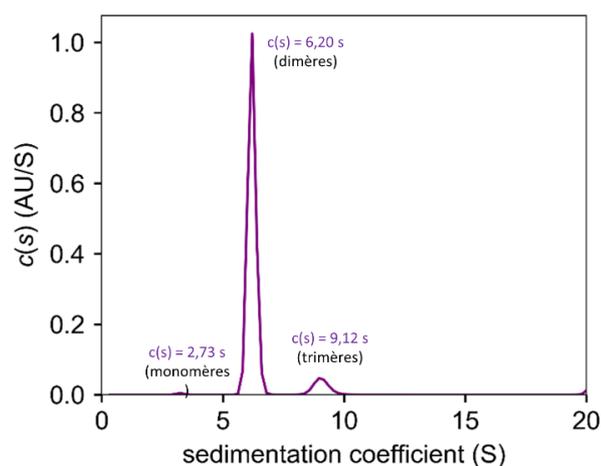
D'une part, l'analyse des résultats d'AUC permet de déterminer que les ARN dilués et repliés dans le tampon A1 ne contient plus de monomère (**figure 50**). En effet la large majorité du signal correspond à une population ayant un  $c(s)$  de 5,69 s. Le calcul du PM expérimental permet de déterminer que ce pic correspond à des ARN dimériques de 75752 Da ce qui est en très bon accord avec les PM calculés pour le dimère d'ARN. De plus, une plus faible proportion d'ARN semble capable de former des populations à plus haut PM, présentant un  $C(s)$  de 8,37 s. Le calcul de PM expérimental associé à ce pic (134855 Da) pourrait correspondre à des formes multimériques de l'ARN.



**Figure 50. Représentation des analyses de vitesse de sédimentation de l'ARN111WT dans le tampon A1**

*L'analyse de la distribution du coefficient de sédimentation  $C_s$  en continu est réalisée à partir des données d'absorbance. Les graphiques sont réalisés en utilisant la plateforme Gussi (Brautigam, 2015).*

D'autre part, l'analyse AUC de l'ARN111<sub>WT</sub> replié dans le tampon A3 permet d'identifier 3 populations d'ARN (**figure 51**). La première population présente un  $c(s)$  de 2,73 s. Le calcul du PM expérimental indique qu'il correspond à 39300 Da qui correspond à la valeur d'un monomère d'ARN111 (36630). Cette population ne représente que 0,37% du signal. Le pic majoritaire, correspondant à 91% du signal, représente un ARN ayant un  $c(s)$  de 6,20s. Le PM associé est de 101957 Da, s'approchant à celui d'un dimère d'ARNg. Finalement, comme pour le tampon A1, une dernière population présente un  $c(s)$  de 9,12 s associé à un PM de 182157 Da. Cette population pourrait correspondre à des multimères d'ARN111<sub>WT</sub>.



**Figure 51. Représentation des analyses de vitesse de sédimentation SV de l'ARN111 WT dans le tampon A1**

L'analyse de la distribution du coefficient de sédimentation  $C_s$  en continu est réalisée à partir des données d'absorbance. Les graphiques sont réalisés en utilisant la plateforme Gussi (Brautigam, 2015).

Ainsi, les analyses d'AUC mettent en évidence que le fragment d'ARN111<sub>WT</sub> multimérise de façon majoritaire dans les solutions tamponnées A1 et A3, bien que l'accord entre le PM théorique du dimère et le PM estimé soit meilleur pour la solution A1 (**tableau 9**). Les complexes entre 6His-Gag et l'ARN111<sub>WT</sub> seront formés dans ces solutions.

	Pics	Coefficient de sédimentation (S)	%	PM (Da)	Rh (nm)
<b>Tampon A1</b>					
<b>ARN111<sub>WT</sub></b>	N°1	5,69	77	75752	3,76
	N°2	8,37	16	134855	4,56
	N°3	11,01	2,48	206697	5,23
<b>Tampon A3</b>					
<b>ARN111<sub>WT</sub></b>	N°1	2,73	0,37	39300	2,70
	N°2	6,20	91	101957	4,06
	N°3	9,12	7,4	182157	4,93

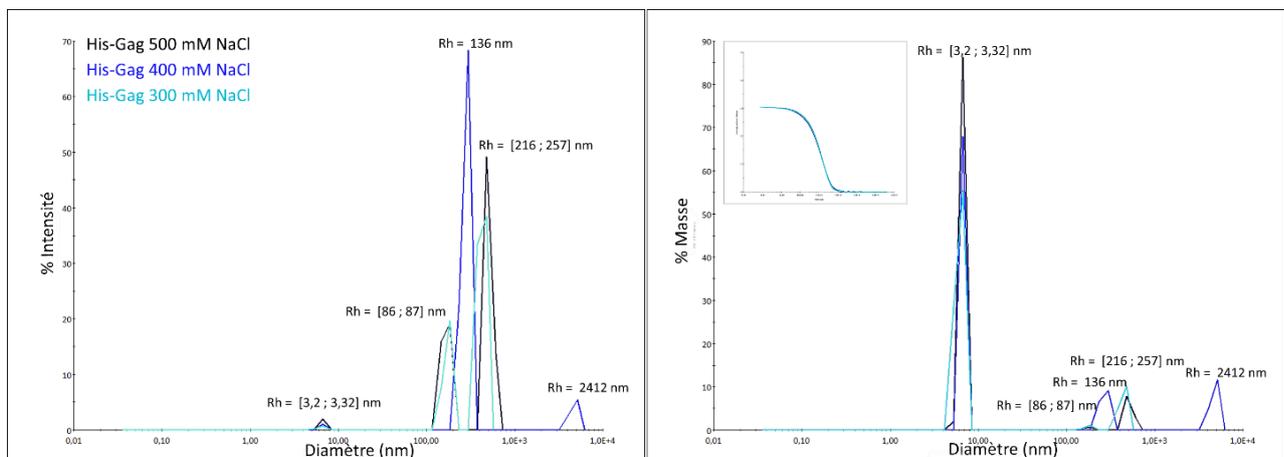
**Tableau 9. Valeurs de  $C_s$  mesurées par AUC pour l'ARN111WT dans les tampons A1 et A3**

Les PM et les Rh associés à chaque pic d'AUC sont calculés à partir de la mesure de coefficient de sédimentation.

### 1.2.2 Caractérisation de l'interaction entre 6His-Gag et l'ARN111<sub>WT</sub> par DLS

Les complexes entre la protéine 6His-Gag et l'ARN 111<sub>WT</sub> sont réalisées par mélange à température ambiante sans incubation. Afin de maximiser les probabilités de formation de complexes entre la protéine 6His-Gag et l'ARN111<sub>WT</sub>, différentes concentrations de NaCl sont testées pour former les complexes (500 mM, A3a ; 400 mM, A3b et 350 mM, A3c). Pour cela, les protéines purifiées dans le tampon A3 sont ensuite diluées dans par un tampon A3 sans NaCl, de manière à atteindre les concentrations de NaCl souhaitées. Les complexes sont formés avec une stœchiométrie telle que 6 molécules de Gag sont présentes pour un dimère d'ARN (**tableau 10**).

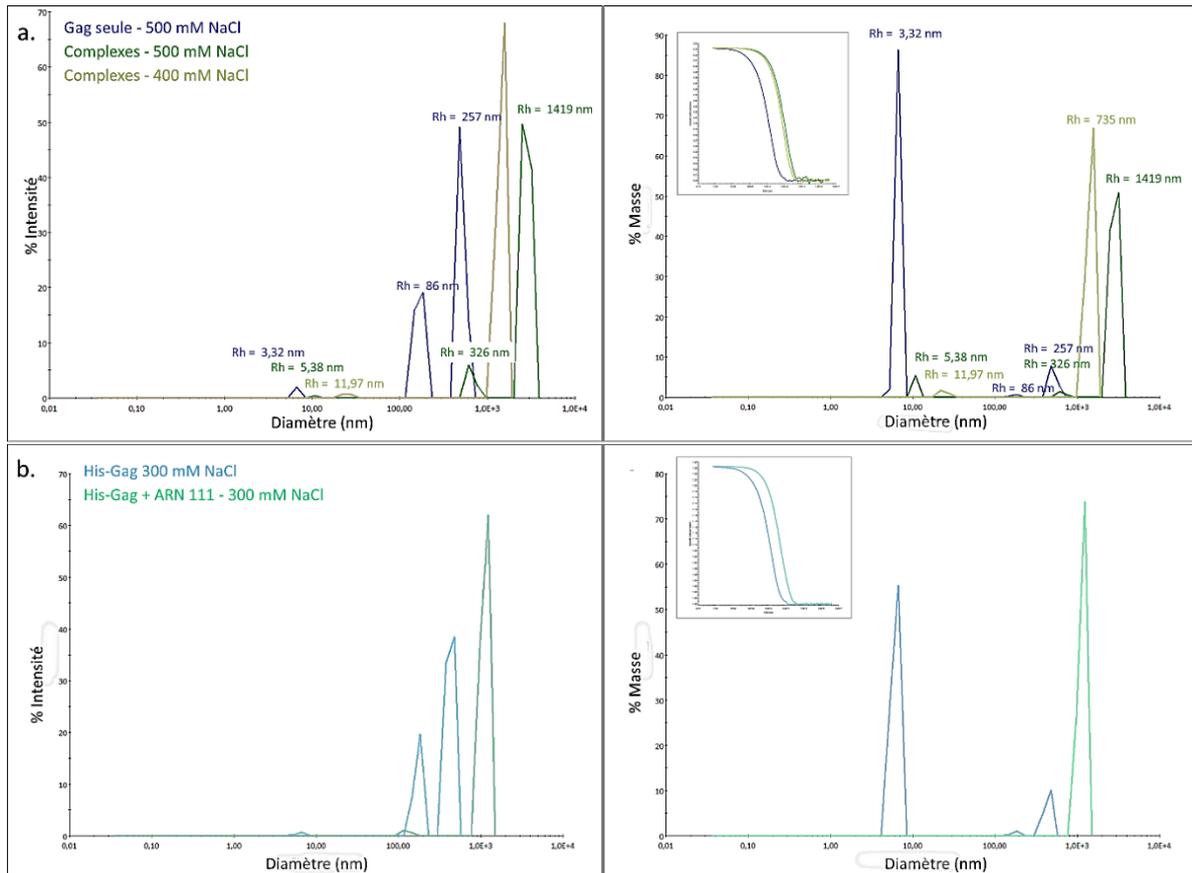
Ainsi, des contrôles de l'impact de cette dilution sur le profil de 6His-Gag en solution sont réalisés par DLS (**figure 52**). Dans l'ensemble, les profils DLS obtenus après l'analyse de la protéine en solution dans le tampon A3a (500 mM NaCl), puis diluées aux concentrations finales de NaCl de 400 mM (tampon A3b) et 300 mM (tampon A3c) sont comparables. Le pic correspondant à la protéine monomérique est majoritaire dans la distribution en fonction de la masse pour les trois concentrations. Néanmoins, la protéine présente également des multimères de taille variable d'une concentration de NaCl à l'autre. A noter que le corrélogramme ne présente dans ce cas, aucun signe d'agrégation.



**Figure 52. Comparaison par DLS de l'effet de la dilution de la concentration de NaCl pour la protéine 6His-Gag en solution dans le tampon A3**

Représentation des distributions en intensité (à gauche) et en masse (à droite) en fonction des diamètres des protéine 6His-Gag en solution. Le Rh est indiqué au-dessus du pic correspondant. Chaque graphique correspond à la moyenne de dix mesures. Le corrélogramme est représenté dans l'encadré. La protéine Gag purifiée dans le tampon A3a (500 mM NaCl) est représentée en noire, la protéine diluée dans 400 mM NaCl (tampon A3b) est représentée en bleu roi et finalement la protéine diluée dans 300 mM NaCl (tampon A3c) est représentée en bleu turquoise.

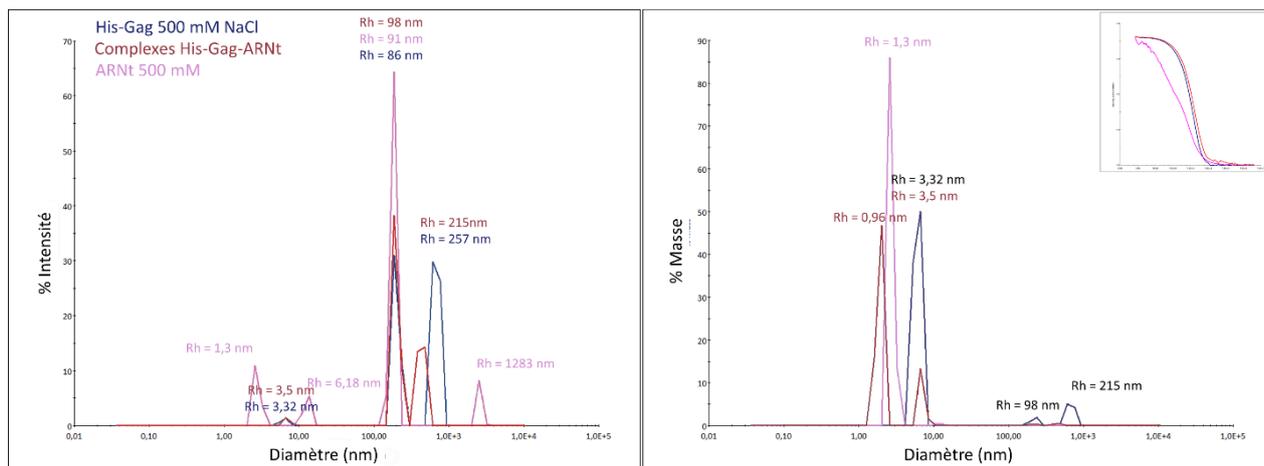
Lorsque l'ARN111<sub>WT</sub> est rajouté à la protéine (courbes vertes, **figure 53**), le chromatogramme issue de l'analyse DLS du complexe présente un décalage de la valeur du Rh du premier pic par rapport au contrôle 6His-Gag seul (courbe bleue, **figure 53 a**). De plus, des pics correspondants à des molécules présentant une grande valeur de Rh apparaissent à la fois dans la distribution par intensité et dans celle par masse. Ce phénomène est encore plus visible lorsque la concentration de NaCl est de 300 mM (**figure 53 b**). Ces pics sont relativement monodisperse et les corrélogrammes présentent peu d'agrégation. Il est possible que l'ajout d'ARN à la protéine Gag induise l'oligomérisation de la protéine caractérisée par l'augmentation de la valeur de Rh associée .



**Figure 53. Comparaison par DLS de l'effet de la concentration de NaCl sur la formation de complexes entre 6His-Gag et l'ARN111WT**

Représentation des distributions en intensité (à gauche) et en masse (à droite) en fonction des diamètres des complexes entre 6His-Gag et l'ARN111<sub>WT</sub> en solution. Le Rh est indiqué au-dessus du pic correspondant. Chaque graphique correspond à la moyenne de dix mesures. Les corrélogrammes sont représentés dans les encadrés. **a.** Le profil DLS de la protéine Gag seule et purifiée dans le tampon A3a (500 mM NaCl) est représenté en bleu foncé. Celui du complexe entre 6His-Gag et l'ARN111<sub>WT</sub> dans le tampon A3a est en vert foncé. Enfin, le dernier correspondant au complexe ribonucléique dans le tampon A3 dont la concentration de NaCl est diluée à 400 mM (tampon A3b) est représenté par une courbe vert clair. **b.** Le profil DLS de la protéine Gag seule et purifiée dans le tampon A3a est représenté en bleu clair et le profil DLS du complexe ribonucléique dans le tampon A3 dont la concentration de NaCl est diluée à 300 mM (tampon A3c) est représenté par une courbe vert turquoise.

Afin de vérifier l'hypothèse selon laquelle le décalage des valeurs de Rh observé dans les différents profils DLS est bien la conséquence d'une interaction spécifique entre la protéine 6His-Gag et l'ARN111<sub>WT</sub>, un contrôle est réalisé en ajoutant un mélange d'ARNt à la protéine. Les complexes sont formés en conservant le même rapport d'une molécule d'ARN pour 6 molécules de protéines Gag recombinantes. Les profils DLS obtenus pour l'ARNt seul sont superposés aux profils de la protéine 6His-Gag seule et à ceux du complexe entre le précurseur et l'ARNt, aucun décalage de Rh n'apparaît et les pics se superposent entre les courbes (**figure 54**).



**Figure 54. Comparaison par DLS de l'effet de l'ajout d'ARNt ou d'ARN111WT à la protéine 6His-Gag dans le tampon A3**

Représentation des distributions en intensité (à gauche) et en masse (à droite) en fonction des diamètres du mélange d'ARNt en solution. Le Rh est indiqué au-dessus du pic correspondant. Chaque graphique correspond à la moyenne de dix mesures. Le corrélogramme est représenté dans l'encadré en haut à droite. Le profil DLS correspondant à la protéine Gag seule est représentée en bleu foncé. Celui correspondant au complexe entre 6His-Gag et l'ARNt est représenté en rouge et l'ARNt seul est représenté en rose.

Cela appuie l'hypothèse que les décalages de la valeur de Rh de 3,32 nm à 5,38 nm dans le tampon A3a (500 mM NaCl) et de 3,32 nm à 11,97 nm dans le tampon A3b (400 mM NaCl) observé lorsque la protéine 6His-Gag et l'ARN111<sub>WT</sub> sont mélangés correspond bien à une interaction fortement spécifique entre ces deux molécules.

Concentration NaCl		Rh 1	Rh 2	Rh 3	Rh 4
ARN111 <sub>WT</sub>	500 mM	<b>(nm)</b>	<b>6,48</b>	<b>258</b>	
		% Intensité	69%	30%	
		% Masse	99%	1%	
6His-Gag	500 mM	<b>(nm)</b>	<b>3,32</b>	<b>86</b>	<b>257</b>
		% Intensité	1,9%	35%	63%
		% Masse	88,4%	0,9%	10,7%
	400 mM	<b>(nm)</b>	<b>3,32</b>	<b>136</b>	<b>2419</b>
		% Intensité	1%	91%	8%
		% Masse	68%	16%	16%
300 mM	<b>(nm)</b>	<b>3,2</b>	<b>87</b>	<b>216</b>	
	% Intensité	1%	27%	72%	
	% Masse	84%	1%	15%	
6His-Gag sans ARN	500 mM	<b>(nm)</b>	<b>3,32</b>	<b>86</b>	<b>257</b>
		% Intensité	2%	35%	63%
		% Masse	88%	1%	11%
6His-Gag + ARN111 <sub>WT</sub>	500 mM	<b>(nm)</b>	<b>5,38</b>	<b>326</b>	<b>1419</b>
		% Intensité	0,4%	8,3%	91,3%
		% Masse	5%	2%	93%
	400 mM	<b>(nm)</b>	<b>11,97</b>	<b>735</b>	
		% Intensité	1,2%	98,9%	
		% Masse	2,7%	97,3%	
300 mM	<b>(nm)</b>	<b>338</b>	<b>1045</b>		
	% Intensité	48%	52%		
	% Masse	23%	77%		
ARN t	500 mM	<b>(nm)</b>	<b>1,3</b>	<b>6,18</b>	<b>91,7</b>
		% Intensité	14%	7%	70%
		% Masse	99,4%	0,5%	0%
6His-Gag + ARN t	500 mM	<b>(nm)</b>	<b>0,96</b>	<b>3,5</b>	<b>98</b>
		% Intensité	0,3%	1,7%	58%
		% Masse	73%	25%	1%

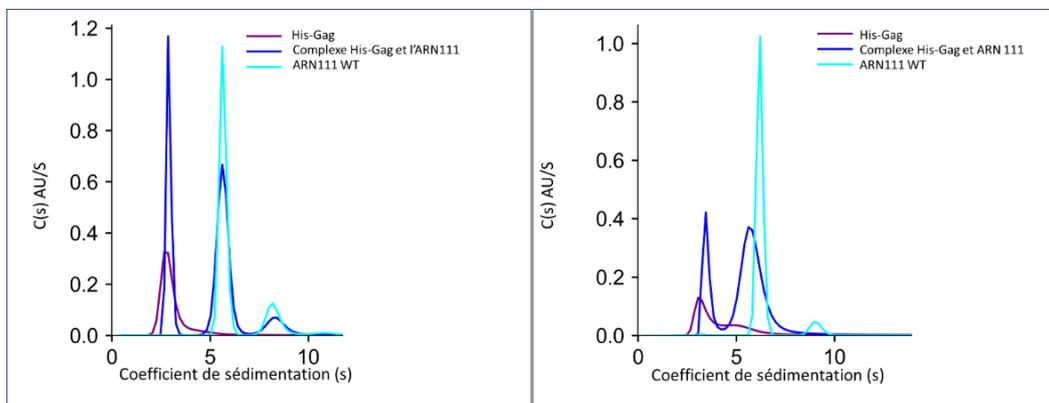
**Tableau 10. Valeurs de Rh déterminés grâce par DLS de la protéine 6His-Gag, des ARN 111WT et ARNt et des complexes protéine/ARN dans différentes dilutions en NaCl de la solution tamponnée A3**

En conclusion, une interaction spécifique entre le fragment d'ARNg et Gag est observable dans la solution tamponnée A3 *in vitro*. La taille des oligomères augmente quand la concentration de NaCl diminue (tampons A3b et A3c). Dans le but de caractériser les paramètres de liaison des complexes (comme la constante d'affinité et la stœchiométrie du complexe protéine-ARN), d'autres analyses sont réalisées notamment par AUC.

### I.2.3 Analyse des complexes formés entre 6His-Gag et l'ARN111<sub>WT</sub> par AUC

Les complexes sont formés entre 6His-Gag produites en cellules BHK21 et l'ARN111<sub>WT</sub> dans les tampons A1 et A3. A cette fin, l'ARN et la protéine sont aussi analysés seuls dans l'une et l'autre solution tamponnées (courbe bleu clair et violet respectivement **figure 55**). Le complexe est formé à une stœchiométrie de 1 : 1 avec une concentration final d'ARN dimérique et de protéines égale à 1  $\mu$ M.

Tout d'abord, l'analyse de vitesse de sédimentation indique que l'ARN111<sub>WT</sub> dans l'une et l'autre solution tamponnée est composé d'une population largement dimérique (**figure 55 et tableau 11**). La protéine 6His-Gag est majoritairement monomérique dans les deux solutions tamponnées, bien qu'une faible proportion de dimère semble apparaître dans la solution tamponnée A3. En superposant les courbes obtenues aucun complexe ribonucléoprotéique dans ces conditions n'est observé.



**Figure 55. Représentation des analyses de vitesse de sédimentation de 6His-Gag mélangé à l'ARN111WT**

L'analyse de la distribution du coefficient de sédimentation  $C_s$  en continu est réalisée à partir des données d'absorbance. Les courbes violettes représentent la protéine 6His-Gag purifiée dans le tampon (A1 ou A3), la courbe bleu turquoise, l'ARN replié dans la solution tamponnée correspondante et la courbe foncée le complexes entre 6His-Gag et l'ARN111<sub>WT</sub>. Les graphiques sont réalisés en utilisant la plateforme Gussi (Brautigam, 2015). **Gauche.** solution tamponnée A1. **Droite.** solution tamponnée A3.

	Pics	PM (Da)	Coefficient de. sedimentation (S)	%	Rh (nm)
<b>Tampon A1</b>					
<b>ARN 111</b>	N°1	52 135	4,12	1	3,02
	N°2	137 684	7,88	91	4,17
	N°3	243 558	11,52	8	5,04
<b>Gag</b>	N°1	59 419	4,51	27	3,11
	N°2	125 631	7,58	71	3,96
<b>Complexes</b>	N°1	59 419	4,51	29	3,11
	N°2	125 631	7,58	71	3,96
<b>Tampon A3</b>					
<b>ARN 111</b>	N°1	75 880	5,70	81	3,69
	N°2	135 627	8,40	16	4,48
<b>Gag</b>	N°1	68 596	3,80	95	4,30
<b>Complexes</b>	N°1	44 904	3,73	37	2,89
	N°2	122 034	5,67	50	3,79
	N°3	233 626	8,43	8,9	4,62

**Tableau 11. Valeurs Cs mesurées en AUC pour l'ARN111WT les différentes solutions tamponnées A1 et A3**

#### I.2.4 Analyse des complexes formés par 6 His-Gag et l'ARN111<sub>WT</sub> par Fidabio

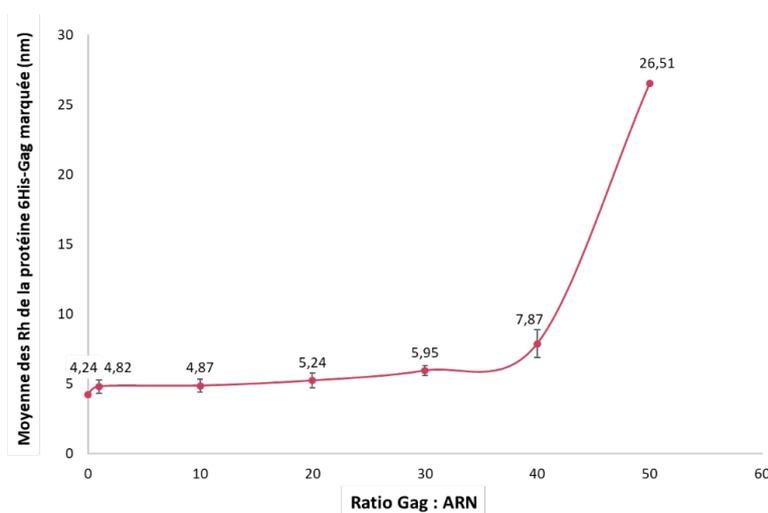
Différentes ratios protéines/ARN entre 6His-Gag et l'ARN111<sub>WT</sub> sont testées à l'aide d'analyses FIDAbio (dont les principes sont décrits dans la session Matériels et Méthode ; §IX.3). Pour cela les complexes sont formés en faisant varier la concentration de l'ARN et en conservant une concentration fixe de la protéine Gag marquée par des sondes ALEXA488 dans le tampon A3 (ici de 350 nM).

Dans un premier temps, l'efficacité de marquage de la protéine est évaluée. Elle varie entre 50% et 80% en fonction des tubes (estimation à partir de la formule présentée dans le Matériels et Méthodes ; §VII.4). Ensuite, le FIDalyser permet d'évaluer le Rh de la protéine recombinante marquée en solution. Cette valeur est corrigée en fonction de la viscosité expérimentale de la solution tamponnée correspondant à 0,00088 Pa.s<sup>-1</sup> (**annexe 3**). Ainsi le Rh de la protéine seule en solution est estimé à 4,24 ± 0,18 nm ce qui est bon accord avec les valeurs de Rh obtenus par DLS et SEC-MALS (**tableau 8**, section Résultats ; §I.1.5). Cette valeur est la moyenne de huit analyses indépendantes.

Lorsque la protéine 6His-Gag marquée est mélangée avec l'ARN111<sub>WT</sub> préalablement replié et dimérique, le Rh observé augmente (**tableau 12 et figure 56**). Les Rh représentés ici correspondent à une moyenne d'au moins 4 mesures, à l'exception du dernier point de stœchiométrie de 50 molécules de Gag pour 1 dimère d'ARN111<sub>WT</sub>. Pour ce complexe, seulement une analyse est réalisée, puisque dans ce cas on observe une certaine agrégation.

Ratio Gag : ARN	Moyenne des Rh (nm)	Ecart type	Nombre de répétition
1 : 0	4,24	0,18	8
1 : 1	4,82	0,48	4
10 : 1	4,87	0,48	6
20 : 1	5,24	0,53	6
30 : 1	5,95	0,37	5
40 : 1	7,87	1,00	5
50 : 1	26,51	X	1

**Tableau 12.** Rh moyens mesurés par FIDAbio à partir de mélanges de 6His-Gag et d'ARN 111WT en fonction des ratios protéine-ARN

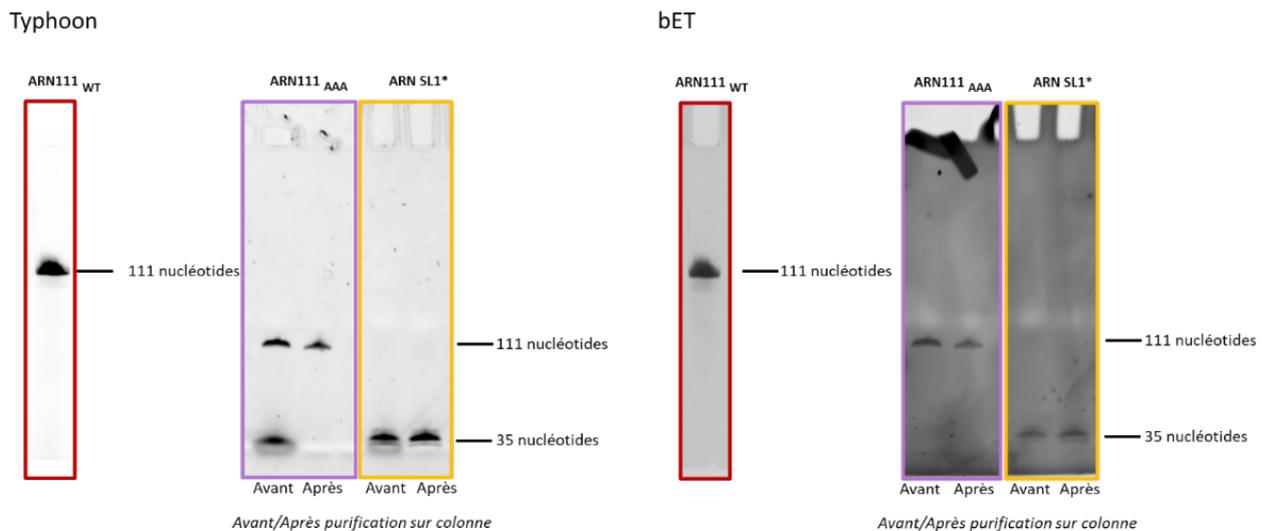


**Figure 56.** Représentation des Rh moyens de la protéine 6His-Gag marquée par une sonde ALEXA488, en présence d'une quantité croissante de dimères d'ARN 111WT obtenus par FIDAbio

Une forte augmentation du Rh apparaît pour les ratios de 40 et 50 molécules de Gag pour 1 molécule ARN. Cependant, ce genre d'analyses sont difficilement reproductibles et la méthode FIDAbio n'est pas optimale pour l'analyse des interactions entre Gag et les fragments d'ARN viraux. En effet son fonctionnement repose sur le déplacement laminaire des complexes en solutions dans un capillaire chargé soit négativement, soit recouvert de charges positives. Dans le premier cas, la présence de protéines Gag chargées positivement rend l'utilisation d'un tel capillaire extrêmement difficile. Dans le second cas, il est impossible de déterminer le pourcentage d'ARN qui pourrait interagir avec le capillaire en interférant ainsi avec la formation du complexe. Pour contrecarrer cette problématique de nouveaux capillaires vont être commercialisés cependant au moment des analyses ils n'étaient pas encore disponibles. Ainsi, nous avons décidé de poursuivre notre analyse par une nouvelle méthode biophysique, l'anisotropie de fluorescence.

### I.2.5 Analyse des complexes formés entre 6His-Gag et ARN 111 WT par Anisotropie de fluorescence

Contrairement à la méthode FIDABio, l'anisotropie de fluorescence utilisée lors de ce travail ne repose pas sur le marquage de la protéine Gag, mais sur celui de l'ARN comme décrit section Matériels et Méthodes ; §III.2. Dans ce but, les deux ARN viraux, ARN111<sub>WT</sub> et ARN<sub>SL135</sub> sont marqués, repliés et dimérisés. Dans un premier temps, l'efficacité de marquage est testée en quantifiant le taux de marquage des fluorophores (FTSC) au niveau de l'extrémité 3' des ARN en réalisant des analyses de SDS PAGE urée qui sont ensuite révélées au Typhoon et par bains de BET. Ensuite, la dimérisation des ARN marqués est contrôlée à l'aide d'un gel d'agarose 1% TMB (figure 49).



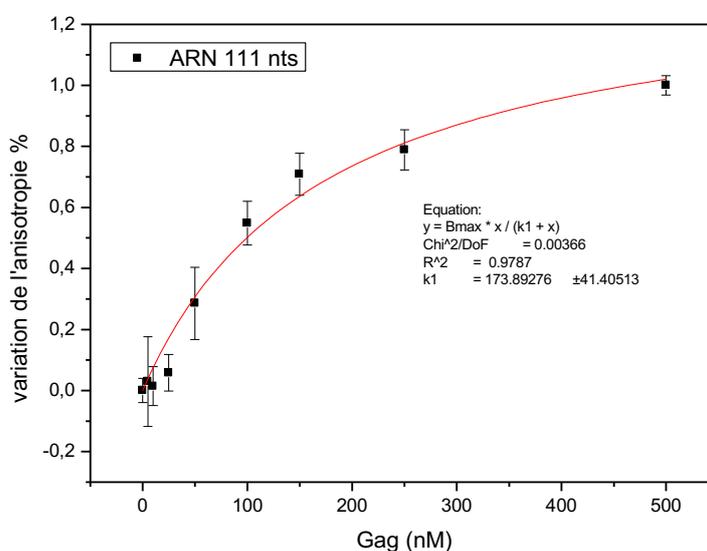
**Figure 57. Analyse par gel dénaturants supplémentés en urée révélés au Typhoon et par un bain de BET du marquage des ARNg**

L'ARN111<sub>WT</sub> est encadré en rouge (à gauche), l'ARN111<sub>AAA</sub> est encadré en violet (au milieu) et l'ARN<sub>SL135</sub> est encadré en jaune (à droite). Pour les ARN111<sub>AAA</sub> et ARN<sub>SL135</sub> la première piste correspond à l'ARN marqué avant l'étape de purification sur colonne suite au marquage fluorescent, la deuxième piste correspond à l'ARN marqué après l'étape de purification sur colonne.

Lorsque les ARN111<sub>WT</sub> fraîchement marqués, repliés et dimériques sont mélangés à des concentrations croissantes de protéines 6His-Gag, la valeur d'anisotropie augmente (**tableau 13 et figure 58**). Chaque ligne du tableau représente la moyenne de 37 mesures. Les courbes sont ensuite normalisées et la moyenne des trois analyses distinctes sont ajustées par le modèle de Scatchard (Scatchard, 1949) de manière à déterminer le Kd. Dans ce cas, le Kd apparent entre l'ARN111<sub>WT</sub> et 6His-Gag est estimée à  $174 \pm 41$  nM (**figure 58**).

Ratio 6His-Gag : ARN111 <sub>WT</sub>	Moyenne de l'anisotropie (mP)	Ecart type	Nombre de répétitions
0 : 1	60	2,29	37
2 : 1	62	3,32	37
5 : 1	68	3,08	37
10 : 1	92	6,93	37
20 : 1	119	6,55	37
30 : 1	140	8,87	37
50 : 1	151	5,90	37
100 : 1	179	4,48	37

**Tableau 13. Valeurs moyennes d'anisotropie de fluorescence pour des complexes formés entre Gag et l'ARN 111WT en fonction du ratio protéine/ARN**



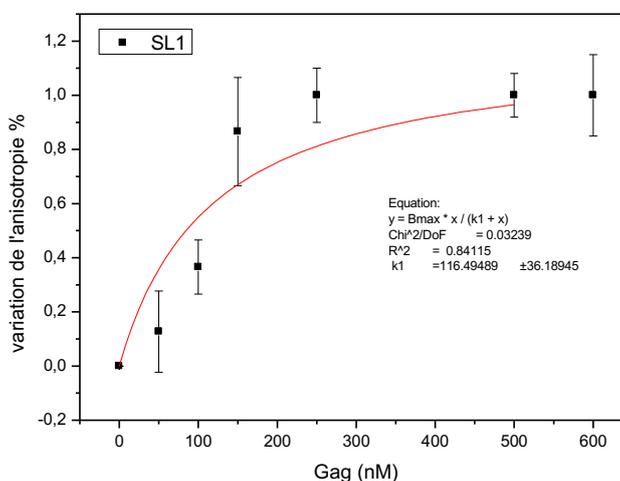
**Figure 58. Représentation des moyennes des valeurs d'anisotropie de l'ARN111WT marqué et dimérique en présence de quantités croissantes de 6His-Gag et ajustement de la courbe par le modèle de Scatchard**

*La courbe représente la moyenne de 37 répétitions (n=3) avec un traitement statistique de Scatchard.*

De même, lorsque les ARNSL1<sub>35</sub> marqués, repliés et dimériques sont mélangés à des concentrations croissantes de protéines 6His-Gag, la valeur d'anisotropie augmente (**tableau 14 et figure 59**). Chaque ligne du tableau représente la moyenne de 8 à 15 mesures. Les courbes ensuite normalisées et ajustées par le modèle de Scatchard ce qui a permis de déterminer le  $K_{d_{app}}$  qui dans ce cas est de  $116 \pm 36$  nM (**figure 59**).

Ratio 6His-Gag : ARN111 <sub>WT</sub>	Moyenne de l'anisotropie (mP)	Ecart type	Nombre de répétitions
0 : 1	67	3,60	6
2 : 1	67	3,08	9
5 : 1	68	3,48	9
10 : 1	71	3,68	9
20 : 1	79	4,21	9
30 : 1	95	17,13	9
50 : 1	100	28,36	6
100 : 1	131	33	9

**Tableau 14. Valeurs moyennes d'anisotropie de fluorescence des complexes entre Gag et l'ARN SL135 en fonction des ratios et du nombre de répétitions**



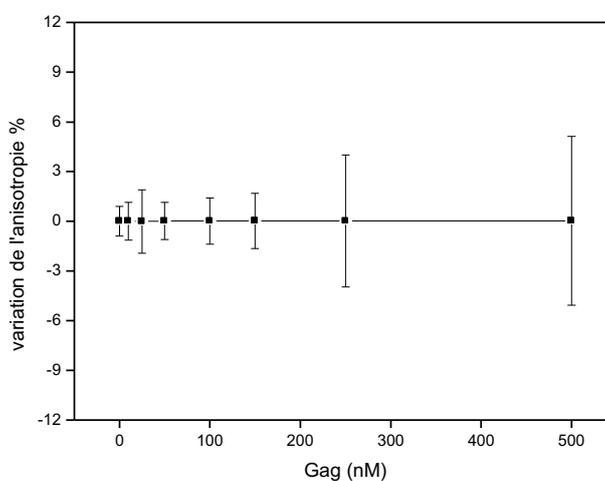
**Figure 59. Représentation des moyennes des valeurs d'anisotropie de l'ARNSL135 marqué et dimérique en présence de quantités croissantes de 6His-Gag et ajustement de la courbe par le modèle de Scatchard**

*La courbe représente la moyenne de 9 répétitions (n=1) avec un traitement statistique de Scatchard.*

Afin de vérifier l'hypothèse selon laquelle l'augmentation de la valeur d'anisotropie observée dans les différentes courbes de titration est bien la conséquence d'une interaction spécifique entre la protéine 6His-Gag et l'ARN111<sub>WT</sub> (**figures 58 et 59**), un contrôle est réalisé en formant les complexes avec le mutant de dimérisation, l'ARN111<sub>AAA</sub> (**tableau 15 et figure 60**). Les valeurs moyennes d'anisotropies pour 8 mesures ne varient pas même lorsque la concentration de Gag augmente, contrairement aux résultats obtenus avec les ARN sauvages (ARN111<sub>WT</sub> et ARNSL1<sub>35</sub>). Ainsi, il est possible de conclure que l'augmentation de la valeur d'anisotropie observée peut être corrélée à une interaction spécifique entre Gag et les fragments d'ARN viraux.

Ratio 6His-Gag : ARN111 <sub>AAA</sub>	Moyenne de l'anisotropie (mP)	Ecart type	Nombre de répétitions
0 : 1	142	0,89	8
2 : 1	140	1,90	8
5 : 1	145	1,13	8
10 : 1	143	1,40	8
20 : 1	145	1,66	8
30 : 1	143	3,99	8
50 : 1	145	5,09	8
100 : 1	154	3,13	8

**Tableau 15. Valeurs moyennes d'anisotropie de fluorescence obtenus à partir de mélanges de 6His-Gag et d'ARN 111AAA en fonction de la stœchiométrie**



**Figure 60. Représentation des moyennes des valeurs d'anisotropie de l'ARN111AAA marqué et replié en présence de quantités croissantes de 6His-Gag**

La courbe représente la moyenne de 8 répétitions ( $n=1$ ) avec un traitement statistique de Scatchard.

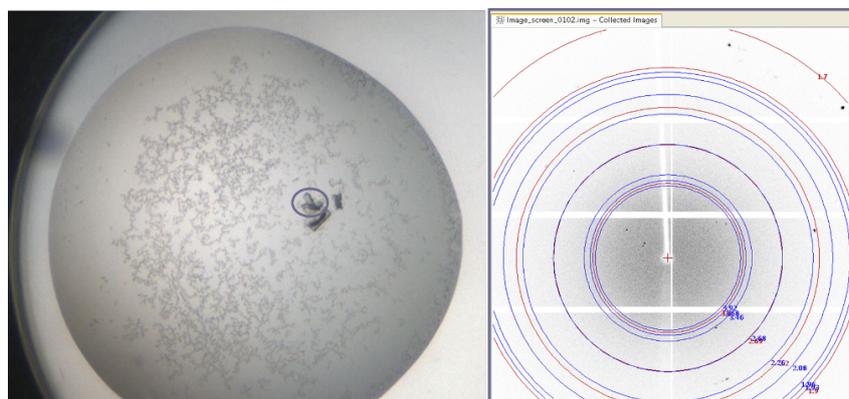
Pour conclure, l'objectif de la deuxième étape de mon travail consiste à déterminer les conditions optimales pour la formation des complexes entre la protéine Gag et les fragments ARN111<sub>WT</sub> et ARNSL1<sub>35</sub> *in vitro*. Ces conditions sont utilisées pour réaliser la structure tridimensionnelle du complexe entre les deux partenaires.

**Résumé :** Le protocole de dimérisation de l'ARNg mis au point par le laboratoire est contrôlé pour les tampons A1 et A3 à l'aide d'un gel d'agarose (TBM) et d'analyses d'AUC sur les ARN111<sub>WT</sub>, ARN111<sub>AAA</sub> et ARNSL1<sub>35</sub>. Les complexes formés entre 6His-Gag et l'ARN111<sub>WT</sub> sont caractérisés par DLS. Ceux formés avec la protéine Gag marquée et l'ARN111<sub>WT</sub> sont analysés par FIDABio. Finalement, le kD est déterminé pour les complexes formés entre la protéine Gag et les ARN111<sub>WT</sub>, ARN111<sub>AAA</sub> ou ARNSL1<sub>35</sub> marqués, par anisotropie de fluorescence

### I. 3. Analyse de la structure de Gag en complexe avec l'ARN

#### I.3.1 Analyse structurale de 6HisGag par cristallographie aux rayons X

Afin d'obtenir la structure du complexe entre 6His-Gag et l'ARN111<sub>WT</sub> par cristallographie aux rayons X, des tentatives de cristallisation de ce complexe sont entreprises. Pour cela, des premiers tests avaient été réalisés au sein du laboratoire avec la protéine 6His-Gag purifiée dans le tampon A1 mais ils n'ont pas permis de mettre en évidence de conditions aboutissant à la formation de cristaux de protéine. Ainsi, une nouvelle tentative est faite avec la protéine 6His-Gag purifiée dans le tampon A1 et dialysée 5 heures à 4°C contre le tampon A3. La concentration de la protéine après dialyse est estimée par mesure d'absorbance au Nanodrop à 2,3 mg/mL. Des cristaux apparaissent uniquement dans la boîte de criblage Wizard I et II (**figure 61**) pour les conditions contenant du di-Ammonium hydrogène phosphate (1 M) comme agent précipitant. Les cristaux sont récupérés et leurs profils de diffraction sont contrôlés aux rayons X. Les clichés de diffraction obtenus sont caractéristiques de cristaux de sel. Ceux-ci sont sans doute issus des réactifs composant la condition de cristallisation et non de la protéine (**figure 61**). Ainsi, à ce jour, l'obtention de cristaux de la protéine 6His-Gag n'a pas encore abouti.



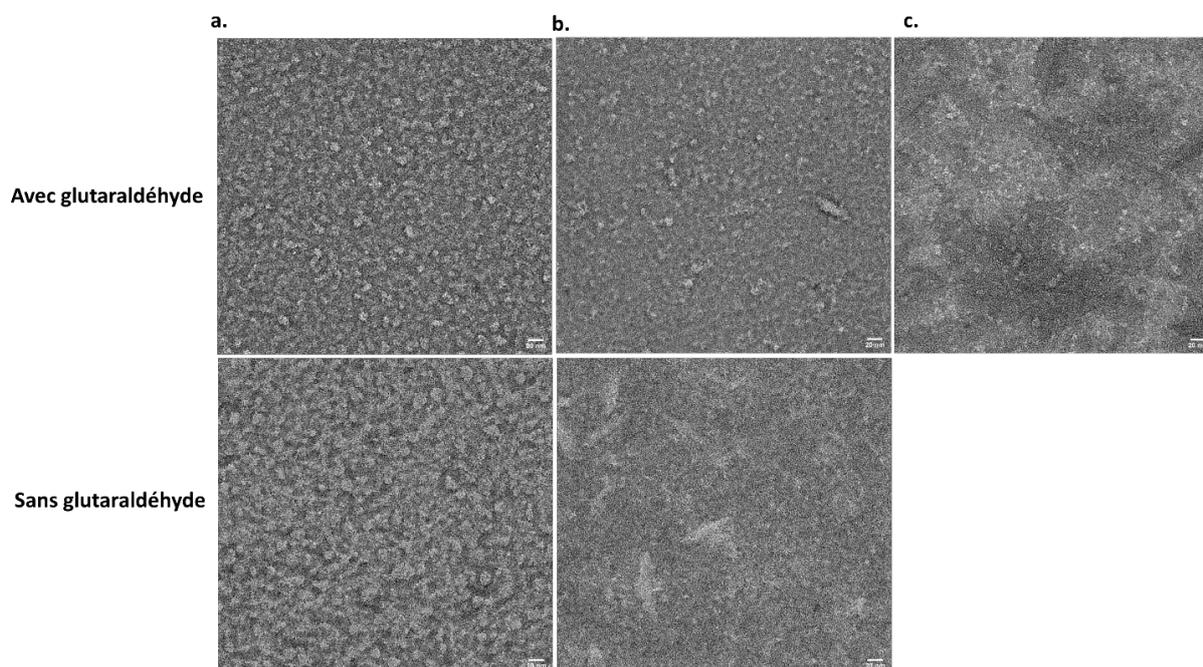
**Figure 61. Analyse par diffraction aux rayons X des cristaux formés en présence de 6His-Gag dans le tampon A3**

*A gauche, formation d'un cristal à partir de 200 nL de protéines et 200 nL d'agents cristallisants contenant du di-ammonium hydrogène phosphate (1 M), imidazole pH 8.0 (100 mM) et chlorure de sodium (200 mM). A droite, profil de diffraction contrôlé aux rayons X.*

#### I.3.2 Mise au point des conditions de formation de grilles de microcopie électronique pour la protéine 6His-Gag par coloration négative

La réalisation de clichés de coloration négative permet de contrôler le comportement de la protéine recombinante 6His-Gag sur des grilles de carbone ainsi que l'effet de la fixation au glutaraldéhyde. Le but est de déterminer les conditions idéales pour réaliser des structures par CryoEM. Différentes concentrations de la protéine purifiée dans le tampon A3 sont déposées sur des grilles en présence ou en absence de glutaraldéhyde (**figure 62**). En conclusion,

la qualité de l'échantillon est meilleure en présence de glutaraldéhyde, elle permet notamment de distinguer des particules isolées (**figure 62 a**).



**Figure 62. Micrographes des grilles de microscopie électronique à coloration négative chargée avec différentes concentration 6His-Gag dans le tampon A3**

Micrographes de différentes concentrations de 6His-Gag (barres d'échelle de 20 nm). **a.** Concentration de 0,06 mg/mL de 6His-Gag. **b.** Concentration de 6His-Gag de 0,30 mg/mL. **c.** Concentration de 6His-Gag de 0,10 mg/mL.

De plus, la concentration idéale est de 0,30 mg/mL. En effet, à cette concentration il est possible de discriminer deux types de population, d'une part des particules ayant un rayon d'environ 4,9 nm et d'autre part des agrégats de plus grande taille. La protéine doit donc être fixée avant d'être déposée sur des grilles de carbone et pas trop concentrée pour éviter son agrégation.

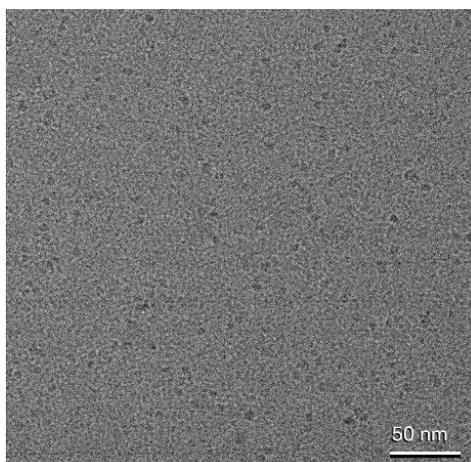
### I.3.3 Analyse de la structure de 6His-Gag par CryoEM

#### I.3.3.1 Criblage de différentes grilles et conditions de préparation des grilles

Différents types de grilles et de conditions de vitrification sont testés pour réaliser les clichés de microscopie électronique (**tableau 16**). Le but de ces criblages est d'identifier les meilleures conditions pour lesquelles la protéine est bien dispersée sur la grille et suffisamment contrastée pour réaliser l'étape de sélection (ou *picking*) lors de la résolution de la structure.

Dans un premier temps des grilles de carbone *Quantifoil* sont utilisées pour déterminer la meilleure concentration de protéines Gag purifiées à déposer sur les grilles, mais aussi la nécessité d'ajouter ou non un agent fixateur

(glutaraldéhyde) et/ou du détergent (DDM). La protéine 6His-Gag s'agrège au niveau du carbone de ces grilles. De plus, le nombre de particules est trop important lorsque 0,25 mg/mL de protéines recombinantes sont déposées sur la grille, mais pas assez importants à 0,15 mg/mL. L'ajout de DDM ainsi que de glutaraldéhyde est nécessaire pour homogénéiser la glace à la surface des grilles même cette glace n'est pas parfaitement homogène. Ensuite, des criblages sont réalisés sur des grilles de carbone C-flat avec une concentration allant de 0,18 mg/mL à 0,3 mg/mL de protéines supplémentées avec du glutaraldéhyde et du DDM. La concentration de 0,18 mg/mL semble idéale pour la répartition de la protéine dans les trous. Même si quelques agrégats subsistent, la qualité des grilles est suffisante pour faire des acquisitions. Une acquisition d'une nuit est réalisée sur l'une de ces grilles (**figure 63**).



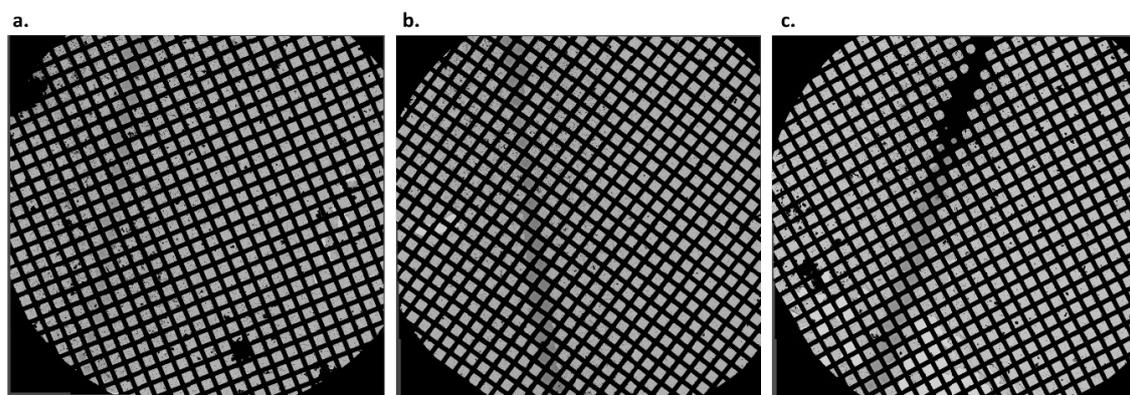
**Figure 63. Micrographe d'une grille C-flat chargée avec 0,18 mg/mL de 6His-Gag purifiée dans le tampon A3 supplémenté avec du DDM et du glutaraldéhyde**

En parallèle, des essais sont réalisés sur des grilles d'or supplémentées ou non avec du PEG. Comme le support change complètement par rapport aux grilles de carbone, les tests de détermination de concentration de protéine Gag purifiée à déposer et de la composition du tampon idéale sont réalisés à nouveau. Comme pour les grilles de carbone, le nombre de particules isolées est trop important lorsque la concentration de Gag est supérieure à 0,20 mg/mL. De plus, le DDM est nécessaire pour homogénéiser la glace à la surface des grilles. Néanmoins, l'utilisation de PEG permet de diminuer la concentration initiale de protéine à 0,15 mg/mL.

Toutes ces grilles sont réalisées en utilisant le Vitrobot avec la même force de séchage (5) et le même temps de séchage (2 secondes). Des tests de variation de temps et de force de séchage sont réalisés en faisant varier la force de 5 à 8 et le temps de 2 à 4 secondes. aucun consensus à partir de ces cribles n'est ressorti. Les grilles séchées avec une force 5 pendant 2 secondes présentent un ATLAS noir synonyme d'un séchage pas assez efficace. Lorsque la même force est appliquée pendant 4 secondes, la qualité de la glace est homogène mais présente de nombreux carreaux cassés, caractéristique d'une étape de séchage trop contraignante. Avec plus de force et un

temps de séchage de 2 secondes la glace sur les grilles est très homogène. Mais trop épaisse pour les grilles séchées avec une force 8 pendant 4 secondes.

Ainsi, pour essayer d'augmenter la reproductibilité de préparation des grilles, des tests sont réalisés en utilisant le Chameleon, robot qui gère automatiquement les étapes d'effluage par plasma, (*glow discharge* ; GD) au dépôt des grilles dans les boîtes de rangement en passant par le dépôt et la vitrification des échantillons. Ainsi des délais avant vitrification allant de 101 ms à 503 ms sont testés. Plus ce délai est court, moins la protéine s'est adsorbée sur la surface de la grille et plus la glace est épaisse. La grille la plus homogène est celle dont le temps avant vitrification est le plus long (**figure 64**).



**Figure 64. Criblages des conditions de préparation des grilles de carbones pour la CryoEM par le Chameleon**

Micrographes des premiers tests de réalisation de grilles de carbones (SPT Labtech) à partir de 0,7 mg/mL de protéines 6His-Gag purifiée dans le tampon A3. **a.** ATLAS de la grille ayant un délai de GD de 30 secondes et de plongeon de 503 ms. **b.** ATLAS de la grille ayant un temps de GD de 30 secondes et délai avant vitrification de 206 ms. **c.** ATLAS de la grille ayant un temps de GD de 30 secondes un délai de 121 ms avant vitrification.

Types de grilles	Concentration de Gag	Tampon	Temps de séchage	Force de séchage	
Quantifoil	0,15	Glutaraldéhyde	2 s	5	
	0,15	Gluta. + DDM	2 s	5	
	0,25	Glutaraldéhyde	2 s	5	
	0,25	Gluta. + DDM	2 s	5	
C-flat	0,18	Gluta. + DDM	2 s	5	Acquisition
	0,18	Gluta. + DDM	4 s	5	
	0,18	Gluta. + DDM	2 s	8	Acquisition
	0,18	Gluta. + DDM	4 s	8	
	0,20	Gluta. + DDM	2 s	5	Acquisition
	0,25	Gluta. + DDM	2 s	5	
	0,30	Gluta. + DDM	2 s	5	
Or	0,2	Glutaraldéhyde	2 s	5	
	0,2	Gluta. + DDM	2 s	5	
Or + Peg	0,05	Gluta. + DDM	2 s	5	
	0,10	Gluta. + DDM	2 s	5	
	0,15	Gluta. + DDM	2 s	5	Acquisition
	0,20	Gluta. + DDM	2 s	5	
Types de grilles	Concentration de Gag		Temps de GD	Temps de plonge	
SptLabtech	0,7		30 s	503 ms	Acquisition
	0,7		30 s	301 ms	
	0,7		30 s	206 ms	
	0,7		1 min	101 ms	
	0,7		1 min	146 ms	
	0,7		30 s	121 ms	

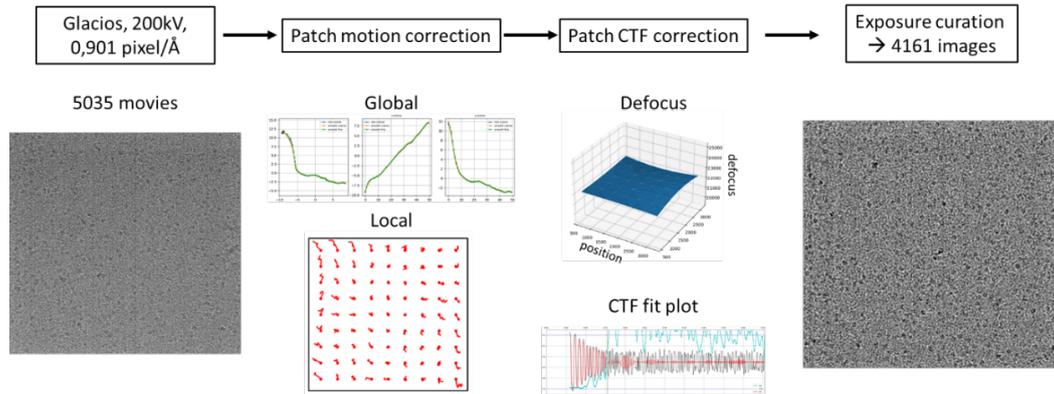
**Tableau 16. Récapitulatif des grilles et des conditions qui sont criblées pour réaliser la structure par CryoEM de 6His-Gag**

*La grille surlignée est celle qui a servie à proposer le modèle ci-dessous (figures 65, 66 et 67).*

### I.3.3.2 Traitement des images CryoEM pour l'ensemble des jeux de données de la protéine Gag seule

Une acquisition sur une nuit est réalisée à partir d'une grille C-Flat contenant 0,18 mg/mL de 6His-Gag supplémentée avec du DDM et fixé avec du glutaraldéhyde. L'acquisition est réalisée sur le microscope Glacios Cryo TEM à 200 kV. Un total de 5035 vidéos sont obtenues. Après alignement des films et correction de la fonction de transfert de contraste (*Contrast Transfer Function, CTF*) le jeu de données est nettoyé pour aboutir à 4161 images **(figure 65)**.

### Data collection (Glacios) and image processing (Cryosparc)

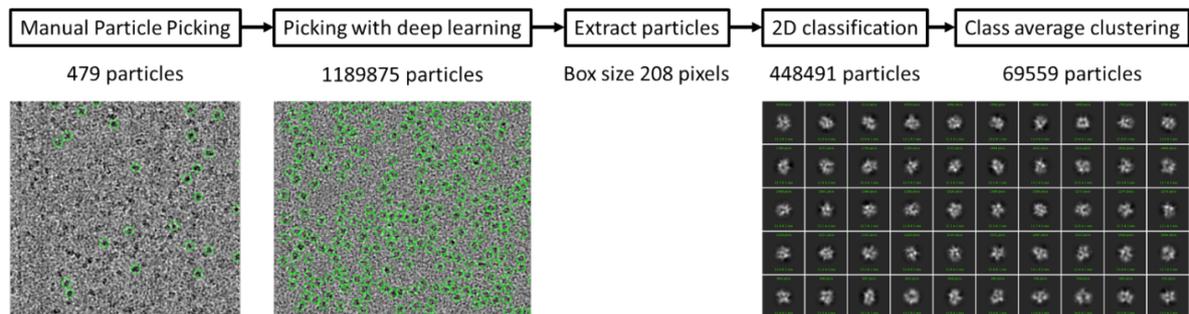


**Figure 65. Reproduction des différentes étapes de construction du modèle de structure de la protéine 6His-Gag (1/3)**

*Alignement des vidéos pour générer les images de microscopie, correction de la CTF et nettoyage du jeu de données. Sélection supervisée de particules à l'aide de modèles d'apprentissage (deep learning).*

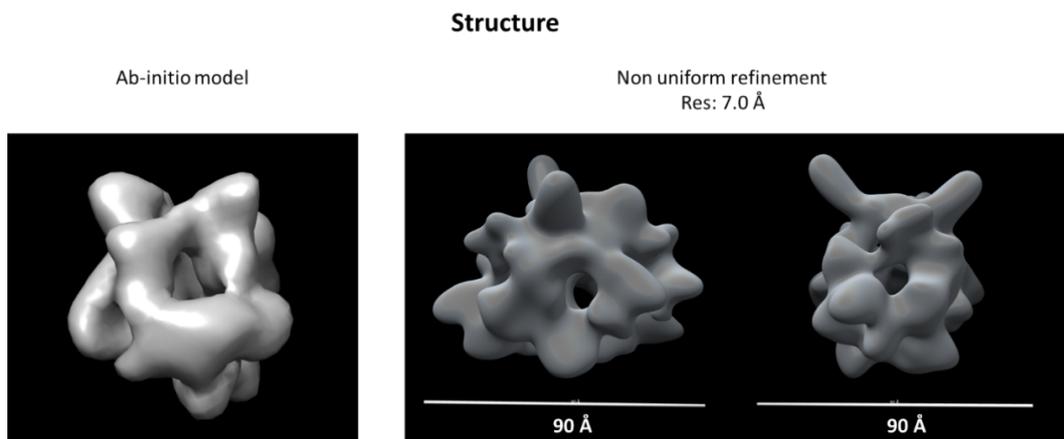
Pour sélectionner et extraire les particules de Gag des images, un choix manuel est réalisé permettant la sélection de 479 particules. Ce jeu de particules est utilisé pour éduquer un programme d'apprentissage de CryoSPARC qui sélectionne 1189875 particules. Ces particules permettent de réaliser des classes 2D. Après plusieurs cycles de nettoyage des mauvaises classes, 448491 particules sont sélectionnées (**figure 66**). Après une étape de « clustering » des classes 2D, une classe comprenant 69559 particules représentant une conformation homogène de Gag est extraite et utilisée pour faire un modèle initial qui est affiné à 7 Å de résolution (**figure 67**).

### Particles picking and cleaning (Cryosparc)



**Figure 66. Reproduction des différentes étapes de construction du modèle de structure de la protéine 6His-Gag (2/3)**

*Extraction des particules sélectionnées d'abord à la main puis en utilisant le programme d'intelligence artificielle de CryoSPARC pour réaliser les classes 2D qui sont ensuite nettoyées.*



**Figure 67. Reproduction des différentes étapes de construction du modèle de structure de la protéine 6His-Gag (3/3)**

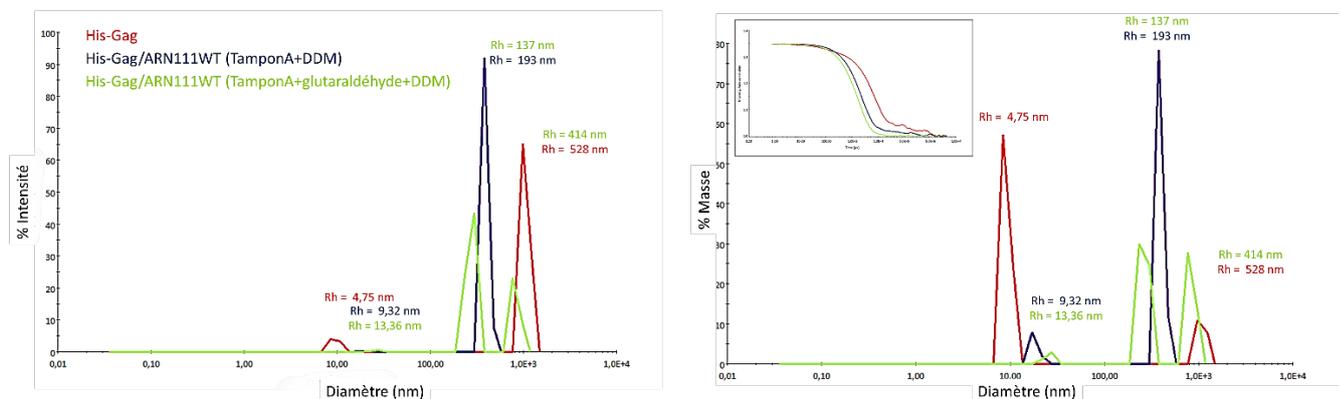
*Un model ab-initio est généré à partir des classes 2D triées et nettoyées. Ce modèle est affiné en utilisant des masques. La structure finale est représentée à droite, à une résolution de 7 Å.*

La résolution de ce modèle est trop faible pour permettre de positionner les structures des domaines de Gag déjà résolues. De nouveaux enregistrements supplémentaires sont nécessaires pour permettre d'avoir un nombre suffisant de particules par classe de conformation. Cependant, il est possible de déduire de ce premier modèle que la protéine 6His-Gag est sous forme de monomère de dimensions 50 x 50 x 90 Å dans le tampon A3.

#### I.3.3.3 Analyse de la structure du complexe entre 6His-Gag et l'ARN111<sub>WT</sub> par CryoEM

Dans le but de résoudre la structure du complexe entre 6His-Gag et l'ARN111<sub>WT</sub>, des premiers tests de préparation de grilles sont réalisés. Pour cela les complexes sont formés comme décrit dans la partie Matériels et Méthodes, §VIII.1 avec une stœchiométrie d'une molécule d'ARN pour dix molécules de Gag (1 : 10) dans le tampon A3b (400 mM). Différentes concentrations sont testées, 0,18 mg/mL (3 µM de protéines Gag) ; 0,2 mg/mL (3,4 µM de protéines Gag) et 0,25 mg/mL (4,2 µM de protéines Gag). Comme pour les grilles ne comportant que la protéine Gag, une fois les complexes formés ils sont fixés avec du glutaraldéhyde et complétés avec du DDM. Ainsi, avant de réaliser les grilles pour la CryoEM, des analyses de DLS sont utilisées pour déterminer l'impact du DDM et du glutaraldéhyde sur la formation du complexe (**figure 68**). Les profils obtenus en masse et en intensité entre la protéine seule dans la solution tamponnée A3 (courbe rouge) sont superposés avec les profils obtenus pour le complexe entre Gag et l'ARN111<sub>WT</sub> dans le tampon A3 supplémenté avec du DDM (courbe bleue) et dans le tampon A3 supplémenté avec du DDM et du glutaraldéhyde (courbe verte). D'une part, 2 pics sont présents sur les profils DLS lorsque du DDM est ajouté dans la solution tamponnée. Le premier pic ne représente que 0,4% de l'intensité et 10 % de la masse mais présente un Rh de 9,32 nm. Il correspondrait à de gros multimères de Gag ou à un complexe entre 6His-Gag et l'ARNg (**tableau 10**, section résultats, partie 2, §I.1.5). L'autre pic est un oligomère de

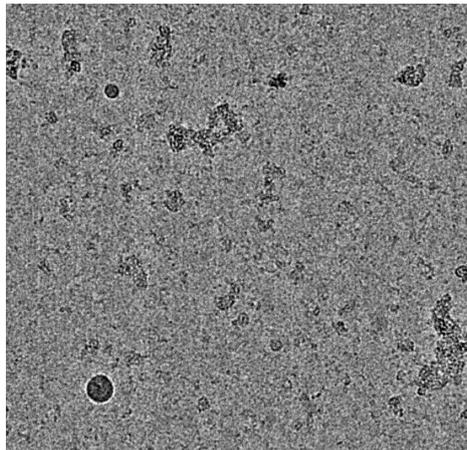
grande taille ayant un Rh de 193 nm. D'autre part, 3 pics sont présents sur les profils DLS lorsque du DDM et du glutaraldéhyde sont ajoutés dans la solution tamponnée. Le premier pic ne représente que 0,8% de l'intensité et 4 % de la masse mais présente un Rh de 13 nm comparable à ce qui est déterminé pour le complexe entre Gag et l'ARNg en DLS (**tableau 10**, section résultats, partie 2, §I.1.5). Les autres pics sont des multimères de grandes tailles ayant des Rh de 132 nm et 420 nm respectivement. Comme caractérisé précédemment, les pics de plus grande taille ne se superposent pas et correspondent tous à des oligomères ou des agrégats de grande taille présentant une faible reproductibilité.



**Figure 68. Analyse par DLS de l'effet du DDM couplé ou non avec du glutaraldéhyde, sur la formation de complexes entre 6His-Gag et l'ARN dans le tampon A3**

La protéine purifiée dans le tampon A3 est représentée en rouge, le complexes entre 6His-et l'ARN et supplémenté avec du DDM est représenté en bleu, finalement le complexes entre 6His-et l'ARN et supplémenté avec du DDM et fixé avec du glutaraldéhyde est représenté en vert. Les distributions en intensité (à gauche) et en masse (à droite) en fonction des diamètres des particules en solution sont représentées pour chacune des protéines testées. Le Rh de chaque protéine est indiqué au-dessus du pic correspondant. Chaque graphique correspond à la moyenne de dix mesures.

Ainsi les différentes concentrations de complexes sont déposées sur les grilles C-flat par le Vitrobot Une acquisition sur la nuit est faite sur la grille contenant 0,2 mg/mL (3  $\mu$ M de protéines Gag).Le complexe Gag/ARN est majoritairement présent sous forme d'agrégats sur les grilles (**figure 69**). L'affinement des conditions de préparation du complexe ainsi que les conditions de préparation des grilles doivent encore être réalisés.

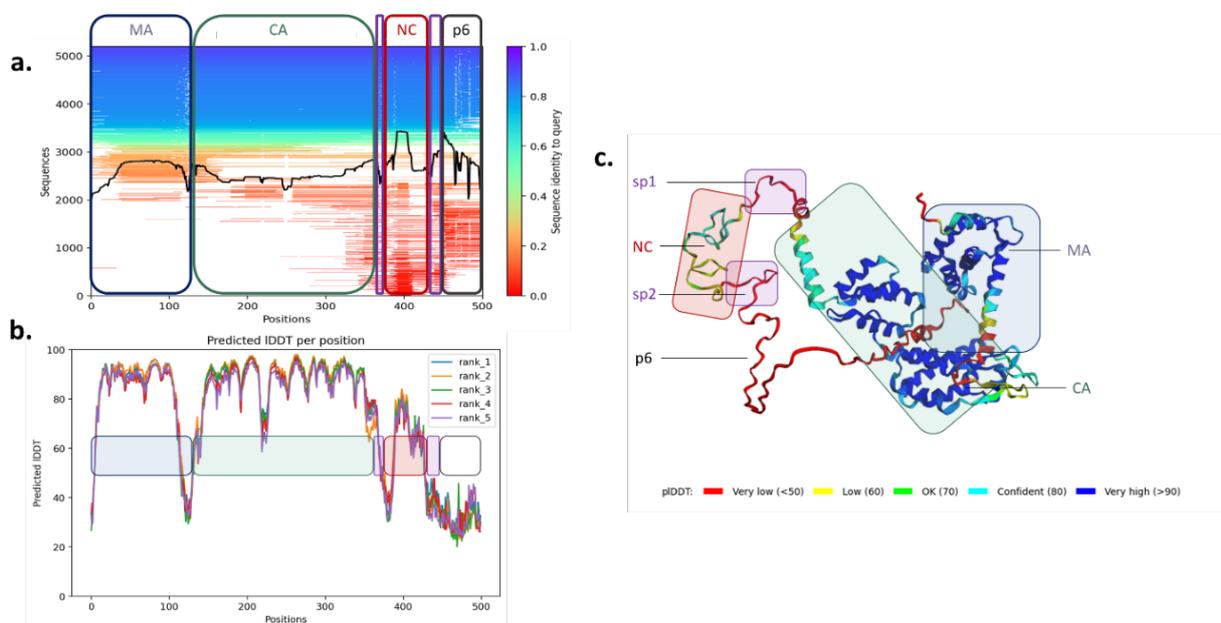


**Figure 69. Micrographe d'une grille C-flat chargée avec 0,20 mg/mL de 6His-Gag en complexe avec l'ARNg (1 : 10) supplémenté avec du DDM et du glutaraldéhyde**

### I.3.4 Proposition d'un modèle de structure de Gag en utilisant le programme Alphafold

Des modèles théoriques sont réalisés grâce la technologie innovante d'intelligence artificielle contenue dans Alphafold. Ainsi des modèles des protéines Gag, avec ou sans l'étiquette dans la partie N-terminale et avec ou sans le domaine p6, sont générés.

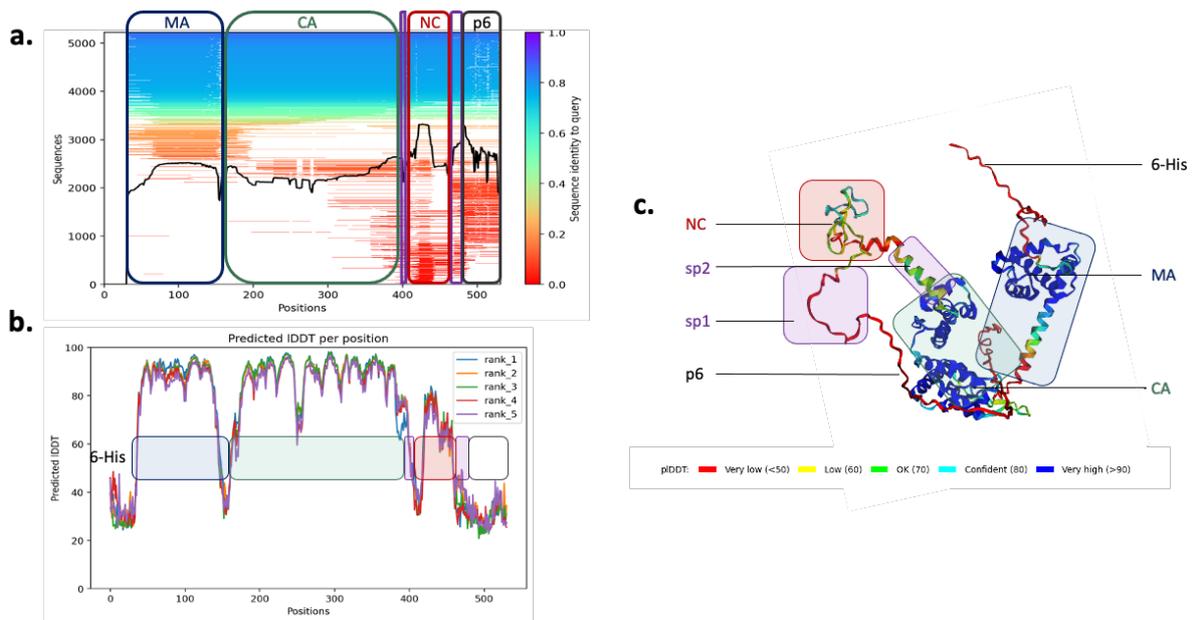
Pour la protéine Gag entière, sans étiquette le modèle présenté est celui ayant le meilleurs score pLDDT de 75,43, soit le modèle numéro 3 (**figure 70**). Les domaines de la MA, CA, NC sont représentés avec un bon score de prédiction. Cependant, les espaceurs et le domaine de p6, qui sont les domaines moins structurés de la protéine, ne présentent pas un bon score pLDDT. Cela signifie que le positionnement des domaines proposés par Alphafold n'est pas sûr.



**Figure 70. Visualisation du modèle proposé par Alphafold pour la structure de la protéine Gag ainsi que les scores associés à la prédiction de la position de chaque résidu**

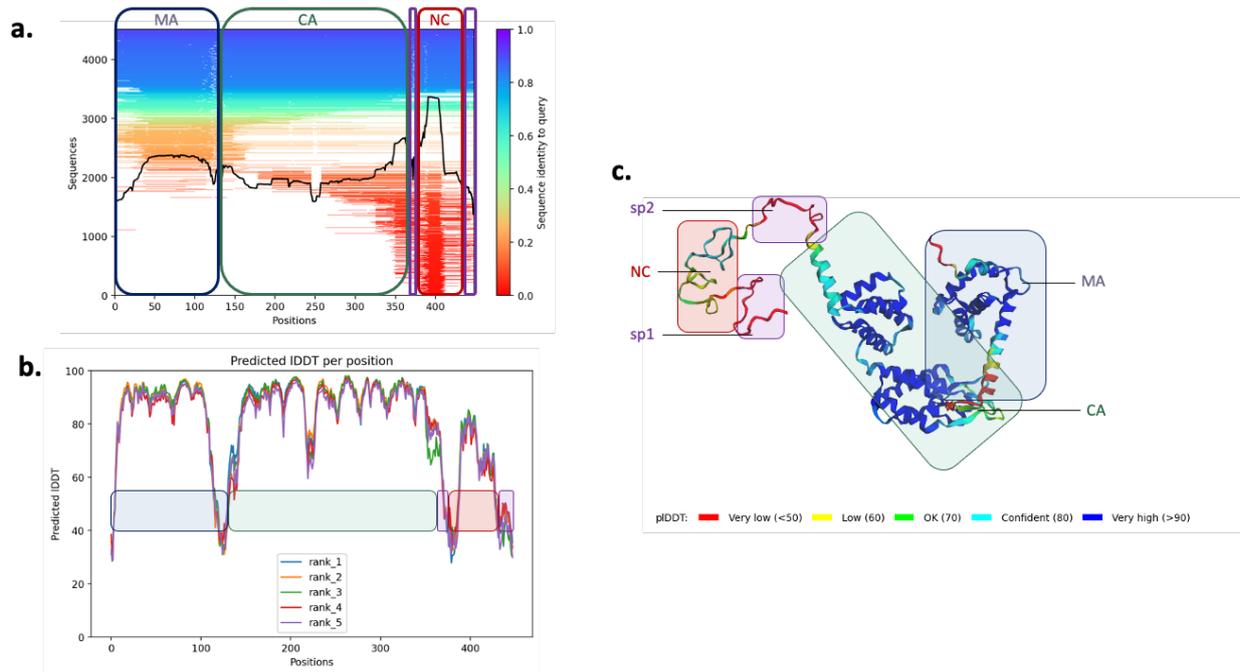
**a.** Représentation de la couverture des alignements multiple de séquence. **b.** IDDT par position en fonction des modèles. **c.** Représentation de la structure de Gag à partir du modèle 3, dont les couleurs varient en fonction du score de IDDT. Les différents domaines de Gag sont représentés en carrés de couleurs.

Concernant la structure de la protéine Gag entière avec l'étiquette 6His dans sa région N-terminale, le modèle présenté est le numéro 5 avec un score pLDDT de 72,68 (**figure 71**). Une proposition de modèle un peu différente de celle obtenu avec la protéine Gag sauvage, néanmoins, les diagrammes de prédiction par positions sont comparables.



**Figure 71. Visualisation du modèle proposé par AlphaFold pour la structure de la protéine 6His-Gag ainsi que les scores associés à la prédiction de la position de chaque résidu**  
**a.** Représentation de la couverture des alignements multiples de séquence. **b.** IDDT par position en fonction des modèles. **c.** Représentation de la structure de Gag à partir du modèle 5, dont les couleurs varient en fonction du score de IDDT. Les différents domaines de Gag sont représentés en carrés de couleurs.

Finalement, la protéine Gag sans le domaine p6 est représentée par le modèle numéro 4 avec un score pLDDT de 81,37 (**figure 72**). Le score est meilleur que les deux modèles précédents. Cependant, le domaine qui est retiré (p6) est un domaine dont la structure est variable, elle apporte donc beaucoup d'incertitude.



**Figure 72. Visualisation du modèle proposé par Alphafold pour la structure de la protéine Gag dépourvue du domaine p6 ainsi que les scores associés à la prédiction de la position de chaque résidu**

**a.** Représentation de la couverture des alignements multiples de séquence. **b.** IDDT par position en fonction des modèles. **c.** Représentation de la structure de Gag à partir du modèle 4, dont les couleurs varient en fonction du score de IDDT. Les différents domaines de Gag sont représentés en carrés de couleurs.

Pour conclure, l'objectif de ces résultats est de résoudre la structure du complexe formé par la protéine Gag et un fragment d'ARNg. La technique CryoEM est privilégiée. Dans un premier temps, les conditions de préparation des grilles avec la protéine Gag seule sont mises au point et permettent de proposer un modèle à 7 Å. Les premiers tests de résolution de la structure du complexe entre Gag et l'ARNg sont réalisés. Cependant, de nouvelles mises au point sont nécessaires.

**Résumé :** différentes méthodes de résolution de structure sont utilisées dans cette étude. Tout d'abord, la cristallographie aux rayons X qui au présent n'a pas encore permis de résoudre la structure de Gag. Puis, la CryoEM qui a permis de proposer un premier modèle de la protéine 6His-Gag à 7 Å. Les premiers tests sont réalisés sur les complexes entre 6HisGag et l'ARN111<sub>WT</sub>. Les structures Alphafold des protéines 6His-Gag, Gag et GagΔp6 sont également représentées.

## II. Discussion : caractérisation des interactions *in vitro* entre le précurseur Gag et ses partenaires protéiques

Le premier objectif de ma thèse a pour but de réaliser la structure tridimensionnelle de la protéine Gag en interaction avec un fragment d'ARNg. Pour cela, la protéine Gag est produite en cellules de mammifères et est caractérisée *in vitro*. Ensuite, des conditions de formation de complexes entre la protéine Gag et l'ARNg sont analysées.

Tout au long des analyses de ce travail, la protéine Gag présentant une étiquette 6-His, dans sa partie N-terminale pour les protéines produites en cellules de mammifères et dans sa partie C-terminales pour les protéines produites en cellules bactériennes. Des travaux précédents montrent que si l'étiquette peut interagir avec des courts fragments d'ADN, l'utilisation de forces ioniques élevées empêche ce phénomène (Bewley et al., 2017). De plus, l'étude de l'interaction entre la protéine Gag et différents ARN (génomique et épissés) réalisées au sein de notre équipe montrent que cette étiquette n'altère ni l'affinité de liaison de Gag à l'ARN, ni la stœchiométrie entre les partenaires (Bernacchi et al., 2017). Néanmoins, dans notre système l'étiquette pourrait être séparée de la séquence de Gag par une région p3C permettant son excision lors des étapes de purification.

### I.1. Purification de 6His-Gag et sa caractérisation par des méthodes biophysiques

#### II.1.1 Choix des solutions tamponnées

Les vecteurs viraux pour la production de la protéine 6His-Gag étaient déjà disponibles au laboratoire. Historiquement, les protéines sont purifiées dans le tampon A1 (**tableau 6**, section Matériels et Méthodes, §IV.2.3) qui est utilisé pour solubiliser une autre protéine du VIH-1 également produite en BHK21, l'IN (Levy et al., 2016). Cependant, cette solution est fortement concentrée en NaCl et en détergent (CHAPS). Après plusieurs tests, une seconde solution tamponnée est mise au point, le tampon A3. Elle présente l'avantage de réduire la concentration de NaCl qui passe de 1 M à 500 mM et de faire disparaître le CHAPS. En effet le NaCl réduit les interactions électrostatiques entre les partenaires en solution et favorise donc la solubilisation de la protéine, de manière similaire au CHAPS, qui en recouvrant la protéine de micelles réduit les interactions entre macromolécules. Or ce travail a pour but de mettre au point les conditions pour réaliser la structure tridimensionnelle par CryoEM du complexe entre Gag et un fragment d'ARNg. La mise au point d'une solution tamponnée qui favorise la formation de complexes et qui est compatible avec la production de grilles de CryoEM est donc nécessaire.

#### II.1.2 Purification de la protéine Gag produite en cellules de mammifères

Ainsi, à partir de 12 L de culot cellulaire, la purification dans le tampon A1 de 6His-Gag permet de purifier 9,36 mg de protéines réparties en deux aliquotes. A partir de 2 L de culot cellulaire, la purification dans le tampon A3 permet de purifier 0,62 mg de protéines. La différence de rendement provient vraisemblablement de l'utilisation de

tampons aux pouvoirs de solubilisation différents. En effet des tests de purification de la protéine 6His-Gag sont réalisés à partir de production de protéine par des cellules BHK21 adhérentes. Différentes solutions tamponnées sont utilisées pour réaliser la purification de Gag sur billes (résultats non montrés). Ainsi, une plus grande quantité de protéines est purifiée dans le tampon A1, par rapport aux purifications réalisées dans le tampon A1 sans CHAPS (1M NaCl – 50 mM HEPES pH7.5, 5 mM MgCl<sub>2</sub>). Les rendements dans cette solution tamponnée sont meilleurs que ceux obtenus dans le tampon A3a, eux même meilleurs que ceux dans le tampon A3c (300 mM NaCl). La présence de CHAPS et la concentration de NaCl joue donc un rôle important pour la solubilisation de la protéine 6His-Gag produite en cellules de mammifères. Lorsque la protéine est purifiée dans le tampon A3, une limite de solubilisation est observée à partir de 0,7 mg/mL. En parallèle, la précipitation de la protéine est visible dans les tubes par la présence de particules blanches. Ainsi, malgré les grosses différences observées, il nous est apparu important de caractériser la protéine dans les deux solutions tamponnées qui permettent d'une part, d'avoir de très bons rendements et d'autre part, de réduire la concentration de NaCl et donc de favoriser la formation de complexes. Le profil biophysique de la protéine 6His-Gag en solution dans l'une et l'autre solution tamponnée est déterminé à l'aide d'un large panel de méthodes biophysiques. Ces mêmes méthodes servent aussi de contrôles lors des investigations pour la formation des complexes entre Gag et l'ARN (objectif 1) et entre Gag et les partenaires de p6 (objectif 2).

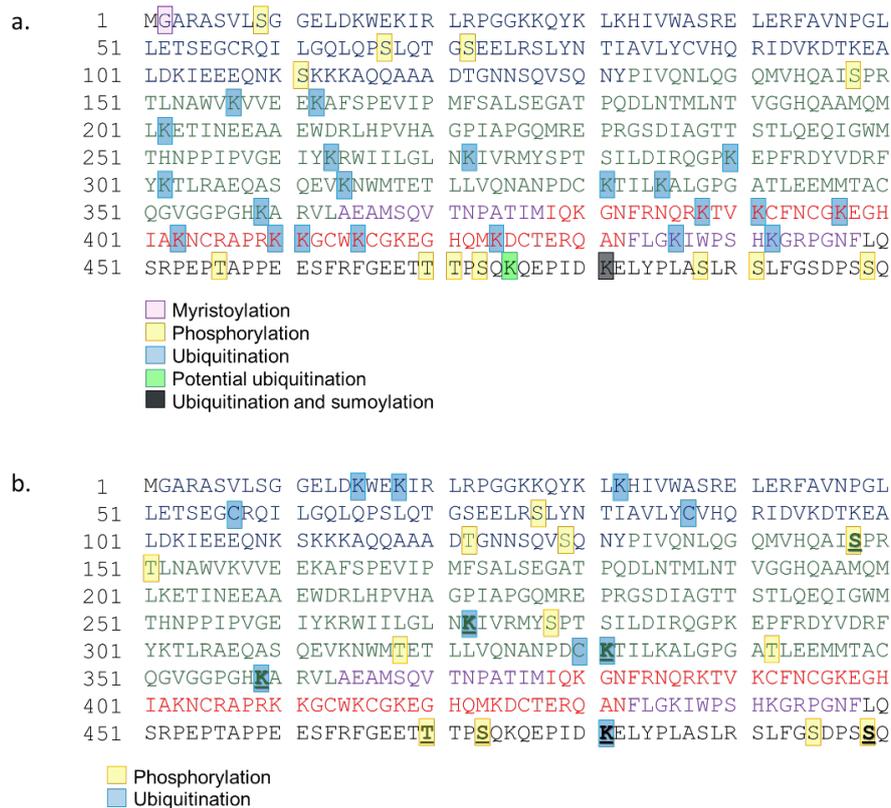
Lorsque la protéine est purifiée dans le tampon A1, à l'issue de l'étape de chromatographie à exclusion de taille, deux profils apparaissent sur SDS-Page (**figure 35 a.2**). D'une part, un profil très concentré en protéines 6His-Gag et présentant très peu de bandes contaminantes (fraction C4 sur la **figure 35, b.2**). D'autre part, un profil qui présentent un enrichissement en bandes contaminantes, représentées par des bandes ayant des tailles supérieures ou inférieures à 59 kDa (fraction C6 sur la **figure 35, b.2**). Les bandes contaminantes peuvent provenir de contaminants cellulaires capables de fixer la colonne de nickel ou d'un profil de dégradation de la protéine 6His-Gag, elle-même. Cependant, cette tendance est moins forte avec le précurseur purifié dans le tampon A3. Il est possible que le caractère moins solubilisant de ce tampon n'est pas permis de conserver solubles les protéines contaminantes ou 6His- Gag tronquées, le cas échéant.

### II.1.3 Détermination de la séquence en aa et des PTM de la protéine Gag produite en cellules de mammifères

Des WB sont utilisés pour identifier la présence de Gag dans les fractions purifiées. Le précurseur Gag produit en cellules bactériennes, déjà caractérisé au laboratoire, est utilisé comme contrôle lors des WB (Bernacchi et al., 2017; Dubois et al., 2018b; El-Wahab et al., 2014 ; **figure 37**). L'ensemble des protéines testées présentent un profil avec différentes bandes à l'issue de l'étape de révélation de la membrane par un AC dirigé contre l'étiquette. La

bande majoritaire migre à un PM attendu pour la protéine Gag (59 kDa) tandis que les autres bandes correspondent à des protéines ayant un plus faible PM. Pour aller plus loin dans l'analyse, les séquences de la protéine 6His-Gag et de ce qui compose la bande contaminante sont séquencées par MS/MS. Le recouvrement de séquences pour la protéine produites en cellules de mammifères est de 94% pour l'échantillon en solution et 98% pour l'échantillon en gel (**figure 38**). La partie N-terminale n'est pas entièrement recouverte. Cependant, lors de l'identification par WB, un anticorps dirigé contre l'étiquette 6His est utilisé. Il est alors possible de conclure que cette extrémité est intacte. Seule une région de 5 acides aminés dans la séquence du domaine NC n'est pas recouverte. ce qui suggère que certains spectres sont absents soit à cause d'un problème de digestion dépendant de la fonctionnalité de l'enzyme trypsine elle-même, soit parce que les fragments produits étaient trop petits ou trop longs due à l'espacement des résidus lysines et arginines. En conclusion, la protéine en solution semble bien correspondre à la protéine Gag. Concernant les contaminants, la bande fortement amplifiée présentant une taille inférieure à 55 kDa est découpé d'un SDS-PAGE et séquencée. De manière intéressante, la séquence de la partie N-terminale de la protéine Gag est bien recouverte pas des spectres MS/MS. Au contraire, la séquence de la partie C-terminale , n'est pas ou peu recouverte. Il est possible de conclure que les bandes présentant une taille inférieure à 59 kDa correspondent à de la protéine 6His-Gag dégradée dans le domaine de CA , s'étendant des position 133 à 363, à partir de la position 289.

Finalement, les PTM sont investiguées en se concentrant sur les ubiquitinations et les phosphorylations. Ces modifications sont particulièrement importantes pour la fonctionnalité de la protéine Gag. Leurs fonctions sont discutées dans ma revue en **annexe 1**, « *Post-Translational Modifications of Retroviral HIV-1 Gag Precursors: An Overview of Their Biological Role* » (Bussienne et a., 2020). Ainsi la protéine 6His-Gag produite en cellules de mammifères présente de nombreuses PTM (**Tableau 7 et figure 73**).



**Figure 73. Représentation des PTM de la protéine Gag en fonction de ces domaines**

La séquence du domaine de MA est représentée en bleu, celle de la CA en vert, la NC est en rouge, les espaceurs sont en violets et finalement p6 est en noir. Les phosphorylations sont surlignées en jaune, les ubiquitinations en bleu, les myristoylations en violet les potentielles ubiquitinations en vert et les positions qui peuvent être à la fois sumoylé et ubiquitinée noir. **a.** Représentation des PTM décrites dans la littérature (d'après Bussienne, et al., 2020). **b.** Représentation des PTM qui sont retrouvées dans au moins deux production de 6His-Gag différentes. Les PTM qui sont retrouvés dans la littérature et dans les productions de 6His-Gag en BHK21 sont soulignées.

Parmi ces modifications, il y a la phosphorylation de la sérine en position 148 de la CA. Cette phosphorylation de la sérine dans le motif « SP » permet l'incorporation dans les particules virales de la kinase ERK2. La déplétion de cette enzyme induit une réduction de l'infectivité notamment en inhibant l'étape de RT (Bussienne et al., 2021; Cartier et al., 1997). De plus, cette enzyme est aussi impliquée dans la phosphorylation des résidus T470, T471, S473, S488, S491 et S499 au niveau du domaine p6. L'inhibition de la phosphorylation de p6 induit l'accumulation de particules virales immatures à la MP (Hemonnot et al., 2004; Radestock et al., 2013). Le domaine p6 des protéines 6His-Gag produites en BHK21 est aussi phosphorylé dans les positions T470, S473 et le S499. De manière analogue à ce qui est observé pour le domaine p6, le caractère redondant des phosphorylations dans le domaine de MA est aussi largement décrit. Plus précisément, les phosphorylations des serines en position 9, 67, 72 et 77 sont impliquées dans la réplication virale (Jacqué et al., 1998). Le domaine de MA des protéines 6His-Gag produites

en cellules BHK21 possèdent des phosphorylations au niveau de la sérine en position 77, comme décrit dans la littérature, mais aussi en position T122 et S129.

Dans la littérature aucune phosphorylation à ce jour n'est décrite dans les domaines NC et dans les deux domaines espaceurs. Ce qui est en accord avec notre analyse sur 6His-Gag produite en BHK21. D'autre part, une plus grande variation de la distribution des ubiquitines entre ce qui est décrit dans la littérature et la protéine 6His-Gag produites en BHK21 est observable. Néanmoins, une fois encore les ubiquitinations présentent des activités redondantes et leur rôle exact pour le cycle viral est actuellement largement méconnu. Ainsi les domaines de CA et NC sont largement décrits comme étant mono-ubiquitinés (Sette et al., 2010). Alors que la protéine 6His-Gag produites en BHK21 présente des nombreuses ubiquitinations dans les domaines MA et CA. De plus, de manière surprenante des ubiquitinations sont identifiées sur des résidus cystéines, les spectres obtenus à l'issue de l'analyse MS/MS sont contrôlés un par un, rendant cette observation fiable. Cependant, aucune ubiquitination des cystéines de Gag n'est décrite dans la littérature. Néanmoins, ces PTM sont moins stables car la liaison de l'ubiquitine se fait au niveau des groupes thiols des résidus cystéines et ces interactions sont souvent perdues lors des analyses d'analyse (McDowell and Philpott, 2013).

En général, les PTM retrouvées dans la production de 6His-Gag en BHK21 sont en accord avec ce qui a été précédemment décrit dans la littérature. Ces PTM devraient assurer le bon fonctionnement de la protéine. Par la suite, il serait intéressant de réaliser de nouvelles analyses pour caractériser les autres PTM (comme la SUMOylation, ISGylation). De plus, l'étiquette 6His est délibérément ajoutée dans la partie N-terminale de la protéine pour empêcher sa myristoylation. Ainsi, la protéine étudiée dans ce travail ne présente pas cette modification.

#### II.1.4 Caractérisation de l'état oligomérique de la protéine Gag produite en cellules de mammifères

Un des challenge de ce travail est de déterminer les conditions expérimentales qui permettent de conserver la protéine 6His-Gag soluble en solution. C'est pourquoi, un panel de méthodes biophysiques est utilisées pour caractériser l'état de la protéine dans les différentes solutions tamponnées (A1 ou A3), que la protéine soit directement purifiée dans l'une ou l'autre solution ou d'abord purifiée dans le tampon A1 puis dialysée contre le tampon A3.

Pour commencer, le profil DLS de la protéine Gag produite en cellules bactériennes montre une espèce majoritaire en solution avec un Rh de 4,76 nm. En appliquant la loi de Stokes-Einstein cela correspondrait au PM de 126 kDa, soit un dimère de Gag. Ces résultats sont en bon accord à ce qui a déjà été observé dans notre laboratoire (Dubois et al., 2018b; El-Wahab et al., 2014). Lorsque la protéine Gag est produites en cellules de mammifères et purifiée

dans le tampon A1 ou A3 elle présente une tendance à l'agrégation (observable sur le profil d'intensité et sur l'autocorrélogramme des **figure 41 et 42**). Lorsqu'elle est purifiée dans le tampon A1, le pic majoritaire dans la représentation en masse possède un Rh de 1,63 nm, ayant un PM de 6 kDa. Ce pic, pourrait correspondre à des micelles de CHAPS car la concentration du détergent dans le tampon A1 est de 7 mM, supérieure à sa CMC (Giudice et al., 2022). De plus, un échantillon ne contenant que du tampon est mesuré par DLS (**annexe 4**). Son profil, présente bien un pic de 1,26 nm. Cependant, lorsque la protéine 6His-Gag est purifiée dans A1, une autre espèce est retrouvée en solution avec un Rh de 6,09 nm dans les distributions en intensité et en masse. Elle présente un PM de 232 kDa qui pourrait correspondre à un tétramère de Gag ou bien à un monomère-dimère recouvert de CHAPS.

En revanche, lorsque la protéine est purifiée dans le tampon A3 elle présente deux types de profils. D'une part, le pic majoritaire observable dans la distribution en masse présente un Rh de 4,56 nm correspondant à un PM de 130 kDa, soit un dimère de 6His-Gag. D'autre part, d'autres analyses montrent la présence au niveau de la distribution en masse, d'un pic majoritaire ayant un Rh de 3,79 nm correspondant à un PM de 76 kDa, sûrement un monomère de Gag (ou un dimère). La divergence entre les PM théorique et expérimentaux mesurés par DLS est probablement due à l'hypothèse imposée par le modèle de Stokes-Einstein, qui suppose que les protéines sont sphériques. En outre, aucune corrélation avec le temps de décongélation, la température de conservation ou les tubes utilisés peut expliquer cette différence de profils, il est possible que les protéines 6His-Gag produites en cellules de mammifères et purifiées dans le tampon A3 soient en équilibre entre des conformations différentes. Des résultats allant dans ce sens sont obtenus lorsque la protéine 6His-Gag est purifiée dans le tampon A1 est dialysée avec le tampon A3 (**figure 43**). En effet la réalisation d'une dialyse à 4°C sur la nuit permet de réduire la concentration de CHAPS jusqu'à faire disparaître le pic correspondant aux micelles et le du pic majoritaire observé dans le tampon A1 passe de 6,09 nm à 5,5 nm après dialyse correspondant à un PM de 166 kDa qui correspondrait à un trimère de Gag.

Des analyses de SEC-MALS sont réalisées en parallèle et montrent que dans le tampon A1, 85% des molécules en solution seraient de monomères et que 15% seraient des micelles de détergent (**figure 44 et 45**). De plus la protéine est relativement monodisperse et semble majoritairement monomérique dans cette solution tamponnée. Tandis que lorsque la protéine 6His-Gag est purifiée dans le tampon A3 elle présente un pic asymétrique et un PM expérimental variant de 77 kDa à 59 kDa (**figure 46**). Il semblerait que ce pic corresponde à un équilibre entre des monomères et des dimères. La masse expérimentale de 77 kDa correspondrait, en effet à un mélange de monomères (59 kDa) et de dimères (118 kDa) de Gag, avec une prévalence de monomères. Ainsi la protéine est

majoritairement monomérique en solution dans les tampons A1 et A3 et dans l'ensemble les résultats de SEC-MALS sont cohérents avec ceux obtenus en DLS. Pour finir, un dernier contrôle est réalisé en utilisant de l'AUC (**figure 47**). De manière intéressante, l'AUC montre la présence d'un dimère pour la protéine purifiée dans le tampon A3 et en plus faible quantité pour la protéine dans le tampon A1. Ces résultats indiqueraient qu'il existe en solution un équilibre entre une majorité de monomères et une minorité de dimères. Cet effet serait plus fortement observable dans le tampon A3, qui est la solution tamponnée qui favoriserait d'avantage les interactions entre macromolécules. Ces résultats sont cohérents avec d'autres études menées sur la protéine Gag (Datta, 2007 ; Dubois, 2018). De plus, l'AUC permet aussi de calculer le rapport de frictions ( $f/f_0$ ) de la protéine Gag. Le rapport  $f/f_0$  peut être défini comme le rapport entre le coefficient de frottement friction ( $f$ ) de la molécule lors de la sédimentation par rapport au coefficient de friction théorique ( $f_0$ ). Le coefficient de friction théorique pour une sphère pleine, non hydratée serait de 1,0. Les déviations de cette valeur indiquent une asymétrie ou une hydratation croissante de la molécule. De nombreuses protéines globulaires ont des valeurs  $f/f_0$  comprises entre 1,05 et 1,35 (Garbette, 2010). Ainsi, les valeurs  $f/f_0$  calculées par le logiciel sont de 1,10 pour 6His-Gag purifiée dans A1, de 1,20 pour 6His-Gag purifiée dans A3 et finalement de 1,48 pour 6His-Gag purifiée dans A1 dialysée dans A3 (**annexe 2**). Ainsi la protéine serait davantage globulaire en présence de fortes concentrations de NaCl et de détergent par rapport aux autres conditions. En revanche la dialyse aurait un effet sur la conformation de la protéine et induirait aussi la perte des populations de dimères. Dans la littérature, l'utilisation d'un mutant incapable de dimériser de la protéine Gag, muté dans le domaine CA (nommé Gag WM), présente un rapport  $f/f_0$  de 1,6 (Datta et al., 2007).

Pour contrôler l'impact du détergent et de la concentration de NaCl sur la stabilité thermique de Gag, des analyses de NanoDSF sont réalisées (**figure 48**). De manière surprenante les  $T_M$  estimées pour 2 analyses différentes réalisées sur 6His-Gag dans le tampon A1 sont de  $36,5 \pm 3^\circ\text{C}$  et de  $46,5 \pm 3,2^\circ\text{C}$ . Lorsque la protéine est purifiée dans le tampon A3, nous observons un  $T_M$  de  $44,1 \pm 0,6^\circ\text{C}$ . Il est possible que la conformation de Gag ait un impact sur la stabilité de la protéine.

## II.2. Analyse de l'interaction de 6His-Gag avec la région 5'UTR de l'ARNg viral in vitro

### II.2.1 Contrôle de la dimérisation de l'ARN

La formation du complexe entre Gag et l'ARNg repose sur la dimérisation préalable de l'ARN. Mon laboratoire a mis au point le protocole de dimérisation de l'ARN (Paillard 1996). Comme les solutions tamponnées utilisées pour la protéine Gag produites en cellules de mammifères sont différentes des tampons habituellement utilisés pour dimériser l'ARN, la première étape de ce travail consiste à contrôler la dimérisation des ARN111<sub>WT</sub> et ARN135 ainsi

que l'absence de dimérisation pour le mutant ARN111<sub>AAA</sub> dans les tampons A1 et A3. La formation des complexes est contrôlée en utilisant différentes méthodes nécessitant le marquage ou non des ARN. Ainsi le contrôle de la dimérisation des ARNg est réalisé à la fois sur des ARN non marqués et à la fois sur des ARN marqués par l'ajout d'une fluorescéine 5-thiosemicarbazide (FTSC) au niveau de l'extrémité région 3'. D'une part, le gel TBM permet de conclure que les ARN111<sub>WT</sub> et ARNSL1<sub>35</sub> sont dimériques, tandis que l'ARN111<sub>AAA</sub> migre comme un monomère d'ARN de 111 nucléotides (**figure 49**). Le marquage par un chromophore n'induit aucune différence de migration. Cependant, lorsqu'une grande quantité d'ARN111<sub>WT</sub> (ici marqué) est déposé sur gel TBM, des bandes correspondant à des espèces de plus grande taille apparaissent. Les analyses d'AUC dans le tampon A3 identifient une majorité d'ARN111<sub>WT</sub> non marqués sont dimériques, bien qu'une plus petite population soit capable de multimériser (**figure 50**). Cependant, elle représente une proportion inférieure à 10% du signal. Il est possible que les bandes observées sur le gel TBM correspondent à ces formes multimériques. Ainsi l'ARN est majoritairement dimérique en solution dans les tampons A1 et A3 (**figures 50 et 51**).

### II.2.2 Formation de complexes entre Gag et l'ARN sans marquage

Les complexes entre 6His-Gag et l'ARN111<sub>WT</sub>, sont formés au ratio de 1 dimère d'ARN pour 6 molécules de Gag (1 : 6) comme décrit dans la littérature (Dubois et al., 2018b). Pour favoriser les interactions entre les deux partenaires des tests sont réalisés dans le tampon A3 dont la concentration de NaCl est diminuée après avoir ajouté l'ARN111<sub>WT</sub> replié et dimérique. Des résultats préliminaires montrent que la protéine seule dans de telles concentrations en NaCl a tendance à former des agrégats de grandes tailles (**figure 52**). Or, il est possible que la formation du complexe entre la protéine et l'ARN stabilise cette dernière. Ainsi, lorsque l'ARN111<sub>WT</sub> est ajouté à la protéine dans les tampons A3a (500 mM NaCl) et A3b (400 mM NaCl), une nouvelle population présentant un Rh de 5,38 nm (A3a) et 11,97 nm (A3b) apparaissent (**figure 53**). Les PM associés à ces rayons pourraient correspondre à un PM typique d'un trimère ou d'un hexamère de Gag ou alternativement au PM d'UN complexe entre Gag et l'ARN111<sub>WT</sub>. Ainsi la diminution de la concentration de NaCl favorise l'interaction entre Gag et l'ARN. De plus, ce pic n'est pas apparent quand la protéine est seule en solution et par conséquent ces résultats tendent à montrer qu'un complexe s'est formé. Dans le tampon A3c (qui correspond au tampon avec la plus faible concentration en NaCl, 300mM NaCl) on observe la présence des complexes de très grande taille (**figure 54**). De manière intéressante, dans ces conditions salines, le pic correspondant à la protéine monomérique (Rh=3,2nm) présent lorsque la protéine est seule en solution a totalement disparue. En outre, lors d'un mélange d'ARNt avec 6His-Gag dans les mêmes conditions salines, aucun pic correspondant à l'interaction protéine-ARN n'est observée. Ainsi la formation des complexes de très grande taille dans le tampon A3c serait spécifique à l'interaction entre Gag et l'ARN111<sub>WT</sub>. Par la suite, des

analyses de vitesse de sédimentation en AUC sont réalisées pour déterminer la meilleure stœchiométrie de formation des complexes protéine-ARN. Comme des contrôles avec la protéine seules sont déjà réalisées dans les tampons A1 et A3, ce sont ces tampons qui sont utilisés pour réaliser les complexes. L'alignement des profils obtenus lors des analyses d'AUC ne permettent pas de conclure sur la présence d'un complexe entre Gag et l'ARN111<sub>WT</sub> dans ces tampons (**figure 55**). A l'avenir des tampons A3a, b,c , ainsi que différentes stœchiométries protéine/ARN seront testés.

### II.2.2 Formation de complexes entre ARNg et la protéine 6His-Gag marquée par FIDABio

La technologie FIDABio repose sur le marquage d'un partenaire pour pouvoir étudier la formation de complexes. Es étape de marquage ont nécessité de nombreuses mises au point permettent de générer très peu de pertes (3%) et de maximiser le pourcentage de marquage qui varie quand même de 50% à 80%. L'un des inconvénients de cette méthode réside dans la présence des sondes libres en solution avec la protéine qui perturbent fortement le signal. Le marquage de 6His-Gag par l'ajout de sondes Alexa génère un grand nombre de sondes libres (*free dye*). Cependant, cette méthode innovante permet de tester de nombreuses conditions en quelques minutes. Nous avons pu tester une large gamme de stœchiométries pour former les complexes entre 6His-Gag et l'ARN111<sub>WT</sub>. Ainsi, lorsque la protéine marquée est mélangée à l'ARN111<sub>WT</sub>, le Rh du possible complexe augmente au fur et mesure qu'augmente la stœchiométrie entre les partenaires (**figure 56**). La courbe n'atteint pas le plateau nécessaire pour calculer la constante de dissociation (Kd). Le Rh de la protéine seule est de 4,5 nm en moyenne, en ajoutant l'ARN111<sub>WT</sub> il passe à 7,87 pour une stœchiométrie de 1 : 40. Cependant, ce genre d'analyses sont difficilement reproductibles et la méthode FIDABio n'est pas optimale pour l'analyse des interactions entre Gag et les ARNg. En effet son fonctionnement repose sur le déplacement de la protéine marquée dans un capillaire qui est soit recouvert de charges négatives, soit recouvert de charges positives. Pour contrecarrer cette problématique de nouveaux capillaires sont commercialisé. Ils n'étaient pas encore disponibles au moment des analyses ils. A ce stade, des analyses de marquage de l'ARN111<sub>WT</sub> (utilisé notamment pour les analyses d'anisotropie) sont envisagés. Le marquage de l'ARN plutôt que celui de la protéine présente plusieurs avantages : tout d'abord l'efficacité de marquage est bien meilleure avec l'ARN. De plus, la quantité initiale de matériel utilisée par analyse FIDABio, est beaucoup plus faible. En effet 50 nM sont suffisant pour avoir un signal correct (contre au minimum 350 mM de Gag). Le Rh de l'ARN111<sub>WT</sub> dimérique est très bien reproductible par cette méthode puisqu'un Rh de  $6,52 \pm 0,18$  nm pour 9 répétitions dans le tampon A3 est retrouvé.

### II.2.3 Formation de complexes entre protéine 6His-Gag et l'ARNg marquée par anisotropie de fluorescence

Finalement, les complexes entre 6His-Gag et l'ARNg sont également étudiés par anisotropie de fluorescence. A l'issue des analyses précédemment réalisées, les complexes sont formés dans le tampon A3b, pour lequel la concentration de NaCl est de 400 mM. Ces analyses nécessitent des étapes de marquages. Ainsi, l'efficacité de marquage pour chaque ARNg est contrôlée à l'aide de gels dénaturants urées. Les gels présentés, **figure 57**, montrent que le marquage est efficace et que l'étape de purification l'est tout autant pour les trois ARN (ARN111<sub>WT</sub>, ARN111<sub>AAA</sub> et ARNSL1<sub>35</sub>). Les premiers complexes sont formés entre Gag et l'ARN111<sub>WT</sub>. L'étude de la variation de l'anisotropie par rapport à la concentration de Gag (en nM) rajoutée à une quantité fixe d'ARN marqué, permet en appliquant le modèle de Scatchard, de déterminer que le kD entre les partenaires. Dans ce cas notre ajustement a permis de déterminer la valeur de  $(174 \pm 41)$  nM. La même analyse avec l'ARNSL1<sub>35</sub>, qui correspond au site minimum d'interaction avec Gag, permet dans les mêmes conditions de déterminer un kD de  $(116 \pm 36)$  nM. Finalement, cette analyse est réalisée avec l'ARN111<sub>AAA</sub> incapable de dimériser. Dans ce cas, aucun complexe n'est formé. Ainsi, la protéine 6His-Gag produites en cellules de mammifères interagit spécifiquement avec l'ARNg dimérique. De plus, elle présente une affinité très similaire pour la région génomique de 111 nucléotides que pour la tige-boucle SL1 de 35 nucléotides.

La plupart des analyses réalisées dans la littérature sont basées sur une protéine Gag produite en cellules bactériennes (McKinstry et al., 2014). Elles mettent en évidence que la protéine Gag est capable de fixer des ARN viraux et cellulaire avec une affinité allant de 2 à 45 nM, mais n'interagit spécifiquement qu'avec l'ARNg (Bernacchi et al., 2017; Dubois et al., 2018b; El-Wahab et al., 2014; Smyth et al., 2015) et notamment l'ARN111<sub>WT</sub> avec une plus haute affinité, de 2,5 nM. Des études de spectroscopie réalisées au laboratoire montrent que Gag produite en cellules bactériennes interagit avec l'ARNSL1<sub>35</sub> et ARN111<sub>WT</sub> avec une affinité proches (Bernacchi et al., 2017). De plus, il est clairement mis en évidence que la boucle interne de SL1 (AGG) correspond au site minimal d'interaction avec la protéine Gag. En effet sa substitution ou sa délétion ont un effet dramatique sur l'interaction entre Gag et l'ARN (Bernacchi et al., 2017; El-Wahab et al., 2014; Smyth et al., 2015). Ainsi les résultats obtenus avec la protéine produite en cellules de mammifères sont cohérents avec ce qui a préalablement été démontré.

### II.3. Analyse structurale de Gag en complexes avec l'ARN

#### II.3.1 Analyse structurale de 6His-Gag par cristallographie aux rayons X

L'avantage de la cristallographie aux rayons X est que cette technique permet de résoudre des structures à résolution atomique. Cependant, elle repose sur la capacité à cristalliser de la protéine qui dépend notamment de sa solubilité et de sa monodispersité. En outre, la polyprotéine Gag possède des domaines flexibles et intrinsèquement désordonnés, notamment les deux espaceurs et le domaine p6, qui peuvent rendre difficile sa cristallisation. De plus, afin d'obtenir de cristaux il est nécessaire de travailler à des très fortes concentrations de protéine. Or, si la protéine 6HisGag produite en cellules de mammifères se concentre bien dans le tampon A1, cette solution tamponnée est fortement concentrée en NaCl et en détergent. Ces éléments peuvent impacter la formation du cristal. Tandis que la protéine purifiée ou dialysée dans le tampon A3 ne se concentre pas au-delà de la limite de 0,7 mg/mL. Ainsi des tests sont réalisés sur la protéine dialysée 4 heures à 4°C. Aucun cristal protéique ne s'est formé même après quelques mois.

En outre, l'ajout de l'ARNg pourrait avoir un rôle stabilisateur pour la protéine. Différentes tailles d'ARN sont disponibles au laboratoire, ARN111<sub>WT</sub>, ARNSL1<sub>35</sub> mais aussi un ARN viral contenant les premiers 600 nucléotides de l'ARNg correspondant à l'extrémité 5'UTR jusqu'au début du gène *gag* (Bernacchi et al., 2017). Ainsi d'autres tests de cristallisation pourront être réalisés.

#### II.3.2 Mise au point des conditions d'acquisition de 6His-Gag par CryoEM

Dans un premier temps, des tests sur des grilles de carbone par coloration négative sont réalisés. Des résultats prometteurs mettent en évidence que la protéine serait monomérique et présenterait un rayon de 3,7 nm en moyenne de manière comparable à ce qui est déterminé par DLS et AUC (**tableau 8**, partie résultat §I.1.5) Ainsi, des analyses de CryoEM sont réalisées nécessitant de nombreuses mises au point autant sur les supports dont les grilles sont formées, que la préparation de celles-ci. Elles sont récapitulées **tableau 16** (résultat partie III. §1.3.2). A l'issue de ces tests deux types de grilles sont sélectionnés. D'une part, les grilles de carbone C-flat, qui permettent de réaliser le modèle présenté en **figure 63**. Elles favorisent la répartition de la protéine Gag dans les trous et permettent de générer une glace homogène. D'autre part, des grilles d'or supplémentées avec du PEG, présentant des résultats encourageants quant à la répartition des protéines et à l'homogénéité de la glace. Une acquisition est réalisée sur ces grilles, mais un problème technique sur le microscope pendant l'acquisition a empêché le traitement approfondi des données. A ce stade, la préparation d'un grand nombre de grilles est nécessaire car une grande variabilité de la qualité des grilles est observée, même lorsque le même échantillon, dans les mêmes conditions de préparation est utilisé. Ce phénomène serait dépendant du Vitrobot. En effet des tests sur l'étape de

séchage et notamment sur la durée et la force de séchage (*blotting*) sont réalisés. Il est attendu que, plus l'étape de séchage est longue et plus la force est élevée, plus les grilles devraient être sèches et la glace fine. Des conditions optimales ont été déterminées mais sont peu reproductibles. Ainsi pour augmenter la reproductibilité de la préparation des grilles, le Chameleon nouvellement acquis à l'IGBMC est utilisé. L'avantage de ce dernier par rapport au Vitrobot, est que toutes les étapes de manipulation des grilles sont automatisées (Levitz et al., 2022). De plus l'étape de *blotting* est évitée. Les grilles produites avec Gag ne sont pas homogènes. Cela provient sûrement d'un certain nombre de facteurs reliés à la manipulation de l'échantillon. Néanmoins, le Chameleon permet de réaliser facilement de nombreuses mises au point comme la durée d'activation des grilles de carbone par GD et la possibilité de réduire le temps entre l'étape de GD et l'immersion de la grille dans l'éthane liquide. De plus, l'échantillon n'est pas fixé. Ainsi, il serait intéressant de réaliser de nouvelles étapes de mise au point en utilisant cette machine.

### II.3.3 Analyse de la structure de 6His-Gag par CryoEM

Une acquisition à partir d'une grille C-flat, est réalisée sur le microscope Glacios. Elle permet de résoudre la structure de Gag et de proposer un modèle à 7 Å. Le Glacios est favorisé par rapport au microscope Titan, car il permet d'augmenter le contraste entre les particules de Gag de 59252 Da et le bruit de fond grâce à son canon à émission de champs de 200 kV. Dans l'ensemble ce premier modèle nous permet d'observer que la protéine 6His-Gag serait sous forme d'un monomère de dimensions 50 x 50 x 90 Å dans le tampon A3. Ces dimensions sont en bon accord avec les Rh déterminés par DLS, AUC et FIDAbio. De plus, la protéine est relativement globulaire de manière comparable aux valeurs de rapport  $f/f_0$  mesurés par AUC. Ainsi ce modèle est conforme au profil biophysique réalisé lors de cette étude. Néanmoins, il serait nécessaire d'augmenter le nombre de particules pour améliorer la résolution. Pour cela, de nouvelles acquisitions sont indispensables.

Si toutefois les conditions ne permettent pas d'augmenter suffisamment la résolution de la protéine, d'autres stratégies peuvent être utilisées. En effet la plateforme de biologie structurale de l'IGBMC a mis au point des grilles de carbones recouvertes d'un cristal 2D de streptavidine (Crucifix et al., 2004). L'utilisation d'une protéine modifiée avec de la biotine pourrait permettre son immobilisation et par conséquent, un meilleur contrôle de la concentration et du nombre de particules par trou.

### II.3.6 Analyse de la structure du complexe formé par 6His-Gag et l'ARN111<sub>WT</sub> par CryoEM

Afin de réaliser les premières grilles de CryoEM contenant le complexe entre la protéine 6His-Gag et l'ARN111<sub>WT</sub> dimérique, les mêmes conditions que celles utilisées pour réaliser le modèle de 6His-Gag seul sont appliquées. Ainsi, des grilles C-flat sont chargées avec des complexes formés d'une molécule de Gag pour 10 dimères

d'ARN111<sub>wt</sub> (1 : 10). La solution composée des complexes est fixée au glutaraldéhyde et du DDM est ajouté juste avant le dépôt sur grille, après la formation du complexe, qui est quasiment immédiate. Les effets de la fixation et du DDM sont contrôlés par DLS dans le tampon A3b (400 mM ; **figure 68**). En absence de fixation mais après ajout de DDM, un décalage de Rh est observé, il passe de 4,75 nm correspondant à la protéine dimérique à 9,32 nm. Quand l'échantillon est fixé puis supplémenté par du DDM, le Rh passe de 4,75 nm à 13,36 nm. Les analyses de DLS réalisées précédemment indiquent un Rh de 11,97 nm pour le complexe entre Gag et l'ARN (**figure 54 et tableau 10**). Or ces analyses sont réalisées sans détergent et le DDM, comme le CHAPS, peut augmenter le Rh de la particule en se fixant autour. Ainsi l'ajout de DDM après fixation ne semble pas perturber la formation du complexe entre Gag et l'ARN. Cependant, le complexe ne représente qu'une partie minoritaire du signal par DLS (**figure 68**). Le même profil est retrouvé sur les grilles (**figure 69**). En effet de nombreux agrégats sont présents mais peu de complexes Gag/ARN isolés. Les conditions dans lesquelles les complexes sont formés (tampon A3) ne semblent pas induire une organisation ordonnée de Gag, sous forme de VLP par exemple (Chen and Cheng, 2017; Datta et al., 2007; McKinstry et al., 2014). Il est donc nécessaire de trouver des conditions qui favorisent la formation des complexes. L'utilisation des grilles or PEGylée, notamment, permettraient de réduire la concentration initiale de protéines. Moins la protéine est concentrée, moins elle devrait être agrégée sur les grilles. De plus, il serait possible d'utiliser les grilles ayant des cristaux 2D de streptavidine. En effet la fixation de Gag sur les cristaux, puis l'interaction avec l'ARN, pourrait permettre de promouvoir la formation des complexes. Pour cela des études préalables de la formation des complexes quand l'un des partenaires est fixé pourraient être réalisées par SPR (Biacore™). Finalement, des mutations favorisant la stabilité comme la mutation dans le domaine SP1-T8I pourraient être envisagées (Mendonça et al., 2021).

#### [1.3.4 Proposition d'un modèle de structure de Gag en utilisant le programme Alphafold](#)

Cette étude utilise l'outil informatique Alphafold pour produire et comparer différents modèles de la protéine Gag. Ainsi, la protéine Gag sauvage, la protéine 6His-Gag et la protéine dépourvue du domaine C-terminal p6 présentent de nombreuses homologies, notamment concernant la structure de chaque domaine. En effet le score de prédiction pour la structure de MA et de la CA dans les différentes protéines Gag est très bon (>80). De la même façon, le score attribué à la structure de la NC est plutôt bon (<60). Cependant, les régions désordonnées entre les domaines présentent des scores mauvais, ce qui signifie que l'organisation 3D de chaque domaine est plutôt bien définie, alors que l'organisation 3D des domaines entre eux n'est pas déterminée par le logiciel. Néanmoins, il ressort que les modèles de la structure 3D de Gag sans partenaire semblent plutôt globulaires ce qui est en bon accord avec notre analyse structurale précédente.



## **Résultats et discussion**

### **objectif n°2**



## **I. Résultats Objectif n°2 : Caractérisation des interactions *in vitro* entre Gag et ses partenaires protéiques**

De la sélection à l'encapsidation de l'ARNg dans les particules virales, des changements de spécificités de liaisons de la protéine à l'ARN viral sont nécessaires (§objectifs). Ainsi l'interaction entre Gag et l'ARNg doit être fortement spécifique au moment de la sélection puis devenir non spécifique lors de l'assemblage de la particule virale à la MP. Notre hypothèse repose sur l'existence d'un interrupteur moléculaire régulant les changements de spécificités d'interactions entre Gag et l'ARNg. Des résultats préliminaires démontrent que le domaine C-terminal de Gag, p6 joue un rôle clé dans la spécificité de l'interaction avec l'ARNg (Dubois et al., 2018b). Ce domaine est aussi impliqué dans le recrutement de protéines cellulaires (TSG101 et ALIX) et virale (Vpr) au moment du bourgeonnement. Ainsi, l'un des objectifs de ce travail est de déterminer les conditions optimales favorisant l'interaction de Gag avec les partenaires de p6 Gag et l'ARNg. Pour cela, les partenaires cellulaires (UEV et ALIX) et le partenaire viral (Vpr) de p6 sont produits dans les cellules mammifères BHK21. Comme pour la protéine Gag, des caractérisations biophysiques sont nécessaires ainsi que des tests visant à optimiser les conditions permettant la formation des complexes.

### **I.1. Production des vecteurs viraux permettant l'expression des partenaires de Gag en cellules de mammifères**

#### **I.1.1 Production des virus recombinants**

Les virus recombinants sont produits comme décrit dans la section matériels et méthodes, §V.2. A l'issue de trois étapes de sélection, la majorité des virus exprimés dans les cellules BHK21 adhérentes n'expriment plus de RFP, signifiant que la recombinaison homologue a eu lieu. Après une étape d'amplification virale, des productions sur des cellules adhérentes sont réalisées. Différentes solutions tamponnées sont testées pour purifier chaque partenaire (**tableau 6**, section matériels et méthodes, §IV.2.3). Les purifications des différentes protéines sont réalisées à l'aide de billes magnétiques MagStrep et l'expression des protéines est contrôlé par des analyses par SDS-PAGE et des WB.

### I.1.2 Contrôle de l'expression par les virus recombinants et optimisation de la purification des protéines partenaires

Ainsi, pour les trois partenaires les cellules sont récupérées et lysées dans une solution tamponnée de lyse. La solubilisation des protéines dans le tampon A1 et A3 (**tableau 6**, section matériels et méthodes, §IV.2.3) est comparée pour chacun des partenaires. Le lysat est alors directement mélangé avec les billes. A l'issue d'une incubation d'1 heure à 4°C les billes sont lavées et éluées. Des échantillons de cellules, de lysat, d'effluent, du lavage, des deux premières étapes d'éluion et les billes sont déposés sur un SDS-PAGE 15% pour Vpr (**figure 74**) et le domaine UEV (**figure 75**) et un SDS-PAGE 10% pour ALIX (**figure 76**). Une bande correspondant aux billes est visible sur gel dénaturant à la taille de 13,5 kDa (IBA). Des analyses par WB sont réalisées en parallèles des gels dénaturants pour contrôler la séquence des protéines.

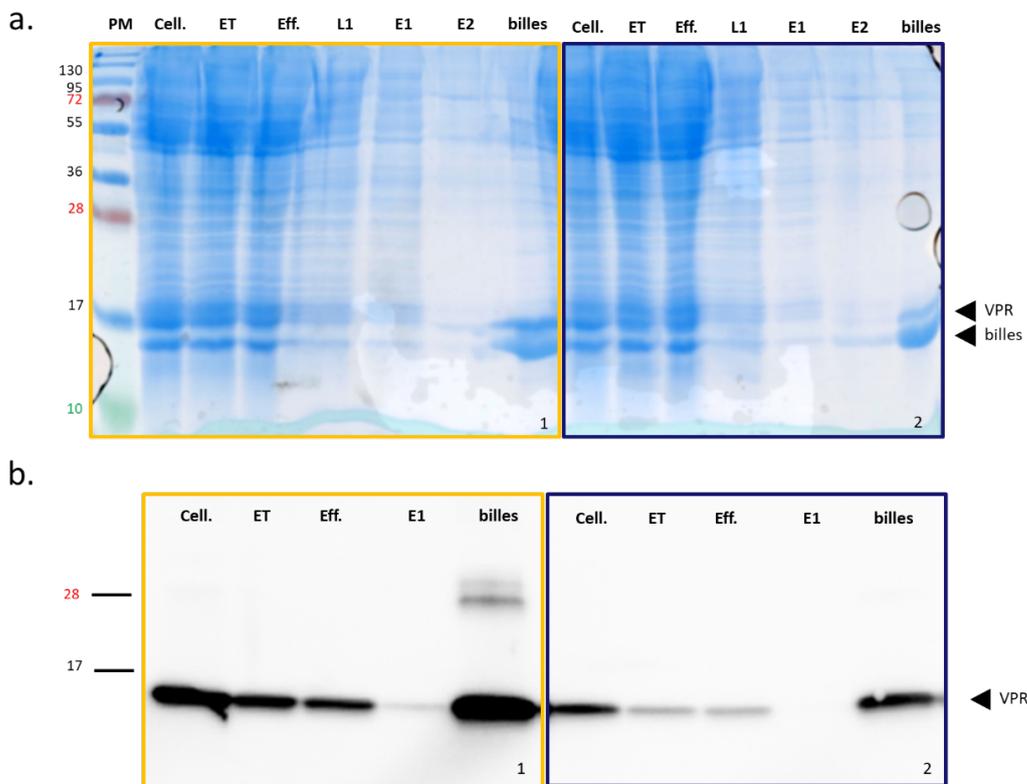
	UEV	Vpr (BHK21)	Vpr (synthétique)	ALIX
<b>PM (Da)</b>	20367	15132	11395	99776
<b>Coefficient d'extinction molaire à 280 nm (<math>M^{-1}.cm^{-1}</math>)</b>	37025	31970	20970	71795

**Tableau 17. PM théoriques des partenaires de p6 et de leurs coefficients d'extinction molaires**

*Le PM est calculé à partir de la séquence des différentes protéines à l'exception de la protéine Vpr produite par synthèse chimique pour laquelle il est obtenu par spectroscopie de masse electrospray (Cornille et al., 1999).*

R-Vpr

Les fractions issues de la purification de la protéine présentent deux bandes d'un PM d'environ 15 et 17 kDa sur SDS-PAGE 15%. Le WB confirme la présence d'une protéine avec un *Strep-tag*<sup>®</sup> dans les cellules, le lysat, l'effluent et accrochée aux billes. Dans le cas de la purification dans le tampon A3, une faible proportion de protéines est éluée alors qu'une grande quantité reste accrochée sur les billes. En comparant les profils des WB, le signal correspondant aux AC fixés aux protéines Vpr semblent plus fort dans le tampon A1 par rapport au tampon A3 (**figure 74**).

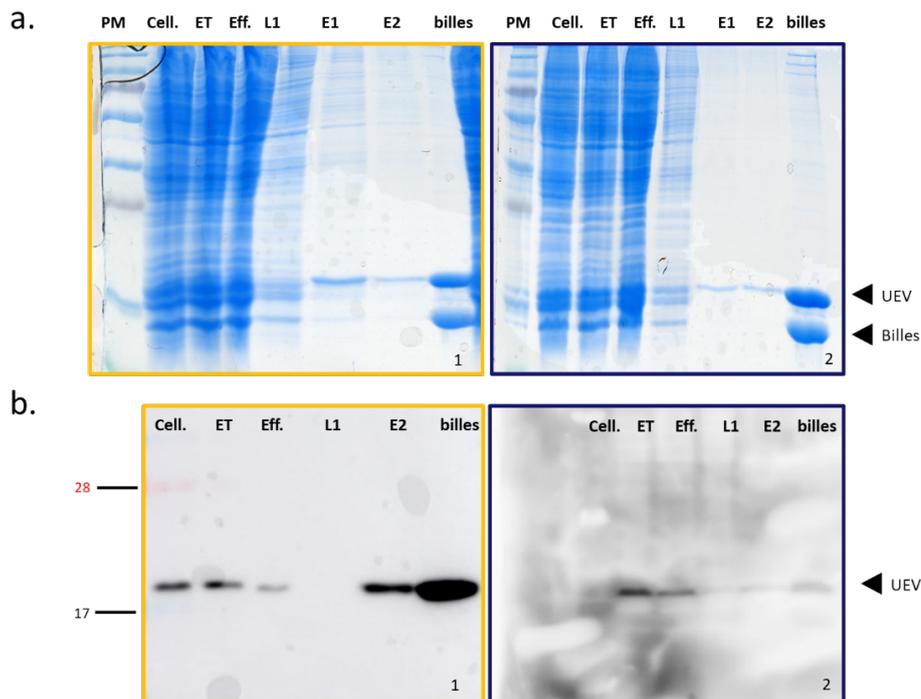


**Figure 74. Analyse par SDS-PAGE de la purification de Vpr et détermination de sa séquence par WB dans les tampons A1 et A3**

Les fractions correspondant à la purification de R-Vpr dans le tampon A1 sont encadrées en jaune (1) et celles à la purification de R-Vpr dans le tampon A3 sont encadrées en bleu (2). **a.** SDS-Page 15% permettant l'analyse de la purification par affinité à l'aide de billes MagStrep Type 3 XT. Le marqueur de poids moléculaire (PM), un échantillon des protéines avant sonication (cell), après sonication (ET), l'effluent (Eff) sont déposés dans les quatre premiers puits du gel. Le premier lavage, les deux premières éluions et les billes se trouvent dans les derniers puits. **b.** Identification de la protéine par WB en utilisant un anticorps dirigé contre l'étiquette *Strep-tag*<sup>®</sup>. Un échantillon des protéines avant sonication (cell), après sonication (ET) et l'effluent (Eff) sont déposés dans les trois premiers puits du gel. Ils sont suivis d'une fraction de la première élution et des billes.

## R-UEV

Les fractions d'élution et correspondant aux billes, issues de la purification du domaine UEV présentent une bande d'environ 20 kDa après migration sur SDS Page 15%. La réalisation d'un WB confirme la présence d'une protéine avec un *Strep-tag*<sup>®</sup> dans les cellules, le lysat, l'effluent, la fraction d'élution et sur les billes. La quantité de protéines déduite après leur détection par WB dans le tampon A3 est nettement inférieure à celle observée dans le tampon A1 (**figure 75**).

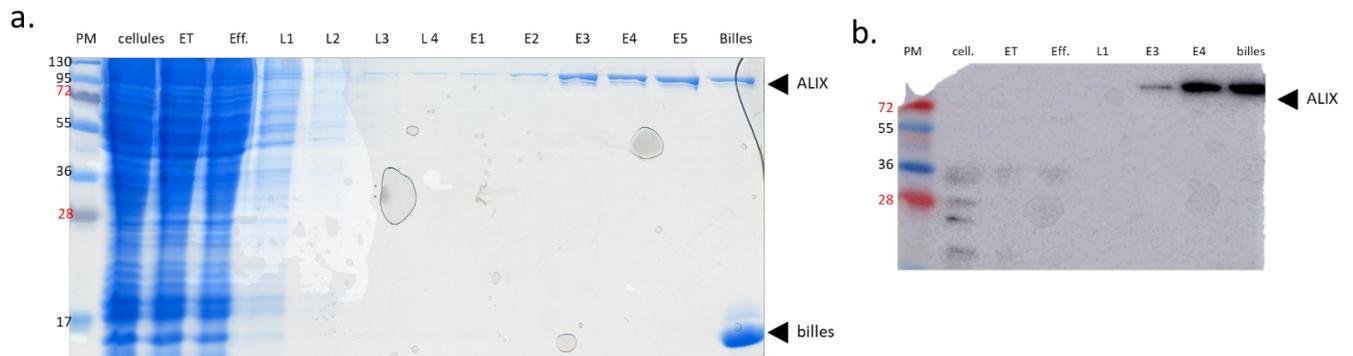


**Figure 75. Analyse par SDS-PAGE de la purification d'UEV et détermination de sa séquence par WB dans les tampons A1 et A3**

L'encadré jaune (1) représente la purification de R-UEV dans le tampon A1 et le carré bleu (2) la purification de R-UEV dans le tampon A3. a. SDS-Page de 15% : Le marqueur de poids moléculaire (PM), un échantillon des protéines avant sonication (cell), après sonication (ET), l'effluent (Eff) sont déposés dans les quatre premiers puits du gel. Le premier lavage, les deux premières éluions et les billes se trouvent dans les derniers puits. b. Identification de la protéine par WB en utilisant un anticorps dirigé contre l'étiquette *Strep-tag*<sup>®</sup>. Un échantillon des protéines avant sonication (cell), après sonication (ET) et l'effluent (Eff) sont déposés dans les trois premiers puits du gel. Ils sont suivis d'une fraction de la première élution et des billes.

## R-ALIX

La protéine ALIX n'est ni soluble dans le tampon A1, ni dans le tampon A3. Elle est alors purifiée dans le tampon ALIX (Fisher et al., 2007) composé de 50 mM de Tris HCl pH 8 et 300 mM NaCl (**figure 76**). Dans ce tampon, les fractions éluées et contenant les billes, issues de la purification de la protéine ALIX présentent une bande d'environ 95 kDa sur le SDS-PAGE 10% dans les fractions d'élué et sur les billes. Le WB confirme la présence d'une protéine avec une molécule de *Strep-tag*<sup>®</sup> dans les fractions d'élué et sur les billes.



**Figure 76. Analyse par SDS-PAGE de la purification d'ALIX et détermination de sa séquence par WB dans le tampon ALIX**

SDS-Page 15% suivant la purification par affinité à l'aide de billes MagStrep Type 3 XT. Le marqueur de poids moléculaire (PM), un échantillon des protéines avant sonication (cell), après sonication (ET), l'effluent (Eff) sont déposés dans les quatre premiers puits du gel. Le premier lavage, les quatre lavages et les 5 élué sont déposés avant les billes qui se trouvent dans les derniers puits. Identification de la protéine R-ALIX par WB en utilisant un anticorps dirigé contre l'étiquette *Strep-tag*<sup>®</sup>. Un échantillon des protéines avant sonication (cell), après sonication (ET) et l'effluent (Eff) sont déposés dans les trois premiers puits du gel. Ils sont suivis d'une fraction du premier lavage et des deux dernières étapes d'élution et enfin, des billes.

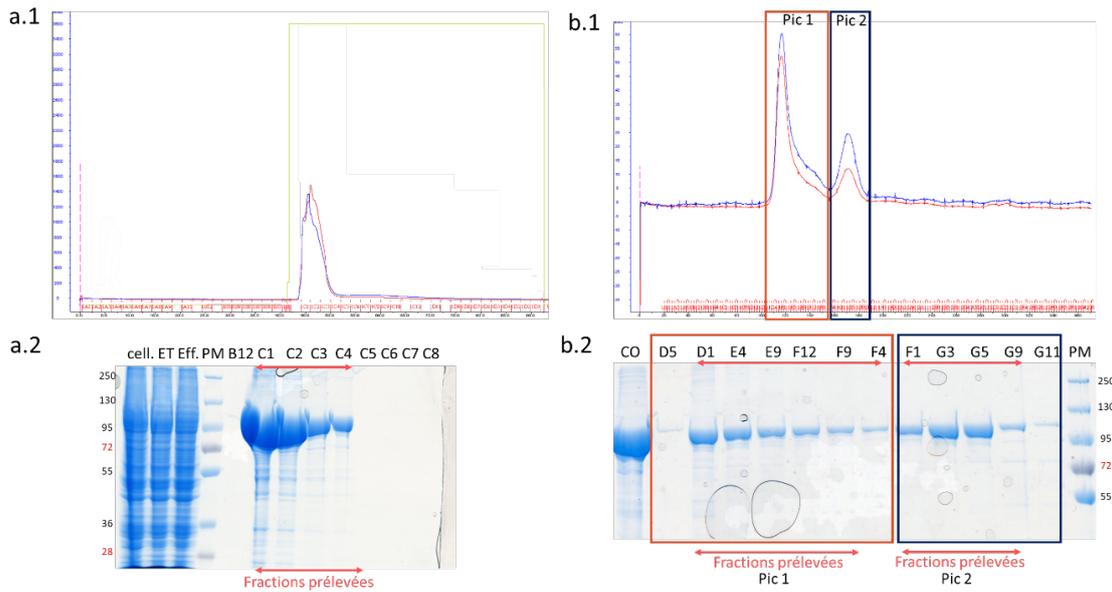
Pour conclure, des bandes de PM attendus par rapport au PM théorique pour chacun des partenaires sont identifiées dans les SDS-PAGE et identifiées par WB. A l'issue de cette première étape de contrôle d'expression des protéines, seules les protéines Vpr pour laquelle le laboratoire dispose également d'une aliquote produite chimiquement et ALIX dont la solution tamponnée est mise au point, sont utilisées dans la suite de ce travail.

## I.2. Purification et caractérisations biophysiques des partenaires ALIX et Vpr produits en cellules de mammifères

Pour aller plus loin dans la caractérisation des protéines R-ALIX et R-Vpr, les protéines produites en cellules BHK21 sont purifiées en utilisant une colonne d'affinité comme décrit dans la section matériels et méthodes, §VI.2.

### I.2.1 Purification d'ALIX

Les culots issus d'une production cellulaire de 2 L exprimant ALIX sont repris dans la solution tamponnée ALIX et purifier comme décrit précédemment (Section Matériels et Méthodes, §VI) Pour les protéines présentant une étiquette R, l'élution se fait en une étape (courbe verte sur le chromatogramme, **figure 77 a.1**). En suivant l'absorbance à 280 nm, un seul pic s'étendant des fractions C1 à C4 est visible sur le chromatogramme. La pureté des fractions sont contrôlées à l'aide d'un SDS-PAGE 10% (**figure 77 a.2**). Des aliquotes contenant les cellules, le lysat et l'effluent sont déposées dans les puits du gel dénaturant pour contrôler la présence de la protéine dans les échantillons de départ. Les fractions ayant une bande de taille attendue (99 kDa) et le moins de bandes contaminantes sont regroupées pour être concentrées. Une deuxième étape de purification est effectuée en utilisant la colonne S200 16/60 équilibrée dans le tampon ALIX (**figure 77 b.1**). Au regard de l'absorbance à 280 nm, deux pics sont visibles sur le chromatogramme. Le premier s'étend des fractions D5 à F9 (appelé ALIX pic 1) et le second des fractions F4 à G11 (appelé ALIX pic 2). La présence de la protéine ALIX est contrôlée dans chacun de ces pics à l'aide d'un SDS-PAGE 10% (**figure 77 b.2**). Le premier puit est utilisé pour contrôler l'enrichissement de l'échantillon par rapport à la purification par affinité, après l'étape de concentration (échantillon « CO »). A l'issue de la migration sur gel dénaturant, une bande correspondant à une protéine de PM ~99 kDa est présente dans les différentes fractions correspondant aux deux pics. D'une part, les fractions D1 à F9 (appelé ALIX pic 1) et d'autre part, de F4 à G9 (appelé ALIX pic 2) sont regroupés et concentrés séparément. A l'issue de l'étape de concentration deux fractions sont récupérées, la première correspondant au pic 1 est composée de 800 µL à une concentration de 2,58 mg/mL. La seconde fraction, correspond au pic 2 est composée de 350 µL à une concentration de 2,66 mg/mL. Ces échantillons sont ensuite utilisés séparément pour la caractérisation biophysique de la protéine ALIX.



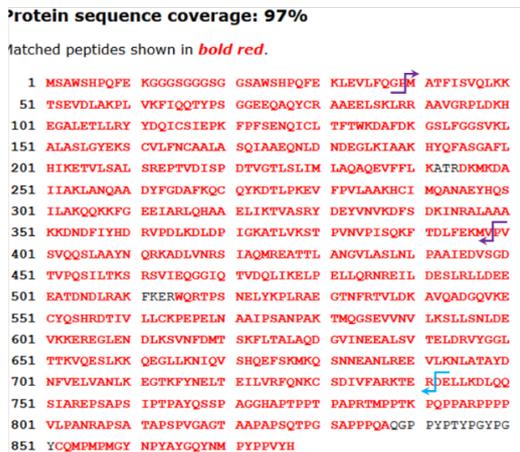
**Figure 77. Purification de R-ALIX dans le tampon ALIX**

a.1. Chromatogramme issue de la purification par affinité de R-ALIX à l'aide d'une colonne StrepTrap®. a.2. SDS-Page 15%. Un échantillon de cellules avant sonication (cell) et après sonication (ET), d'effluent (Eff) et le marqueur de poids moléculaire (PM) sont déposés dans les quatre premiers puits du gel. Les fractions prélevées et utilisées pour être concentrées et purifiées par chromatographie à exclusion de taille sont soulignées par une flèche rouge. b.1. Chromatogramme issue de la purification chromatographie à exclusion de taille à l'aide d'une colonne S200 16/60. Les encadrés orange (pic 1) et bleu (pic 2) représentent les différents pics observés sur le chromatogramme. b.2. Gel SDS-Page 15%. Un échantillon issu des fractions concentrées (CO) est déposé dans le premier puit. Le PM est déposé dans le dernier puit. Les fractions prélevées et utilisées pour être concentrées et conservées sont soulignées par une flèche rouge.

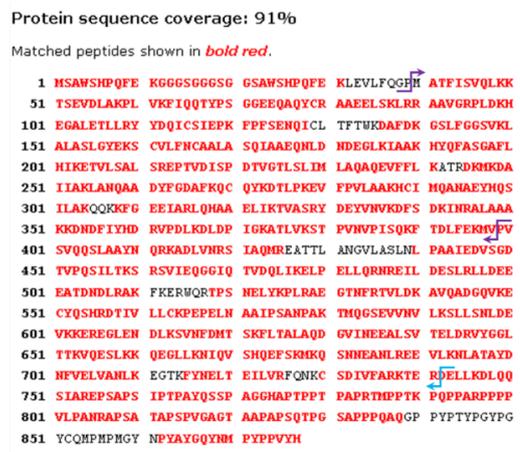
### 1.2.2 Contrôle de la séquence et des PTM d'ALIX

Une fois la protéine purifiée des analyses de MS/MS sont réalisées pour contrôler la séquence des protéines et identifier l'ensemble des PTM. Pour chacune de ces analyses, les protéines ALIX sont analysées en gel et en solution. La séquence est confirmée par un recouvrement compris entre 97% pour les préparations en solution et 91% pour la préparation en gel, avec un nombre de spectres supérieur à 3700 dans les deux cas (figure 78).

En solution



En gel



**Figure 78. Couverture de la séquence de la protéine R-ALIX pour les préparations en solution et en gel**

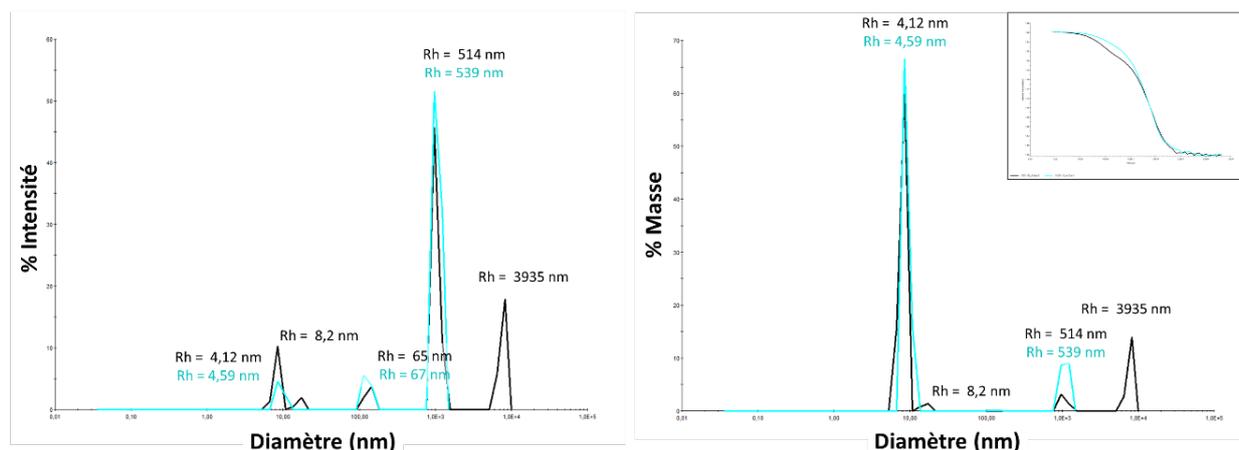
Lors de la caractérisation des PTM d'ALIX, une attention particulière est portée sur les phosphorylations et les ubiquitinations, qui sont importantes pour la fonctionnalité de la protéine comme indiquée dans la littérature (Dowlatshahi et al., 2012; Sette et al., 2013; Votteler et al., 2009). Toutes les modifications obtenues en fonction du type d'échantillon (en gel ou en solution) après digestion par la trypsine sont répertoriées dans le **tableau 18**.

Enzymes		Trypsine
En gel	Recouvrement de séquences	97%
	Phosphorylations	S147, S405, S462, T518, S596
	Ubiquitinations	K564
En solution	Recouvrement de Séquence	91%
	Phosphorylations	S147, T518
	Ubiquitinations	K58, K99, C270 K287, C289, K308, K378, S401, K525, K564, K729

**Tableau 18. Les PTM observées par MS/MS pour ALIX produite en BHK21**

### 1.2.3 Analyse par DLS des protéines ALIX

La solubilité et la capacité à multimériser des protéines ALIX dans la solution tamponnée de purification sont caractérisées par DLS. Ceci permet de déterminer la qualité des protéines avant de réaliser les complexes. Ainsi, les protéines recombinantes ALIX purifiées dans la solution tamponnée ALIX issues des fractions pic 1 et pic 2 (figure 77), sont analysées et comparées par DLS en suivant le protocole décrit dans la section matériels et méthodes, §IX.1. Les profils DLS pour l'un ou l'autre échantillon sont assez similaires et les pics observés sont majoritairement monodisperses (<12% de polydispersité ; **figure 79**). Cependant, le profil DLS correspondant à l'échantillon pic 1 présentent cinq pics alors que celui de l'échantillon pic 2 n'en présente que trois. En effet, deux pics correspondant à des molécules de Rh de 8,4 nm et de 3935 nm apparaissent sur le profil DLS de l'échantillon pic 1 (courbe noire) mais sont absents du profil DLS de l'échantillon pic 2. Ce profil est identique à celui observé en distribution de masse. Néanmoins, la macromolécule majoritaire en masse pour l'un et l'autre profil DLS possède un Rh de 4,12 nm et 4,59 nm respectivement pour l'échantillon issue des échantillons pic 1 et pic 2. En appliquant la loi de Stokes-Einstein les PM correspondant sont de 93 kDa et 119 kDa, soit un monomère d'ALIX (99 kDa).



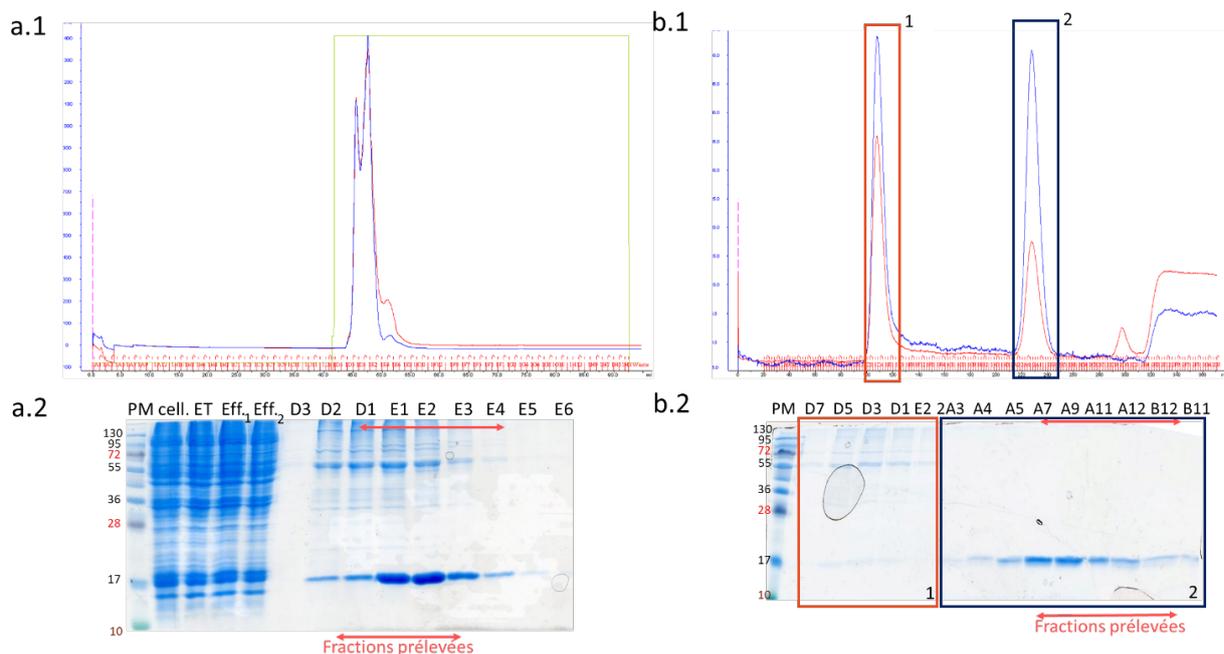
**Figure 79. Comparaison des profils DLS des fractions nommées pic1 et pic2 obtenues à l'issue de la purification par chromatographie de taille d'ALIX**

Représentation des distributions en intensité (à gauche) et en masse (à droite) en fonction des diamètres des protéines en solution. La courbe noire correspond à l'échantillon sorti en premier de la colonne de chromatographie, pic 1. La courbe turquoise correspond à l'échantillon sorti en dernier de la chromatographie, pic 2. Le Rh est indiqué au-dessus du pic correspondant. Chaque graphique correspond à la moyenne de dix mesures. Les corrélogrammes sont représentés dans l'encadré en haut à droite.

#### 1.2.4 Purification R-Vpr

Tout d'abord, les culots issus de 4 L de culture cellulaire exprimant Vpr sont repris dans le tampon A3 (tableau 6, section matériels et méthodes, paragraphe, IV.2.3). Pour maximiser l'attachement des protéines à la matrice l'effluent est chargé une nouvelle fois sur la colonne Streptactine. A l'issue de l'élution, deux pics s'étendant des fractions D3 à D1 et fractions D1 à E3, avec un coude fraction E4 (figure 80 a.1) sont présent sur le chromatogramme.

Un SDS Page 15% (figure 80 a.2) est réalisé pour contrôler le profil des aliquotes contenant les cellules (cell), le lysat (ET) et les deux effluents. Les fractions présentant le plus de matériels de PM attendu (15 kDa) et le moins de contaminants sont regroupées pour être concentrées. Ainsi les fractions de D1 à E4 sont regroupées. Une deuxième étape de purification est effectuée en utilisant la colonne S200 26/60 équilibrée dans le tampon A3 (figure 80 b.1). Au regard de l'absorbance à 280 nm, deux pics s'étendant des fractions D7 à E2 et des fractions 2A3 à 2B11 sont présents sur le chromatogramme. La pureté de la protéine VPR est contrôlée à en réalisant un SDS Page 15% (figure 80 b.2). A l'issue de la purification de Vpr, un volume de 1 mL de protéines ayant une concentration de 1,6 mg/mL est conservé après une congélation rapide dans de l'azote liquide à  $-80^{\circ}\text{C}$ .



**Figure 80. Purification de R-Vpr dans le tampon A2**

**a.1** Chromatogramme issue de la purification par affinité à l'aide d'une colonne StrepTrap®. **a.2.** SDS-PAGE 10% : Le marqueur de poids moléculaire (PM), ainsi qu'un échantillon de cellules avant sonication (cell) et après sonication (ET) et les deux effluent (Eff<sub>1</sub> et Eff<sub>2</sub>) sont déposés dans les premiers puits du gel. Les fractions prélevées et utilisées pour être concentrées et purifiées par chromatographie à exclusion de taille sont soulignées par une flèche rouge. **b.1** Chromatogramme issue de la purification chromatographie à exclusion de taille à l'aide d'une colonne S200 26/60. Les encadrés orange (1) et bleu (2) représentent les différents pics observés sur la chromatogramme. **b.2.** SDS-PAGE 10% : Le PM est déposé dans le premier puit. Les fractions prélevées et utilisées pour être concentrées et conservées sont soulignées par une flèche rouge.

### 1.2.5 Contrôle de la séquence et des PTM de Vpr

Une fois la protéine purifiée des analyses de MS/MS sont réalisées pour contrôler la séquence des protéines et identifier les PTM. Pour les analyses de MS/MS, les protéines VPR sont uniquement analysées en gel. La séquence est confirmée par un recouvrement compris de 52% pour un nombre de spectres de 631 (**figure 81**).

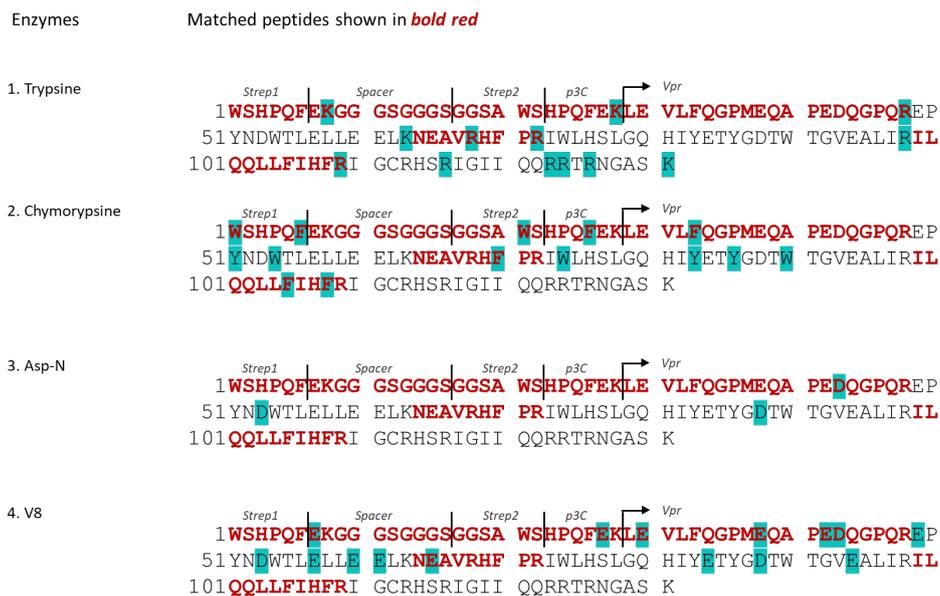
Proteins sequence coverage: 52%

Matched peptides shown in **bold red**

	Strep1	Spacer	Strep2	p3C	Vpr
1	<b>WSHPQFEKGG</b>	<b>GSGGGSGGSA</b>	<b>WSHPQFEKLE</b>	<b>VLFQGPMEQA</b>	<b>PEDQGPQREP</b>
51	YNDWTLELLE	ELK <b>NEAVRHF</b>	PR <b>IWLHSLGQ</b>	HIYETYGDTW	TGVEALIR <b>IL</b>
101	<b>QQLLFIFRI</b>	GCRHSRIGII	QRRTRNGAS	K	

**Figure 81. Couverture de la séquence de la protéine R-Vpr pour les préparations en gel**

Une absence des spectres est observée au niveau de 48% de la séquence de Vpr. La séquence de Vpr étant riche en arginine, la digestion par la trypsine induit des fragments dont la taille est au-delà des limites de détection du spectromètre. Il en est de même pour la chymotrypsine et aussi pour les enzymes Asp-N et V8 disponibles sur la plateforme (**figure 82**). Néanmoins, l'analyse de la taille de Vpr sur gel montrent que les protéines sont intactes.



**Figure 82. Couverture de la séquence de la protéine R-Vpr et mise en évidence des sites de coupures des enzymes de digestion disponibles à la plateforme**

La partie de la séquence de R-Vpr recouverte par des peptides est représentée en gras et en rouge. Les résidus reconnus par les enzymes de digestions sont surlignés en turquoise.

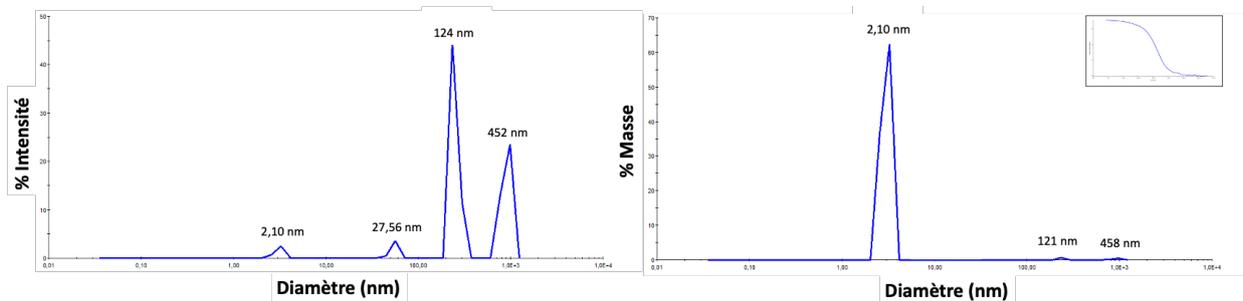
Aucune PTM n'est identifiées sur la séquence de Vpr. Néanmoins, l'étiquette R en possède. D'une part, 13 spectres sur un total de 133 présentent une double ubiquitination au niveau de l'étiquette selon 4 combinaisons possibles, S2 et K8, ou K8 et S12, ou K8 et S16, ou encore S13 et S16. D'autre part, 27 spectres sur un total de 123 présentent une phosphorylation selon 3 combinaisons soit, la position S12 est phosphorylée, soit, la position S16, soit, les deux positions S12 et S16.

### I.2.6 Analyse par DLS des protéines Vpr

La solubilité et la capacité à multimériser des protéines Vpr produites en cellules BHK21 et par synthèse chimique, sont caractérisées par DLS

Premièrement, les protéines recombinantes Vpr produites en cellules de mammifères et purifiées dans la solution tamponnée A2, sont analysées et comparées par DLS en suivant le protocole décrit dans la section Matériels et Méthodes, §IX.1 (**figure 83**). Le profil DLS est composé de pics majoritairement monodisperses (<11% de polydispersité) et est composé de 4 pics correspondant à des molécules ayant un Rh de 1,87 nm, 28 nm, 124 nm et 452 nm. Le profil DLS en masse est composé d'un seul correspondant à une molécule de Rh de 1,87 nm. Ce pic

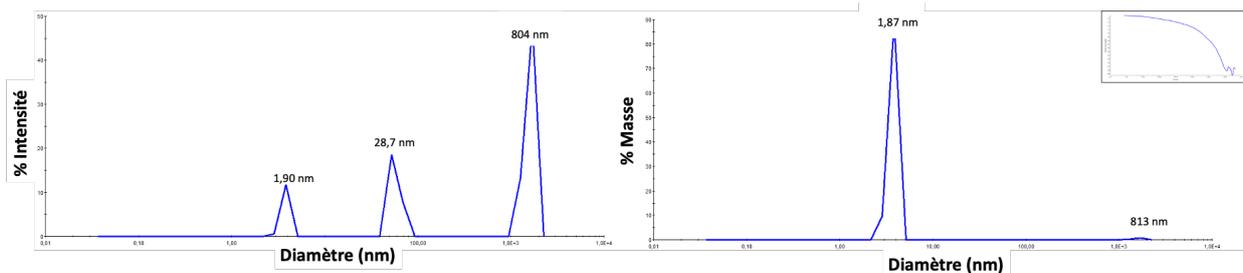
correspondrait à un monomère de R-Vpr, puisque l'équation de Stokes-Einstein permet de déterminer un PM de 9 kDa. Le PM théorique est de 11395 Da. De plus, le corrélogramme ne présente pas de signes d'agrégation.



**Figure 83. Profils DLS de la protéine Vpr produite en cellules de mammifères**

Représentation des distributions en intensité (à gauche) et en masse (à droite) en fonction des diamètres des protéines en solution. Le Rh est indiqué au-dessus du pic correspondant. Chaque graphique correspond à la moyenne de dix mesures. Le corrélogramme est représenté dans l'encadré en haut à droite.

Deuxièmement, les profils DLS des protéines Vpr produites par synthèse chimique et conservée dans la solution tamponnée Vpr composée de 40 mM de phosphates de potassium pH 6,5, 50 mM de NaCl et 1 mM de DTT sont composés de pics monodisperses (<14% de polydispersité). Le profil DLS en intensité est divisé en 3 pics correspondant à des molécules ayant un Rh de 2,10 nm, 29 nm, 124 nm et 804 nm. Le chromatogramme en masse ne représente qu'une population de Rh égal à 2,10 nm. Ce pic correspondrait à un monomère de R-Vpr dont l'équation de Stokes-Einstein permet de déterminer le PM est de 14,4 kDa. Malgré nos efforts pour la conservation de la protéine nous observons au niveau du corrélogramme une certaine tendance à l'agrégation (**figure 84**).



**Figure 84. Profils DLS de la protéine Vpr produite par synthèse chimique**

Représentation des distributions en intensité (à gauche) et en masse (à droite) en fonction des diamètres des protéines en solution. Le Rh est indiqué au-dessus du pic correspondant. Chaque graphique correspond à la moyenne de dix mesures. Le corrélogramme est représenté dans l'encadré en haut à droite.

L'ensemble des résultats obtenus lors de cette analyse par DLS est représenté dans le tableau ci-dessous (**tableau**

**19**).

	UEV	Vpr (BHK21)	Vpr (par synthèse chimique)	ALIX
<b>PM (Da)</b>	20367	15132	11395	99776
<b>Solution tamponnée</b>	A1 et A3	A2	- 40 mM de phosphates de potassium pH 6.5 - 50 mM de NaCl - 1 mM de DTT	- 50 mM de Tris HCl pH 8 - 300 mM NaCl
<b>DLS Rh (nm)</b>	n.d.	2,10 ± 0,06	1,87 ± 0,30	4,36 ± 0,33

**Tableau 19. Caractéristiques des partenaires de p6 en fonction de la solution tamponnée de purification**

Le PM est calculé à partir de la séquence des différentes protéines à l'exception de la protéine Vpr produite par synthèse chimique pour laquelle il est obtenu par spectroscopie de masse electrospray (Cornille et al., 1999). Le Rh associé à la DLS représente la moyenne de 10 mesures.

### I.3 Formation et caractérisation des complexes entre 6His-Gag et ses partenaires protéiques

Les complexes entre la protéine 6His-Gag et ses partenaires R-ALIX et Vpr sont étudiés en utilisant la méthodologie FIDABio. Tout d'abord, le Rh de 6His-Gag marquée seule dans les différentes solutions tampons est mesuré et récapitulé ci-dessous (**tableau 20**).

Solution tamponnée	Cellules de production	Moyenne des Rh (nm)	Nombre de répétition
Tampon A2	Mammifères	4,38 ± 0,17	15
Tampon A2	Bactéries	4,45 ± 0.42	5
Tampon ALIX	Mammifères	3.42	1
Tampon A3	Mammifères	4,24 ± 0,18	8

**Tableau 20. Valeurs moyennes des Rh mesurées par FIDABio de la protéine 6His-Gag marquée dans les différentes solutions tamponnées**

#### I.3.1 Analyse FIDABio des complexes formés entre 6His-Gag et R-ALIX

Différentes stœchiométries pour former les complexes entre 6His-Gag et R-ALIX sont analysées à l'aide d'analyses FIDABio (dont les principes sont décrits dans la session Matériels et Méthodes ; §IX.3). Pour cela des concentrations variables d'ALIX sont mélangées à une concentration fixe de la protéine 6His-Gag marquée par une sonde ALEXA (ici 1 µM) dans le tampon A3. Le FIDalyser permet alors d'évaluer le Rh de la protéine recombinante marquée en solution. Cette valeur est corrigée en fonction de la viscosité expérimentale de la solution tamponnée. Ainsi, le Rh de 6His-Gag marquée en solution dans le tampon A3 est estimé à 3,42 nm. Cette valeur est issue d'une seule analyse. Dans la même solution tamponnée, le Rh de Gag est estimé à 3,80 nm ou 4,75 nm par DLS (**Tableau 9**). Lorsque la protéine Gag marquée est mise en contact avec la protéine ALIX préalablement diluée dans le tampon A3, le Rh observé augmente (**tableau 21**). Les Rh représentés ici correspondent à une mesure.

Ratio Gag : ALIX	Moyenne des Rh (nm)	Nombre de répétition
1 : 0	3,42	1
1 : 10	4,25	1
1 : 20	4,15	1
1 : 30	3,81	1

**Tableau 21. Valeurs de Rh mesurées par FIDAbio de la protéine 6His-Gag marquée en présence d'ALIX en fonction des ratios protéine-protéine**

### I.3.2 Analyse par FIDABIO du complexe entre Gag et Vpr

Pour commencer, différentes stœchiométries pour former les complexes entre 6His-Gag et R-Vpr sont testées à l'aide d'analyses FIDAbio. Pour cela des concentrations variables de Vpr et sont mélangées à une concentration fixe de Gag marquée par une sonde ALEXA (ici 1  $\mu$ M) dans le tampon A2. Tout d'abord, le Rh de la protéine recombinante 6His-Gag marquée en solution dans le tampon A2 est estimé à  $4,26 \pm 0,34$  nm. Cette valeur est la moyenne de 3 analyses réalisées avec trois protéines marquées de manière indépendantes. Des analyses DLS sont réalisées sur la protéine purifiée dans le tampon A2 (non présentées) et indique un Rh de  $4,54 \pm 0,14$  nm (moyennes obtenues à partir de 10 mesures DLS réalisées sur 5 échantillons protéiques différents).

Différentes méthodes de formations de complexes entre les protéines 6His-Gag et R-Vpr sont testées (**tableau 22**). D'abord, elles sont mises en contact à température ambiante et déposées dans le FIDalyser comme décrit dans la session Matériels et Méthodes ;§IX.3. Ensuite, les protéines sont mélangées et incubées 30 minutes à 37°C pour favoriser les échanges. Finalement, les protéines sont mélangées et dialysées sur la nuit à 4°C ensemble contre le tampon A2 pour lequel la concentration de NaCl est diminuée à 400 mM (A2b).

Aucune modification du Rh de la protéine 6His-Gag n'est observée lorsque les protéines sont mélangées à température ambiante et à 37°C. Tandis que lorsqu'elles sont mélangées et dialysées contre le tampon A2b, le Rh correspondant à la protéine Gag marquée augmente (**tableau 22**).

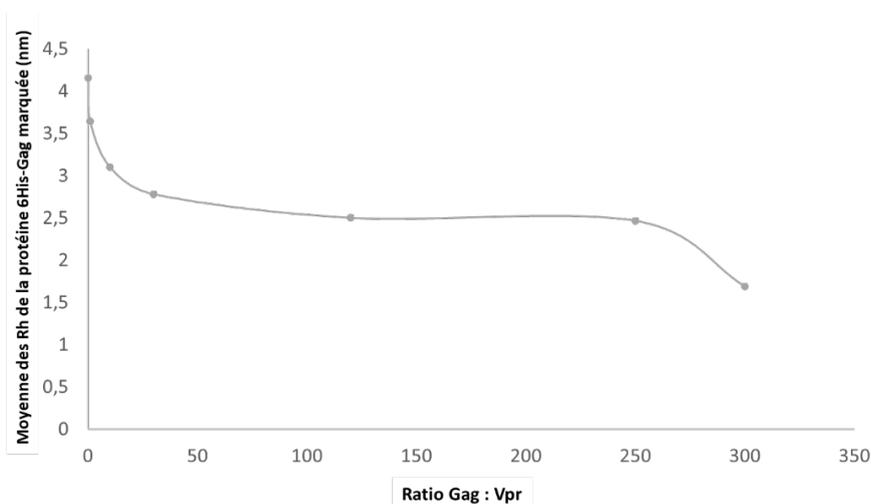
Ratio Gag : Vpr	Moyenne des Rh (nm)	Nombre de répétition
<b>Mélange à température ambiante</b>		
1 : 0	4,66 ± 0,25	5
1 : 10	4,01 ± 0,01	3
1 : 20	5,27 ± 0,98	3
1 : 30	4,47 ± 0,33	3
1 : 50	4,77 ± 0,84	3
1 : 100	4,37 ± 0,24	3
<b>Mélange à 37°C pendant 30 minutes</b>		
1 : 0	4,07 ± 0,10	4
1 : 1	4,05 ± 0,25	2
1 : 10	4,12 ± 0,11	2
1 : 20	3,99 ± 0,21	2
1 : 30	4,02 ± 0,07	2
1 : 50	4,03 ±	2
<b>Formation par dialyse</b>		
1 : 0	4,07 ± 0,37	2
1 : 10	5,11 ± 0,28	2
1 : 20	6,48 ± 0,34	3

**Tableau 22. Valeurs de Rh déterminées par FIDAbio de la protéine 6His-Gag marquée en présence de R-Vpr en fonction des ratios protéine-protéine et des conditions de formation des complexes**

Ensuite, différentes stœchiométries pour former les complexes entre 6His-Gag produite en cellules de mammifères et la protéine Vpr produite par synthèse chimique sont analysées. Les complexes sont de nouveau formés en faisant varier la concentration de Vpr et en conservant une concentration fixe de la protéine Gag marquée (ici 500 nM) dans le tampon A2. Tout d'abord, le Rh de la protéine 6His-Gag marquée en solution dans le tampon A2 est estimée à (4,36 ± 0,04) en très bon accord avec ce qui est observé dans les analyses précédemment. Cette valeur est la moyenne de 4 analyses réalisées avec quatre protéines marquées de manière indépendantes (**tableau 23**). Différemment de ce qu'on a observé précédemment avec la protéine R-Vpr produite en BHK21, lorsque les concentrations de Vpr ajoutées à la protéine Gag augmentent, le Rh de la protéine 6His-Gag diminue (**figure 85**).

Ratio Gag : Vpr	Moyenne des Rh (nm)	Nombre de répétition
<b>Mélange à température ambiante</b>		
1 : 0	4,36 ± 0,04	4
1 : 1	3,31 ± 0,25	4
1 : 10	2,38 ± 0,36	3
1 : 30	2,38 ± 0,94	4
1 : 60	2,6 ± 0,83	2
1 : 100	3,11 ± 0,26	2
1 : 120	2,5	1
1 : 250	2,47 ± 1,02	2
1 : 300	1,69	1

**Tableau 23. Valeurs de Rh déterminées par FIDAbio de la protéine 6His-Gag marquée en présence de Vpr produite par synthèse chimique en fonction des ratios protéine-protéine**

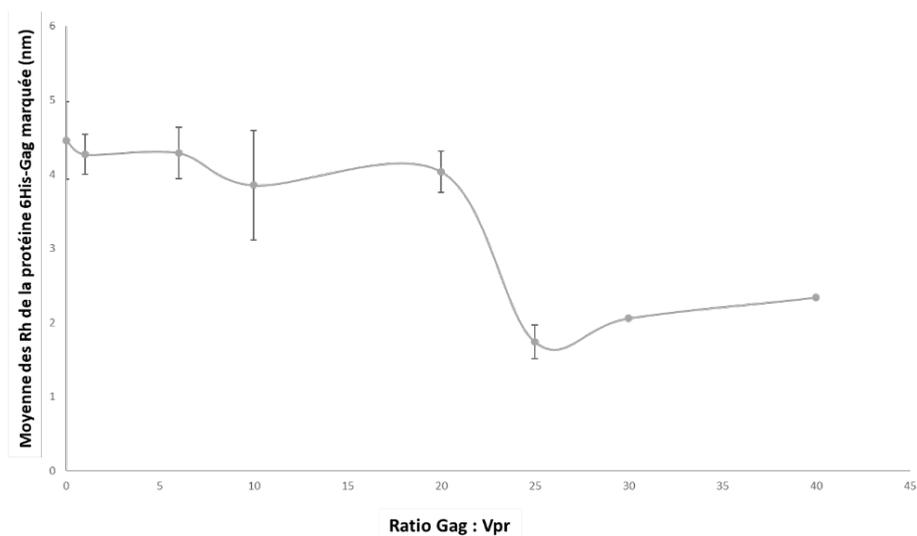


**Figure 85. Analyse par FIDAbio de l'effet de l'ajout de quantité croissante de protéine Vpr synthétisée chimiquement sur le Rh moyens de la protéine 6His-Gag marquée par une sonde fluorescente**

En parallèle, la capacité de la protéine R-Vpr produites en cellules de mammifères à interagir avec la protéine Gag produite en cellules bactériennes est contrôlée de la même manière. Les complexes sont aussi formés par mélange à température ambiante, en faisant varier la concentration de R-Vpr et en conservant une concentration fixe de la protéine Gag marquée A (ici 300 mM) dans le tampon A2. Tout d'abord, le Rh de la protéine recombinante Gag marquée est estimé à (4,45 ± 0,52) en très bon accord avec ce qui est observé dans les analyses de DLS. Cette valeur est la moyenne de 5 analyses réalisées avec quatre protéines marquées de manière indépendantes (**tableau 24**). Lorsque les concentrations de R-Vpr augmentent, le Rh de la protéine Gag-6His diminue (**figure 86**). Ces résultats sont en accord avec ceux obtenus lorsque la protéine 6His-Gag produites en cellules de mammifères est mélangé avec la protéine Vpr produite synthétiquement.

Ratio Gag : Vpr	Moyenne des Rh (nm)	Nombre de répétition
<b>Mélange à température ambiante</b>		
1 : 0	4,45 ± 0,52	5
1 : 1	4,27 ± 0,26	5
1 : 6	4,29 ± 0,34	2
1 : 10	3,85 ± 0,74	3
1 : 20	4,04 ± 0,27	2
1 : 25	1,74 ± 0,25	2
1 : 30	2,06	1
1 : 40	2,34	1

**Tableau 24. Valeurs de Rh déterminées par FIDAbio de la protéine Gag-6His produite en E. coli marquée en présence de R-Vpr en fonction des ratios protéine-protéine**



**Figure 86. Analyse par FIDAbio de l'effet de l'ajout de quantité croissante de protéine R-Vpr sur le Rh moyens de la protéine Gag-6His, produite en cellules bactériennes, marquée par une sonde fluorescente**

Pour conclure, l'objectif de ces analyses est de déterminer les conditions de formations de complexes *in vitro* entre la protéine Gag et ses partenaires. L'obtention de ces résultats sont la première étape pour réaliser des complexes entre la protéine Gag, un de ses partenaire protéique et l'ARNg afin de pouvoir analyser l'effet des ligands de p6 sur la spécificité d'interaction entre le précurseur Gag et l'ARN viral.

**Résumé :** les MVA recombinants permettant la production de R-Vpr, R-ALIX et R-UEV en BHK21, sont mis au point. Les Rh, les séquences et les PTM sont caractérisées pour les protéines R-ALIX et R-Vpr. Cependant, le recouvrement de la séquence de Vpr par MS/MS est partiel (50%). Finalement, les premiers essais de formation complexes entre la protéine 6His-Gag et l'un ou l'autre de ces partenaires sont analysés par FIDABio. Aucune modification du Rh n'est relevée pour les essais des mélanges entre 6His-Gag et R-ALIX à température ambiante. Le même résultat est obtenu pour le mélange entre 6His-Gag et R-Vpr à température ambiante et à 37°C. Cependant le Rh augmente lorsque 6His-Gag et R-Vpr sont dialysés ensemble et diminue lorsque 6His-Gag est mélangée avec la protéine Vpr synthétisée chimiquement et lorsque R-Vpr est mélangée à la protéine Gag-6His produite en cellules bactériennes.

## II. Discussion Objectif n°2 : Caractérisation des interactions *in vitro* entre Gag et ses partenaires protéiques

Le second objectif de ma thèse est d'optimiser les conditions de formation des complexes formés par Gag et les partenaires de p6 dans le but d'apporter une meilleure compréhension de leur rôle dans la spécificité d'interaction entre Gag et l'ARNg. Pour cela, les partenaires sont produits en cellules de mammifères et caractérisés *in vitro*. Ensuite, des conditions de formation de complexes entre la protéine Gag et ces partenaires sont investiguées.

### I.1. Production des vecteurs viraux permettant l'expression des partenaires de Gag en cellules de mammifères

Tout d'abord, j'ai mis au point les vecteurs viraux MVA pour permettre l'expression des différents partenaires de p6 en cellules de mammifères. La première étape de mon projet est de contrôler la qualité des virus produits. Dans un premier temps, la perte de l'expression de la RFP est contrôlée pour chacun des vecteurs viraux, caractéristique de la recombinaison. Trois étapes de sélections sont nécessaires (décrit section Matériels et Méthodes, §V.2). A l'issue de la sélection des virus recombinants, l'expression des protéines d'intérêts par les virus est d'abord contrôlée à l'aide d'électrophorèse sur SDS-Page et d'un WB (**figures 74, 75 et 76**). Alors, une production de protéine à partir de cellules BHK21 adhérentes est réalisée pour chacun des trois partenaires. Si ce type de production de protéines permet de réaliser des contrôles d'expression, la quantité de protéines obtenues ne suffit pas pour réaliser des étapes de purification par colonne d'affinité et de chromatographie à exclusion de taille. A la place, les protéines sont purifiées par affinité à l'aide des billes *Magstrep*<sup>®</sup>. Différents problèmes sont observés lors de ces étapes, d'une part, les SDS-Page présentent de nombreux contaminants. D'autre part, l'étape d'élution est peu efficace. En effet pour chacun des 3 partenaires, peu de protéines sont récupérées lors des étapes d'élution et beaucoup de protéines restent accrochées sur les billes. Néanmoins, les WB permette d'identifier les trois partenaires dans les puits correspondants. En première attention les solutions tamponnée A1 et A3 sont testées pour purifier les protéines de manière similaire à ce qui est fait pour la protéine Gag recombinante. Il était attendu que le tampon A1 présente un meilleur rendement que le tampon A3 grâce à la présence de détergents et d'une plus forte concentration de NaCl. Or, ces conditions ne permettent pas de solubiliser la protéine ALIX, qui d'après la littérature est soluble dans des pH basiques (Fisher et al., 2007). Ainsi, après avoir purifié la protéine avec une solution tamponnée composée de 50 mM de Tris HCl pH 8 et 300 mM NaCl, l'étape d'élution permet de récupérer de plus grande quantité de protéines par rapport aux autres partenaires (2 fois plus). A cette étape, l'élution est réalisée avec de la biotine, or il est documenté que cette molécule est plus stable à pH basique (Lin et al., 2020). De la même façon, des études montrent que Vpr est plus soluble à pH acide (Cornille et al., 1999). Ainsi le tampon A2 (pH 6,8) est mis au point pour la purification de Vpr. Le tampon de purification optimal d'UEV doit encore être optimisé. Les conditions de solubilisation d'ALIX étant mise au point et ayant à disposition la protéine VPR

synthétisée chimiquement, ces deux protéines ont été prioritairement utilisées dans la suite de l'étude. Ainsi, la séquence en acides aminés, les PTM et l'état oligomérique en solution de ces deux partenaires sont caractérisés après une purification de protéines plus importante.

## I.2 Purification et caractérisations biophysiques des partenaires ALIX et Vpr produits en cellules de mammifères

### II.2.1 Caractérisation de la protéine ALIX

A partir des culots de cellules provenant d'une culture de 2 L, la purification d'ALIX permet de purifier 3 mg de protéines réparties en deux aliquotes. Des pertes peuvent avoir lieu à différentes étapes de la purification notamment au moment de la liaison de la protéine sur la colonne d'affinité. En effet il n'y a pas de différence dans les profils de migration sur gel dénaturant entre les échantillons extrait total et l'effluent (**figure 77**). Cela pourrait être amélioré en faisant passer deux fois l'effluent sur la colonne (comme réalisé avec la protéine VPR). De plus, l'élution est réalisée en une étape induisant le décrochement rapide de la protéine. Elle est alors très concentrée dans les fractions et peut précipiter. Finalement, les deux pics (**figure 77**) présent sur le chromatogramme peuvent correspondre à des états multimériques différents de la protéine en solution. Pour contrôler cela des analyse par MS/MS et DLS sont réalisées.

Les résultats de l'analyse de MS/MS sur les échantillons en gel et en solution permettent d'obtenir un recouvrement de séquences compris entre 91 et 97%, respectivement (**figure 78**). Ces résultats permettent de conclure que la protéine en solution correspond bien à ALIX. Même si des régions de 3 à 14 résidus à l'intérieur de la séquence d'ALIX ne sont pas recouverts. Certains spectres pourraient être absents à cause d'un problème de digestion dépendant de la fonctionnalité de la trypsine. Ou car les fragments produits seraient trop petits ou trop longs due au positionnement des résidus lysines et arginines. En conclusion, il est possible que la protéine soit à l'équilibre entre des multimères et des monomères en solution et ceci expliquerait la présence de plusieurs pics sur le chromatogramme (**figure 77**). Cette tendance est confirmée à l'issue des analyses DLS. La molécule majoritaire présente dans la distribution en masse des pics a et b, a un Rh de 4,12 nm. La loi de Stokes-Einstein permet de déterminer que le PM associé est de 93 kDa, soit la masse attendue pour un monomère d'ALIX (95 kDa ; **figure 79**). Cependant, une molécule présentant un Rh de 8,2 nm, est également présent en solution dans l'échantillon correspondant au pic a. Ce pic correspondrait à un dimère d'ALIX. Ces résultats sont en très bon accord avec des analyses de MALS et de diffusion de rayons X aux petits angles (SAXS) qui montrent qu'ALIX est présente sous forme monomérique ou dimérique en solution dans des conditions de pH similaires. Le domaine V de ALIX est le domaine qui induit la dimérisation de la protéine (Pires et al., 2009).

## II.2.2 Caractérisation de la protéine VPR

A partir des culots de cellules provenant d'une culture de 4 L, la purification de VPR permet de purifier 1,6 mg de protéines. Des pertes de matériel peuvent être causées par un mauvais accrochement de la protéine sur la colonne d'affinité. En effet il n'y a pas de différence dans les profils des échantillons extraits total et les effluents sur le SDS-PAGE malgré le double passage de l'effluent (**figure 79**). Comme pour ALIX, l'élution induit le décrochement rapide de la protéine qui pourrait entraîner sa précipitation. Il est aussi possible que la solution tamponnée ne soit pas optimale pour la purification de R-VPR.

La protéine est décrite dans la littérature comme étant capable d'oligomériser *in vitro* (Bourbigot et al., 2005; Sabbah et al., 2006) et *in vivo* (Fritz et al., 2010). Plusieurs résidus sont identifiés comme favorisant la dimérisation de Vpr, les résidus des positions 34 à 51 (Sabbah et al., 2006) et ceux s'étendant des positions 60 à 91, qui correspondent à un domaine riche en leucines et en isoleucines dans la partie C-terminale de Vpr. De plus, des études par RMN montrent que le domaine C-terminal de Vpr adopte une structure dimérique antiparallèle (Bourbigot et al., 2005). De plus, des études par microscopie confocale mettent en évidence que l'oligomérisation de Vpr est essentielle à l'interaction avec le précurseur Gag permettant le transport de Vpr à la MP. Au contraire, la multimérisation de Gag n'est pas nécessaire au recrutement de Vpr (Fritz et al., 2010).

Ainsi, l'état oligomérique de Vpr en solution dans le tampon A3 est étudié par des analyses de DLS. Une seule espèce d'un Rh de 1,5 nm est présente sur le profil en masse, le PM correspondant est de 9 kDa (**figure 80**). Le PM théorique de la protéine monomérique est de 15 kDa. Cette différence entre PM estimé et PM théorique pourrait être due à la conformation de la protéine Vpr qui ne serait pas complètement globulaire. Le profil de distribution en intensité, quant à lui, témoigne de la présence de plusieurs oligomères de grandes tailles. Néanmoins, le corrélogramme ne représente pas d'agrégation. Cette tendance est aussi observée avec la protéine synthétisée chimiquement. De même, le profil en distribution des masses ne représente qu'une espèce ayant avec un Rh de 1,8 nm correspondant à un PM de 15 kDa et donc à la protéine monomérique. Il est alors possible de conclure que l'une et l'autre protéine est majoritairement monomérique en solution, bien qu'elles conservent la capacité de former des oligomères de plus grande taille. La structure par RMN montre que la protéine Vpr est composée de 3 hélices- $\alpha$  repliées autour d'un noyau hydrophobe, entourées par des domaines N-terminaux et C-terminaux flexibles (Morellet et al., 2003). La présence de l'étiquette pourrait modifier l'aspect globulaire de la protéine. De plus, il est possible que l'étape de décongélation induise la multimérisation de la protéine. Pour contrôler ce phénomène, des analyses de DLS pourraient être réalisées sur de la protéine fraîchement purifiée, ainsi que sur un échantillon de protéines congelés rapidement dans de l'azote liquide et décongelés. Comparer les résultats de

ces analyses permettrait de déterminer l'impact de la congélation rapide dans l'azote sur l'état multimérique de Vpr ainsi que l'effet du temps après décongélation sur la protéine.

Ensuite, la séquence de la protéine R-Vpr produite en cellules de mammifères est contrôlée par MS/MS, le recouvrement de la séquence n'est que partiel (52% ; **figure 81**). Contrairement aux autres protéines, l'analyse de Vpr est réalisée avec 4 protéases différentes. Tout d'abord, la trypsine et la chymotrypsine qui coupe après les résidus aromatiques sont utilisées (**figure 82**) les fragments générés par l'une ou l'autre enzymes sont trop petits ou trop grands pour être correctement séquencés. Ensuite, les enzymes Asp-N qui coupe après les résidus acides aspartiques et v8 qui coupe après les résidus d'acides aspartiques et les acides glutamiques, génèrent de nouveau des fragments trop grands ou trop petits. Néanmoins, la partie N-terminale de la protéine est confirmée. Malgré ces problèmes, le PM associé à la protéine après migration sur gel dénaturant et transfert sur membrane PVDF correspond bien à la protéine complète qui ne serait pas dégradé (15 kDa).

### II.2.3 Le rôle des PTM dans le recrutement des partenaires de Gag

Les PTM jouent à la fois un rôle dans la régulation des interactions entre les protéines, en régulant les différentes étapes de cycle cellulaires et dans notre cas, viral. Elles sont notamment nécessaires au recrutement des partenaires cellulaires et viraux de p6 (**annexe 1**).

TSG101 et Alix sont toutes les deux ubiquitinés (Bussienne et al., 2021; Sette et al., 2010; Watanabe et al., 2020) et capables de se lier à des ubiquitines, notamment présentes sur Gag (Bussienne et al., 2021; Dowlatshahi et al., 2012; Pornillos et al., 2002). En effet le domaine V d'ALIX et UEV de TSG101 possèdent une poche qui permet l'interaction avec les ubiquitines de Gag. Dans la suite de la discussion, je vais me concentrer sur ALIX, pour laquelle des analyses de mutagenèses montrent que cette poche favorise l'interaction entre ALIX et Gag. En parallèle, des études montrent que l'ubiquitination des facteurs ESCRT améliore la fonction de cette machinerie (Dowlatshahi et al., 2012). Ainsi, ALIX est mono-ubiquitylée *in vivo* (Votteler et al., 2009). Deux E3 ubiquitines ligases sont identifiées comme impliquées dans cette ubiquitination, la protéine POSH et la protéine Nedd4-1. Ainsi, 33 des 65 lysines d'ALIX sont identifiées comme étant accessibles et des cibles potentielles. Des études *in vitro* montrent que la protéine POSH cible même plusieurs résidus lysines de la séquence d'ALIX (Votteler et al., 2009). En outre, il est mis en évidence que le recrutement de Nedd4-1 par ALIX, nécessite non seulement l'ubiquitination d'ALIX et aussi celle du *late domaine* LYPXnL de p6 (Sette et al., 2010). De nombreuses études mettent en évidence le rôle de Nedd4-1 dans la libération des particules virales néoformées du VIH-1 et d'autres rétrovirus comme le RSV, par exemple (Martin-Serrano et al., 2004; Sette et al., 2010; Strack et al., 2003, 2000). La protéine ALIX que j'ai produite en BHK21 présente de nombreuses ubiquitinations. Dont notamment l'ubiquitination de la lysine en position 564

qui est retrouvée à la fois dans l'échantillon en solution et en gel (**tableau 18**). De plus, il est décrit que la partie N-terminale d'ALIX contient un site de phosphorylation par les tyrosines kinases de la famille des Src (Kato et al., 2003). Notamment, des études structurales suggèrent que la Src kinase se lie à la fois au domaine Bro1 et phosphoryle des résidus tyrosines dans la queue C-terminale d'ALIX (Schmidt et al., 2005). Au regard des résultats de MS/MS obtenus sur l'échantillon d'ALIX purifié, des résidus sont phosphorylés dans les parties N- et C-terminales. Ces résultats sont ainsi cohérents avec ce qui est décrit dans la littérature (Schmidt et al., 2005). De plus, la serine en position 147 (dans le domaine Bro) et la tyrosine en position 518 (dans le domaine V) sont phosphorylés à la fois dans l'échantillon en gel et en solution. De nouvelles analyses de MS/MS obtenus à partir de purification d'ALIX produites en BHK21, sont nécessaires pour contrôler la répétabilité des PTM d'une production à l'autre. Néanmoins, ces résultats sont encourageants quant à la production d'une protéine fonctionnelle.

Vpr est décrite dans la littérature comme jouant un rôle important dans le détournement de la machinerie post traductionnelle de l'hôte. En effet elle s'associe avec des ubiquitines ligases pour induire la dégradation des protéines par la voie du protéasome, comme par exemple la protéine CTIP2 associée à l'extinction des gènes du VIH-1 (Forouzanfar et al., 2019) ou l'*helicase-like transcription factor* (HTLF), une protéine qui induit la réparation de l'ADN après la réplication (Lubow and Collins, 2020). Pour cela, il est mis en évidence que Vpr lie simultanément la protéine DCAF1 du complexe ubiquitine ligase Cul4A-DDB1-DCAF1 et son substrat. De la même manière, Vpr est capable de médier l'arrêt du cycle cellulaire en hyper-phosphorylant la protéine Cdc2. Cette protéine est un régulateur de point de contrôle des dommages à l'ADN. Sa phosphorylation inhibe son activation (Lubow and Collins, 2020). Aucune ubiquitination de Vpr n'est actuellement décrite dans la littérature. Enfin la phosphorylation de la serine en position 79 de Vpr est importante pour le blocage du cycle cellulaire G2/M. La protéine kinase A (PKA) est associée à cette phosphorylation (Barnitz et al., 2010). Comme pour la protéine Gag, la phosphorylation présente un caractère redondant, ainsi les serines en position 94 et 96 peuvent aussi être phosphorylées et elles jouent aussi un rôle dans l'arrêt de cycle en phase G2/M (Zhou and Ratner, 2000). En parallèle, une autre protéine kinase, la protéine kinase C (PKC- $\alpha$ ) phosphoryle la Serine en position 487 du domaine p6 (Bussienne et al., 2021; Kudoh et al., 2014). Cette phosphorylation est nécessaire à l'interaction entre Gag et Vpr et permet l'incorporation de Vpr dans les particules virales. En effet des analyses d'inhibition de son activité enzymatique dans les macrophages primaires humains, réduit les niveaux de Vpr dans les virions. (Kudoh et al., 2014). Lors des analyses par MS/MS réalisées sur VPR, aucune PTM n'est identifiée sur la séquence de la protéine. Il n'est pas exclu que cela provienne du recouvrement partiel de la séquence de la protéine. En outre, l'étiquette R est largement

modifiée. En effet des phosphorylations et/ou des ubiquitinations sont clairement identifiées. Ces modifications pourraient alors expliquer la faible proportion de protéines capables de s'accrocher sur la colonne *StrepTrap*<sup>®</sup>. L'étiquette d'ALIX ne présente pas de modification et d'ailleurs nous avons observé une meilleure capacité de la protéine à être retenue sur la colonne par rapport à Vpr. La modification de l'étiquette dans un cas et pas dans l'autre, pourrait être expliquée par la disponibilité de l'étiquette après l'étape de traduction. En effet il est possible que la protéine ALIX puisse se replier dans une conformation que ne permet pas d'interaction entre son étiquette et la machinerie post traductionnelle des cellules BHK21. Cela expliquerait aussi l'absence de modification dans la partie N-terminale avant le résidu 147. Des analyse par AUC pourraient permettre de déterminer la forme de la protéine *in vitro*. Néanmoins, la modification de l'étiquette de Vpr est un problème majeur. Cette étiquette, a été choisie pour faciliter les analyses de formation de complexes entre Vpr et 6His-Gag. Précisément, pour les étapes de sélection éventuelles par affinité. Au regard de ces résultats, il est préférable de changer de type d'étiquette. Cependant, le but de ce travail étant de former des complexes entre Gag et Vpr, une autre stratégie de formation de complexe est envisagée et sera discutée dans le paragraphe suivant. Cette nouvelle stratégie permettrait de trouver une solution aux problèmes liés à la formation des complexes et à l'étiquette de Vpr.

### **I.3. Formation et caractérisation des complexes entre 6His-Gag et ses partenaires protéiques**

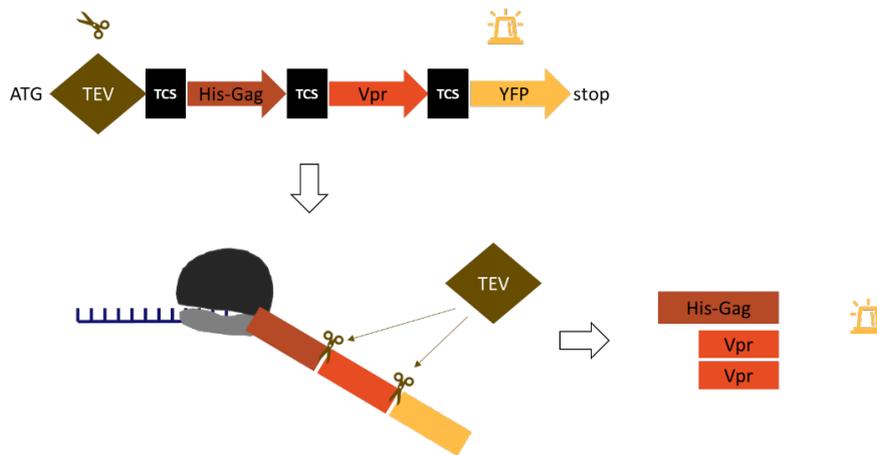
Les complexes entre la protéine produites en cellules de mammifères et ses partenaires sont étudiés en utilisant FIDAbio. Tout d'abord, le Rh de 6His-Gag marquée seule dans le tampon A correspondant est mesuré (récapitulé **tableau 19**). Les différences de Rh observées entre les mesures FIDAbio et en DLS peuvent être inhérentes aux méthodes. Il est intéressant de noter que le Rh est légèrement plus grand à partir des mesures en DLS par rapport aux mesures FIDAbio. De plus la variabilité de taille observée en DLS est retrouvée avec FIDAbio. Ainsi, il est possible de conclure que le marquage n'influence pas la conformation de Gag.

Les résultats préliminaires entre 6His-Gag et R-ALIX tendent à montrer que le Rh de la protéine Gag en présence d'ALIX augmente. Il passerait ainsi d'un Rh de 3,42 à 4,25 nm pour un ratio de 1 molécule de Gag pour 10 molécules d'ALIX (1:10). Puis resterait stable pour les ratios de 1 : 20 et 1 : 30. Des analyses supplémentaires sont nécessaires pour confirmer ces résultats notamment pour les plus petits ratios 1 : 1 ; 1 : 3 etc. De plus, la concentration de Gag initiale est 1  $\mu$ M or des analyses réalisées en ITC après dialyse d'ALIX dans une solution tamponnée composée de 10 mM HEPES (pH 7,4) et 150 mM NaCl ont permis d'identifier une *K<sub>D</sub>* entre Gag et ALIX de 8  $\mu$ M (Munshi et al., 2007). Ainsi, il serait intéressant d'utiliser une plus grande quantité de matériel initial. En outre, les complexes sont

réalisés par mélange à température ambiante dans le tampon A3, il est possible de réaliser les complexes dans d'autres conditions, par dialyse ou à 37°C comme cela a été fait pour les complexes entre Gag et Vpr.

D'autre part, le ratio d'interaction entre Gag et Vpr est discuté dans la littérature. Certaines analyses mettent en évidence *in vitro* que 5 à 7 molécules de Gag interagissent avec 1 molécule de Vpr (Müller et al., 2000; Singh et al., 2000). Alors que d'autres montrent que l'oligomérisation de Vpr est nécessaire au recrutement de Gag. Dans ce cas, les deux partenaires interagissent grâce à des interactions hydrophobes (Fritz et al., 2010). Ces derniers résultats sont intéressants car ils tendent à montrer que le complexe entre Gag et Vpr est résistant à de fortes concentrations en NaCl (900 mM) (Bachand et al., 1999). Or lors de cette étude les complexes sont réalisés à une concentration de 500 mM NaCl (tampon A2). Ainsi dans un premier temps des complexes entre la protéine 6His-Gag et R-Vpr sont formés *in vitro* en testant différentes conditions. Ni le mélange des partenaires à température ambiante, ni le mélange à 37°C permet d'observer un changement du Rh de 6His-Gag marquée. L'absence de formation de complexe peut être due à la modification de l'étiquette R de Vpr ou aux conditions expérimentales et notamment à la présence de CHAPS et de NaCl dans la solution tamponnée. C'est pour cela, qu'un test de formation de complexes par dialyse sur la nuit à 4°C est réalisé. A l'issue de la dialyse une augmentation du Rh de 4,07 à 6,48 nm pour une stœchiométrie de 1 molécule de Gag pour 20 molécules de Vpr (1 : 20), est observée (**tableau 22**). Il est impossible de conclure si le changement de Rh est la conséquence de la multimérisation de Gag ou celle de la formation d'un complexe entre les deux partenaires car le contrôle réalisé avec la protéine seule a complètement précipité. Ainsi, il faudrait recommencer cette analyse pour déterminer si la présence de Vpr stabilise la protéine Gag lors de la dialyse ou non. Il aurait pu être intéressant de marquer Vpr, or l'étape de marquage utilisée dans ce travail nécessite une solution tamponnée basique (pH 8) alors que Vpr montre une meilleure solubilité dans une solution tamponnée acide (pH 6,8). De plus, la réalisation des complexes entre Gag et la protéine Vpr produite par synthèse chimique indique que le Rh de 6His-Gag diminue lorsque la concentration de Vpr augmente. La fixation de Vpr sur la protéine induirait une compaction de la conformation de Gag et ainsi la diminution de Rh. Ces résultats très intéressants sont répétés plusieurs fois (de 2 à 4 fois en fonction des points, **tableaux 23, figure 85**). De plus, la même tendance est observée lorsque le complexe est formé entre la protéine R-Vpr produites en BHK21 et la protéine Gag produite en cellules de bactéries (**tableau 24, figure 86**). En effet pour les deux types de complexes, une réduction du Rh est observée pour un ratio de 1 molécule de Gag pour 20 molécules de Vpr. Il serait à l'avenir important de vérifier cette tendance en utilisant d'autres méthodes biophysiques comme l'AUC ou le SEC MALS. Enfin, une nouvelle stratégie est envisagée pour la formation de complexes entre les protéines produites en cellules de mammifères (Drillien et al., 2022). Elle consiste à produire un nouveau vecteur MVA qui coderait pour

la protéine Gag et son partenaire dans le même plasmide (**figure 85**). L'avantage de cette méthode est que les complexes peuvent se former directement dans la cellule après l'étape de traduction. Ainsi les partenaires seront récupérés à l'aide de l'étiquette 6His présente à l'extrémité N-terminale de Gag et la protéine Vpr ne serait alors plus étiquetée. Cette méthode peut aussi être envisagée avec les autres partenaires comme ALIX et par la suite UEV.



**Figure 87. Modèle de la stratégie d'expression de multi protéines en BHK21 par un MVA recombinant**

Représentation de la séquence génomique de la région d'intérêt. Elle code pour la protéase TEV et les protéines d'intérêts, ici 6HisGag, VPR et YFP séparée par un site de coupure TEV (TCS). La protéine YFP sert de contrôle de recombinaison homologue. Dans la cellule, la TEV est capable de s'exciser et de sépare les protéines pendant l'étape de traduction aboutissant à la surexpression des différentes protéines dans le cytoplasme (figure inspirée du travail de Drillien et al., 2022).

## **Conclusion & perspectives**



En conclusion, mon projet de thèse a pour but d'apporter une meilleure compréhension des interactions entre Gag et l'ARNg permettant l'encapsulation spécifique de l'ARNg dans les particules virales. Pour mieux comprendre les mécanismes subjacents à cette étape fondamentale, notre stratégie consiste à combiner des analyses structurales et fonctionnelles des différents complexes protéine/ARN impliqués. Pour cela, la protéine Gag est produite en cellules de mammifères. Une telle production assure les modifications post-traductionnelles de la protéine qui pourraient jouer un rôle dans son interaction avec l'ARN. Dans ce cadre, j'ai réalisé une revue qui relate les différentes PTM des protéines et plus particulièrement les rôles des modifications de Gag dans le cycle viral et notamment dans les étapes d'assemblage (**annexe 1**). Les analyses de spectrométrie de masse en tandem (MS/MS) réalisées sur la protéine 6His-Gag montrent que les PTM (phosphorylations et ubiquitylations) sont reproductibles entre les différentes productions et sont en accord avec la littérature.

Des résultats très préliminaires de caractérisation de la protéine produite en cellules de mammifères avaient été réalisées au laboratoire avant le début de ce travail. Cependant, la caractérisation biophysique de la protéine dans différentes solutions tamponnées s'est révélée nécessaire dans la mise en place des conditions de formation des complexes Gag/ARN et de leurs études par CryoEM. Ainsi ce travail a permis d'optimiser la production de Gag dans les cellules BHK21 et de caractériser la protéine 6HisGag en déterminant son Rh, sa TM, son coefficient de sédimentation et son état oligomérique dans les tampons A1 et A3. En parallèle, je me suis intéressée aux conditions de formation de complexes entre la protéine 6His-Gag et l'ARNg. Mes résultats montrent que les complexes se forment entre un oligomère de Gag et un dimère d'ARN, ayant une affinité d'environ 174 nM pour l'ARN de 111 nucléotides et de 116 nM pour l'ARN SL1 dans le tampon A3. En outre, les conditions de formation de complexes sont cruciales pour la réalisation de grilles et l'obtention de la structure par Cryo-microscopie électronique (CryoEM). À ce jour, aucune structure de Gag en interaction avec l'ARNg n'est disponible. Dans ce but, je me suis focalisée sur les conditions optimales pour la préparation de grilles de CryoEM pour la protéine Gag seule, me permettant de réaliser un premier enregistrement et d'obtenir une structure préliminaire à une résolution de 7 Å.

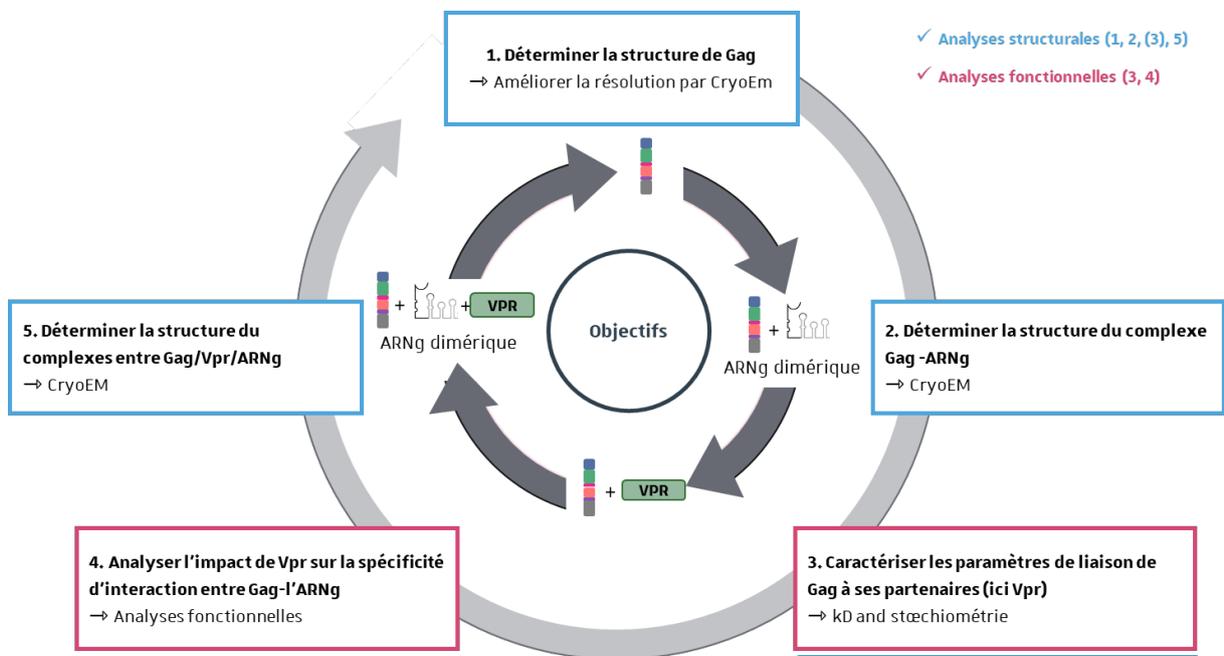
En outre, mon laboratoire d'accueil a montré que Gag se lie spécifiquement à l'ARNg avec une forte affinité tandis que la protéine dépourvue du domaine p6, Gag $\Delta$ p6 se lie aux ARN non génomiques de manière non spécifique avec une affinité plus faible (Dubois et al., 2018b). Ces résultats suggèrent que le domaine p6 de Gag pourrait réguler la spécificité de fixation de Gag aux ARN, qui au moment de l'assemblage doit devenir non-spécifique. Notre

hypothèse est que l'interaction de p6 avec ces partenaires module la spécificité d'interaction Gag/ARNg favorisant son encapsidation. En ce sens, j'ai mis au point la production des partenaires de Gag en utilisant le même système d'expression en cellules de mammifères. L'homogénéité et la stabilité des protéines sont des facteurs clés dans la détermination de leurs structures à haute résolution et dans la formation des complexes. Ainsi, j'ai aussi réalisé un contrôle qualité sur les partenaires de p6 en portant une attention particulière à l'état oligomérique et aux PTM de ces protéines.

En perspective à court terme, des grilles de CryoEM sont déjà préparées au laboratoire, elles peuvent être utilisées pour réaliser de nouvelles acquisitions au microscope Glacios afin d'augmenter le nombre de particules et ainsi la résolution de la structure de la protéine Gag. Obtenir cette structure serait un atout crucial pour résoudre la structure de Gag en interaction avec l'ARNg en réalisant notamment du remplacement moléculaire. Ainsi le deuxième objectif à court terme est d'améliorer les conditions de formation de complexes entre Gag et l'ARNg en trouvant des conditions qui favorisent la proportion de complexes par rapport aux protéines seules en solution. Pour cela la formation de complexes par dialyse continue peut-être envisagée. Cette méthode consiste à utiliser une pompe qui va prélever le tampon final pour échanger progressivement le tampon initial. Cela présente l'avantage de permettre une dialyse progressive et de favoriser les contacts entre les molécules en diminuant la concentration de NaCl de manière moins brusque qu'avec les méthodes classiques. En parallèle, la fixation d'un des partenaires sur une surface peut être une solution pour empêcher le phénomène d'agrégation et favoriser les interactions spécifiques entre les partenaires. Pour cela, des analyses de SPR pourraient être réalisées. Une fois les conditions de formation de complexes améliorées, elles pourront être utilisées pour réaliser les grilles de CryoEM. Si la structure par CryoEM devait présenter des difficultés, il y a la possibilité de réaliser des enregistrements par diffusion des rayons X aux petits angles (SAXS) combinés à de la spectrométrie de masse structurale (HDX-MS) cela permettra de déterminer une enveloppe à basse résolution qui donnera une information sur l'agencement des domaines et les zones d'interactions.

En parallèle, les conditions de formations de complexes entre Vpr et la protéine Gag doivent être approfondies. Notamment entre les deux partenaires produits en cellules BHK21. Pour cela, il est envisagé de mettre au point le vecteur viral permettant l'expression des deux partenaires à partir d'un gène polycistronique. Les complexes pourraient être purifiés dans le tampon A2 comme décrit pour les partenaires séparément.

A moyen terme, des analyses pour comparer la capacité de fixation de Gag à l'ARNg en présence et en absence des partenaires de p6 seront réalisées. La prochaine étape est donc d'optimiser les conditions de la formation du complexe Gag/ARN/Vpr et d'analyser leurs paramètres de liaison (stœchiométrie et constante de dissociation). Ces études fonctionnelles permettront d'établir l'influence du partenaire sur la spécificité de liaison entre Gag et l'ARNg. Les contrôles seront réalisés à l'aide d'un fragment d'ARN incapable de dimériser (ARN111<sub>AAA</sub>) et de la protéine Gag donc le domaine de liaison à Vpr aura été délété, Gag $\Delta$ p6. En parallèle, des tests de résolution de la structure par CryoEM du complexe seront réalisés. Ainsi les études fonctionnelles et structurales des différents complexes donneront des informations cruciales pour comprendre comment l'ARN est recruté par la protéine Gag et quel est le rôle des partenaires de p6 dans cette interaction. Une fois la stratégie d'étude mise au point pour la protéine Vpr, elle pourra être utilisée avec les partenaires cellulaires ALIX et TSG101 (**figure 88**).



**Figure 88. Stratégie expérimentales proposée pour apporter une meilleure compréhension des interactions entre Gag et l'ARNg permettant l'encapsidation spécifique de l'ARNg et du rôle des partenaires de p6 dans l'encapsidation spécifique de l'ARNg**

Notre stratégie consiste à combiner des analyses structurales et fonctionnelles des différents complexes protéines/ARN pour déterminer de l'impact de l'interaction des partenaires de p6 sur la spécificité de liaison de Gag à l'ARNg. Les étapes qui nécessitent une analyse structurale par CryoEM sont encadrées en bleu, celles nécessitant des analyses fonctionnelles sont encadrées en roses.

Les résultats obtenus avec la protéine soluble 6His-Gag pourront être comparés à ceux obtenus avec la protéine myristoylée mis au point au laboratoire. En outre, la mise au point d'un test permettant de déterminer 1) l'efficacité de l'encapsidation de l'ARNg par rapport à des autres ARN viraux et cellulaires et 2) le rôle des partenaires de p6 sur l'encapsidation de l'ARNg. Ce test s'appuierait sur la capacité de Gag à s'assembler en VLP en présence d'acides nucléiques non spécifiques, comme l'ARNt et en absence de membranes (Chameettachal et al., 2018; McKinstry et al., 2014). Ainsi il sera possible d'analyser les aliquotes des VLP par coloration négative et Cryo Tomographie. Une fois que les conditions nécessaires à la formation des VLP seront mises au point, cette méthode permettra de vérifier si les ARNg assemblent plus efficacement les VLP que les ARN hétérologues et de déterminer l'impact des partenaires de p6 sur la formation de VLP.

Dans l'ensemble ces informations devraient permettre à plus long terme de dessiner des nouveaux inhibiteurs dirigés contre l'encapsidation de l'ARNg du VIH-1.





# Bibliographie

- Aldovini, A., Young, R.A., 1990. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J. Virol.* 64, 1920–1926. <https://doi.org/10.1128/JVI.64.5.1920-1926.1990>
- Amartely, H., Avraham, O., Friedler, A., Livnah, O., Lebendiker, M., 2018. Coupling Multi Angle Light Scattering to Ion Exchange chromatography (IEX-MALS) for protein characterization. *Sci. Rep.* 8, 6907. <https://doi.org/10.1038/s41598-018-25246-6>
- Auvert, B., Taljaard, D., Lagarde, E., Sobngwi-Tambekou, J., Sitta, R., Puren, A., 2005. Randomized, Controlled Intervention Trial of Male Circumcision for Reduction of HIV Infection Risk: The ANRS 1265 Trial. *PLoS Med.* 2, e298. <https://doi.org/10.1371/journal.pmed.0020298>
- Auxilien, S., Keith, G., Le Grice, S.F., Darlix, J.L., 1999. Role of post-transcriptional modifications of primer tRNA<sup>Lys,3</sup> in the fidelity and efficacy of plus strand DNA transfer during HIV-1 reverse transcription. *J. Biol. Chem.* 274, 4412–4420. <https://doi.org/10.1074/jbc.274.7.4412>
- Ayouba, A., Souquières, S., Njinku, B., Martin, P.M.V., Müller-Trutwin, M.C., Roques, P., Barré-Sinoussi, F., Maucière, P., Simon, F., Nerrienet, E., 2000. HIV-1 group N among HIV-1-seropositive individuals in Cameroon: *AIDS* 14, 2623–2625. <https://doi.org/10.1097/00002030-200011100-00033>
- Bachand, F., Yao, X.-J., Hrimech, M., Rougeau, N., Cohen, É.A., 1999. Incorporation of Vpr into Human Immunodeficiency Virus Type 1 Requires a Direct Interaction with the p6 Domain of the p55 Gag Precursor \*. *J. Biol. Chem.* 274, 9083–9091. <https://doi.org/10.1074/jbc.274.13.9083>
- Baird, H.A., Galetto, R., Gao, Y., Simon-Lorière, E., Abreha, M., Archer, J., Fan, J., Robertson, D.L., Arts, E.J., Negroni, M., 2006. Sequence determinants of breakpoint location during HIV-1 intersubtype recombination. *Nucleic Acids Res.* 34, 5203–5216. <https://doi.org/10.1093/nar/gkl669>
- Barnitz, R.A., Wan, F., Tripuraneni, V., Bolton, D.L., Lenardo, M.J., 2010. Protein kinase A phosphorylation activates Vpr-induced cell cycle arrest during human immunodeficiency virus type 1 infection. *J. Virol.* 84, 6410–6424. <https://doi.org/10.1128/JVI.02273-09>
- Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 220, 868–871. <https://doi.org/10.1126/science.6189183>
- Bartesaghi, A., Merk, A., Borgnia, M.J., Milne, J.L.S., Subramaniam, S., 2013. Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* 20, 1352–1357. <https://doi.org/10.1038/nsmb.2711>
- Basmaciogullari, S., Pizzato, M., 2014. The activity of Nef on HIV-1 infectivity. *Front. Microbiol.* 5, 232. <https://doi.org/10.3389/fmicb.2014.00232>
- Basu, V.P., Song, M., Gao, L., Rigby, S.T., Hanson, M.N., Bambara, R.A., 2008. Strand transfer events during HIV-1 reverse transcription. *Virus Res., Retroviral Reverse Transcription* 134, 19–38. <https://doi.org/10.1016/j.virusres.2007.12.017>
- Bbosa, N., Kaleebu, P., Ssemwanga, D., 2019. HIV subtype diversity worldwide. *Curr. Opin. HIV AIDS* 14, 153–160. <https://doi.org/10.1097/COH.0000000000000534>
- Bell, I., Ashman, C., Maughan, J., Hooker, E., Cook, F., Reinhart, T.A., 1998. Association of simian immunodeficiency virus Nef with the T-cell receptor (TCR) zeta chain leads to TCR down-modulation. *J. Gen. Virol.* 79 ( Pt 11), 2717–2727. <https://doi.org/10.1099/0022-1317-79-11-2717>
- Berkowitz, R.D., Luban, J., Goff, S.P., 1993. Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J. Virol.* 67, 7190–7200. <https://doi.org/10.1128/JVI.67.12.7190-7200.1993>
- Bernacchi, S., Abd El-Wahab, E.W., Dubois, N., Hijnen, M., Smyth, R.P., Mak, J., Marquet, R., Paillart, J.-C., 2017. HIV-1 Pr55Gag binds genomic and spliced RNAs with different affinity and stoichiometry. *RNA Biol.* 14, 90–103. <https://doi.org/10.1080/15476286.2016.1256533>
- Bewley, M.C., Reinhart, L., Stake, M.S., Nadaraja-Hoke, S., Parent, L.J., Flanagan, J.M., 2017. A non-cleavable hexahistidine affinity tag at the carboxyl-terminus of the HIV-1 Pr55Gag polyprotein alters nucleic acid binding properties. *Protein Expr. Purif.* 130, 137–145. <https://doi.org/10.1016/j.pep.2016.10.001>
- Blakemore, R.J., Burnett, C., Swanson, C., Kharytonchyk, S., Telesnitsky, A., Munro, J.B., 2021. Stability and conformation of the dimeric HIV-1 genomic RNA 5'UTR. *Biophys. J.* 120, 4874–4890. <https://doi.org/10.1016/j.bpj.2021.09.017>

- Bleul, C.C., Wu, L., Hoxie, J.A., Springer, T.A., Mackay, C.R., 1997. The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* 94, 1925–1930. <https://doi.org/10.1073/pnas.94.5.1925>
- Bourbigot, S., Beltz, H., Denis, J., Morellet, N., Roques, B.P., Mély, Y., Bouaziz, S., 2005. The C-terminal domain of the HIV-1 regulatory protein Vpr adopts an antiparallel dimeric structure in solution via its leucine-zipper-like domain. *Biochem. J.* 387, 333–341. <https://doi.org/10.1042/BJ20041759>
- Boutant, E., Bonzi, J., Anton, H., Nasim, M.B., Cathagne, R., Réal, E., Dujardin, D., Carl, P., Didier, P., Paillart, J.-C., Marquet, R., Mély, Y., de Rocquigny, H., Bernacchi, S., 2020. The two zinc fingers in the nucleocapsid domain of the HIV-1 Gag precursor are equivalent for the interaction with the genomic RNA in the cytoplasm, but not for the recruitment of the complexes at the plasma membrane. *Biophys. J.* <https://doi.org/10.1016/j.bpj.2020.05.035>
- Brautigam, C.A., 2015. Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data, in: Cole, J.L. (Ed.), *Methods in Enzymology, Analytical Ultracentrifugation*. Academic Press, 5, 109–133. <https://doi.org/10.1016/bs.mie.2015.05.001>
- Brown, J.D., Kharytonchyk, S., Chaudry, I., Iyer, A.S., Carter, H., Becker, G., Desai, Y., Glang, L., Choi, S.H., Singh, K., Lopresti, M.W., Orellana, M., Rodriguez, T., Oboh, U., Hijji, J., Ghinger, F.G., Stewart, K., Francis, D., Edwards, B., Chen, P., Case, D.A., Telesnitsky, A., Summers, M.F., 2020. Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science* 368, 413–417. <https://doi.org/10.1126/science.aaz7959>
- Burdick, R.C., Li, C., Munshi, M., Rawson, J.M.O., Nagashima, K., Hu, W.-S., Pathak, V.K., 2020. HIV-1 uncoats in the nucleus near sites of integration. *Proc. Natl. Acad. Sci. U. S. A.* 117, 5486–5493. <https://doi.org/10.1073/pnas.1920631117>
- Bussienne, C., Marquet, R., Paillart, J.-C., Bernacchi, S., 2021. Post-Translational Modifications of Retroviral HIV-1 Gag Precursors: An Overview of Their Biological Role. *Int. J. Mol. Sci.* 22. <https://doi.org/10.3390/ijms22062871>
- Carnes, S.K., Zhou, J., Aiken, C., 2018. HIV-1 Engages a Dynein-Dynactin-BICD2 Complex for Infection and Transport to the Nucleus. *J. Virol.* 92. <https://doi.org/10.1128/JVI.00358-18>
- Cartier, C., Deckert, M., Grangeasse, C., Trauger, R., Jensen, F., Bernard, A., Cozzone, A., Desgranges, C., Boyer, V., 1997. Association of ERK2 mitogen-activated protein kinase with human immunodeficiency virus particles. *J. Virol.* 71, 4832–4837.
- Cavallari, I., Rende, F., D'Agostino, D.M., Ciminale, V., 2011. Converging Strategies in Expression of Human Complex Retroviruses. *Viruses* 3, 1395–1414. <https://doi.org/10.3390/v3081395>
- Cervený, L., Murthi, P., Staud, F., 2021. HIV in pregnancy: Mother-to-child transmission, pharmacotherapy, and toxicity. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* 1867, 166206. <https://doi.org/10.1016/j.bbadis.2021.166206>
- Chameettachal, A., Pillai, V.N., Ali, L.M., Pitchai, F.N.N., Ardah, M.T., Mustafa, F., Marquet, R., Rizvi, T.A., 2018. Biochemical and Functional Characterization of Mouse Mammary Tumor Virus Full-Length Pr77Gag Expressed in Prokaryotic and Eukaryotic Cells. *Viruses* 10, 334. <https://doi.org/10.3390/v10060334>
- Chamontin, C., Rassam, P., Ferrer, M., Racine, P.-J., Neyret, A., Lainé, S., Milhiet, P.-E., Mougel, M., 2015. HIV-1 nucleocapsid and ESCRT-component Tsg101 interplay prevents HIV from turning into a DNA-containing virus. *Nucleic Acids Res.* 43, 336–347. <https://doi.org/10.1093/nar/gku1232>
- Chen, J., Grunwald, D., Sardo, L., Galli, A., Plisov, S., Nikolaitchik, O.A., Chen, D., Lockett, S., Larson, D.R., Pathak, V.K., Hu, W.-S., 2014. Cytoplasmic HIV-1 RNA is mainly transported by diffusion in the presence or absence of Gag protein. *Proc. Natl. Acad. Sci.* 111, E5205–E5213. <https://doi.org/10.1073/pnas.1413169111>
- Chen, J., Rahman, S.A., Nikolaitchik, O.A., Grunwald, D., Sardo, L., Burdick, R.C., Plisov, S., Liang, E., Tai, S., Pathak, V.K., Hu, W.-S., 2016. HIV-1 RNA genome dimerizes on the plasma membrane in the presence of Gag protein. *Proc. Natl. Acad. Sci.* 113, E201–E208. <https://doi.org/10.1073/pnas.1518572113>
- Chen, L., Kwon, Y.D., Zhou, T., Wu, X., O'Dell, S., Cavacini, L., Hessell, A.J., Pancera, M., Tang, M., Xu, L., Yang, Z.-Y., Zhang, M.-Y., Arthos, J., Burton, D.R., Dimitrov, D.S., Nabel, G.J., Posner, M.R., Sodroski, J., Wyatt, R., Mascola, J.R., Kwong, P.D., 2009. Structural basis of immune evasion at the site of CD4 attachment on HIV-1 gp120. *Science* 326, 1123–1127. <https://doi.org/10.1126/science.1175868>
- Chen, Z., Cheng, W., 2017. Reversible aggregation of HIV-1 Gag proteins mediated by nucleic acids. *Biochem. Biophys. Res. Commun.* 482, 1437–1442. <https://doi.org/10.1016/j.bbrc.2016.12.054>
- Chouaib, R., Safieddine, A., Pichon, X., Imbert, A., Kwon, O.S., Samacoits, A., Traboulsi, A.-M., Robert, M.-C., Tsanov, N., Coleno, E., Poser, I., Zimmer, C., Hyman, A., Le Hir, H., Zibara, K., Peter, M., Mueller, F., Walter, T., Bertrand, E., 2020. A Dual Protein-mRNA Localization Screen Reveals Compartmentalized Translation and Widespread Co-translational RNA Targeting. *Dev. Cell* 54, 773–791.e5. <https://doi.org/10.1016/j.devcel.2020.07.010>

- Chukkapalli, V., Hogue, I.B., Boyko, V., Hu, W.-S., Ono, A., 2008. Interaction between the Human Immunodeficiency Virus Type 1 Gag Matrix Domain and Phosphatidylinositol-(4,5)-Bisphosphate Is Essential for Efficient Gag Membrane Binding. *J. Virol.* 82, 2405–2417. <https://doi.org/10.1128/JVI.01614-07>
- Cihlar, T., Fordyce, M., 2016. Current status and prospects of HIV treatment. *Curr. Opin. Virol.* 18, 50–56. <https://doi.org/10.1016/j.coviro.2016.03.004>
- Cimarelli, A., Sandin, S., Höglund, S., Luban, J., 2000. Basic Residues in Human Immunodeficiency Virus Type 1 Nucleocapsid Promote Virion Assembly via Interaction with RNA. *J. Virol.* 74, 3046–3057.
- Clavel, F., Guétard, D., Brun-Vézinet, F., Chamaret, S., Rey, M.A., Santos-Ferreira, M.O., Laurent, A.G., Dauguet, C., Katlama, C., Rouzioux, C., 1986. Isolation of a new human retrovirus from West African patients with AIDS. *Science* 233, 343–346. <https://doi.org/10.1126/science.2425430>
- Clavel, F., Orenstein, J.M., 1990. A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J. Virol.* 64, 5230–5234.
- Coffin, J., Swanstrom, R., 2013. HIV Pathogenesis: Dynamics and Genetics of Viral Populations and Infected Cells. *Cold Spring Harb. Perspect. Med.* 3, a012526. <https://doi.org/10.1101/cshperspect.a012526>
- Coffin, J.M., Hughes, S.H., Varmus, H.E., 1997. *A Brief Chronicle of Retrovirology*, Retroviruses. Cold Spring Harbor Laboratory Press.
- Collins, K.L., Chen, B.K., Kalams, S.A., Walker, B.D., Baltimore, D., 1998. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature* 391, 397–401. <https://doi.org/10.1038/34929>
- Cornille, F., Wecker, K., Loffet, A., Genet, R., Roques, B., 1999. Efficient solid-phase synthesis of Vpr from HIV-1 using low quantities of uniformly <sup>13</sup>C-, <sup>15</sup>N-labeled amino acids for NMR structural studies. *J. Pept. Res. Off. J. Am. Pept. Soc.* 54, 427–435. <https://doi.org/10.1034/j.1399-3011.1999.00129.x>
- Craigie, R., 2012. The molecular biology of HIV integrase. *Future Virol.* 7, 679–686. <https://doi.org/10.2217/FVL.12.56>
- Crawford, L.V., Crawford, E.M., 1961. The properties of Rous sarcoma virus purified by density gradient centrifugation. *Virology* 13, 227–232. [https://doi.org/10.1016/0042-6822\(61\)90056-3](https://doi.org/10.1016/0042-6822(61)90056-3)
- Crucifix, C., Uhring, M., Schultz, P., 2004. Immobilization of biotinylated DNA on 2-D streptavidin crystals. *J. Struct. Biol.* 146, 441–451. <https://doi.org/10.1016/j.jsb.2004.02.001>
- Cujec, T.P., Cho, H., Maldonado, E., Meyer, J., Reinberg, D., Peterlin, B.M., 1997. The human immunodeficiency virus transactivator Tat interacts with the RNA polymerase II holoenzyme. *Mol. Cell. Biol.* 17, 1817–1823.
- Cullen, B.R., 1998. Retroviruses as model systems for the study of nuclear RNA export pathways. *Virology* 249, 203–210. <https://doi.org/10.1006/viro.1998.9331>
- D'arc, M., Ayouba, A., Esteban, A., Learn, G.H., Boué, V., Liegeois, F., Etienne, L., Tagg, N., Leendertz, F.H., Boesch, C., Madinda, N.F., Robbins, M.M., Gray, M., Cournil, A., Ooms, M., Letko, M., Simon, V.A., Sharp, P.M., Hahn, B.H., Delaporte, E., Mpoudi Ngole, E., Peeters, M., 2015. Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proc. Natl. Acad. Sci. U. S. A.* 112, E1343–1352. <https://doi.org/10.1073/pnas.1502022112>
- Das, A.T., Klaver, B., Berkhout, B., 1999. A Hairpin Structure in the R Region of the Human Immunodeficiency Virus Type 1 RNA Genome Is Instrumental in Polyadenylation Site Selection. *J. Virol.* 73, 81–91.
- Dash, P.K., Kaminski, R., Bella, R., Su, H., Mathews, S., Ahooyi, T.M., Chen, C., Mancuso, P., Sariyer, R., Ferrante, P., Donadoni, M., Robinson, J.A., Sillman, B., Lin, Z., Hilaire, J.R., Banoub, M., Elango, M., Gautam, N., Mosley, R.L., Poluektova, L.Y., McMillan, J., Bade, A.N., Gorantla, S., Sariyer, I.K., Burdo, T.H., Young, W.-B., Amini, S., Gordon, J., Jacobson, J.M., Edagwa, B., Khalili, K., Gendelman, H.E., 2019. Sequential LASER ART and CRISPR Treatments Eliminate HIV-1 in a Subset of Infected Humanized Mice. *Nat. Commun.* 10, 2753. <https://doi.org/10.1038/s41467-019-10366-y>
- Datta, S.A.K., Clark, P.K., Fan, L., Ma, B., Harvin, D.P., Sowder, R.C., Nussinov, R., Wang, Y.-X., Rein, A., 2016. Dimerization of the SP1 Region of HIV-1 Gag Induces a Helical Conformation and Association into Helical Bundles: Implications for Particle Assembly. *J. Virol.* 90, 1773–1787. <https://doi.org/10.1128/JVI.02061-15>
- Datta, S.A.K., Curtis, J.E., Ratcliff, W., Clark, P.K., Crist, R.M., Lebowitz, J., Krueger, S., Rein, A., 2007. Conformation of the HIV-1 Gag Protein in Solution. *J. Mol. Biol.* 365, 812–824. <https://doi.org/10.1016/j.jmb.2006.10.073>
- Datta, S.A.K., Temeselew, L.G., Crist, R.M., Soheilian, F., Kamata, A., Mirro, J., Harvin, D., Nagashima, K., Cachau, R.E., Rein, A., 2011. On the Role of the SP1 Domain in HIV-1 Particle Assembly: a Molecular Switch? *J. Virol.* 85, 4111–4121. <https://doi.org/10.1128/JVI.00006-11>
- De Oliveira, F., Cappy, P., Lemée, V., Moisan, A., Pronier, C., Bocket, L., Bouvier-Alias, M., Chaix, M.-L., Gault, E., Morvan, O., Poveda, J.-D., Schneider, V., Wirlden, M., Alessandri-Gradt, E., Mourez, T., Plantier, J.-C., 2018. Detection of numerous HIV-1/MO recombinants in France. *AIDS Lond. Engl.* 32, 1289–1299. <https://doi.org/10.1097/QAD.0000000000001814>

- De Oliveira, F., Mourez, T., Vessiere, A., Ngoupo, P.-A., Alessandri-Gradt, E., Simon, F., Rousset, D., Plantier, J.-C., 2017. Multiple HIV-1/M + HIV-1/O dual infections and new HIV-1/MO inter-group recombinant forms detected in Cameroon. *Retrovirology* 14, 1. <https://doi.org/10.1186/s12977-016-0324-3>
- Deeks, S.G., Walker, B.D., 2004. The immune response to AIDS virus infection: good, bad, or both? *J. Clin. Invest.* 113, 808–810. <https://doi.org/10.1172/JCI200421318>
- Dey, S., Batisse, C., Shukla, J., Webster, M.W., Takacs, M., Saint-André, C., Weixlbaumer, A., 2022. Structural insights into RNA-mediated transcription regulation in bacteria. *Mol. Cell* 82, 3885–3900.e10. <https://doi.org/10.1016/j.molcel.2022.09.020>
- Dharan, A., Bachmann, N., Talley, S., Zwickelmaier, V., Campbell, E.M., 2020. Nuclear pore blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus. *Nat. Microbiol.* 5, 1088–1095. <https://doi.org/10.1038/s41564-020-0735-8>
- Dharan, A., Talley, S., Tripathi, A., Mamede, J.I., Majetschak, M., Hope, T.J., Campbell, E.M., 2016. KIF5B and Nup358 Cooperatively Mediate the Nuclear Import of HIV-1 during Infection. *PLoS Pathog.* 12, e1005700. <https://doi.org/10.1371/journal.ppat.1005700>
- Dilley, K.A., Ni, N., Nikolaitchik, O.A., Chen, J., Galli, A., Hu, W.-S., 2011. Determining the frequency and mechanisms of HIV-1 and HIV-2 RNA copackaging by single-virion analysis. *J. Virol.* 85, 10499–10508. <https://doi.org/10.1128/JVI.05147-11>
- Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Heaphy, S., Karn, J., Lowe, A.D., Singh, M., Skinner, M.A., Valerio, R., 1989. Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 86, 6925–6929. <https://doi.org/10.1073/pnas.86.18.6925>
- Dowlatshahi, D.P., Sandrin, V., Vivona, S., Shaler, T.A., Kaiser, S.E., Melandri, F., Sundquist, W.I., Kopito, R.R., 2012. ALIX is a Lys63-specific Polyubiquitin Binding Protein That Functions in Retrovirus Budding. *Dev. Cell* 23, 1247–1254. <https://doi.org/10.1016/j.devcel.2012.10.023>
- Drillien, R., Pradeau-Aubretton, K., Batisse, J., Mezher, J., Schenckbecher, E., Marguin, J., Ennifar, E., Ruff, M., 2022. Efficient production of protein complexes in mammalian cells using a poxvirus vector. *PloS One* 17, e0279038. <https://doi.org/10.1371/journal.pone.0279038>
- Dubois, N., Khoo, K.K., Ghossein, S., Seissler, T., Wolff, P., McKinstry, W.J., Mak, J., Paillart, J.-C., Marquet, R., Bernacchi, S., 2018a. The C-terminal p6 domain of the HIV-1 Pr55Gag precursor is required for specific binding to the genomic RNA. *RNA Biol.* 15, 923–936. <https://doi.org/10.1080/15476286.2018.1481696>
- Dubois, N., Marquet, R., Paillart, J.-C., Bernacchi, S., 2018b. Retroviral RNA Dimerization: From Structure to Functions. *Front. Microbiol.* 9, 1–19. <https://doi.org/10.3389/fmicb.2018.00527>
- Eisele, E., Siliciano, R.F., 2012. Redefining the viral reservoirs that prevent HIV-1 eradication. *Immunity* 37, 377–388. <https://doi.org/10.1016/j.immuni.2012.08.010>
- El-Amine, R., Germini, D., Zakharova, V.V., Tsfasman, T., Sheval, E.V., Louzada, R.A.N., Dupuy, C., Bilhou-Nabera, C., Hamade, A., Najjar, F., Oksenhendler, E., Lipinski, M., Chernyak, B.V., Vassetzky, Y.S., 2018. HIV-1 Tat protein induces DNA damage in human peripheral blood B-lymphocytes via mitochondrial ROS production. *Redox Biol.* 15, 97–108. <https://doi.org/10.1016/j.redox.2017.11.024>
- Ellermann, V., Bang, O., 1909. Experimentelle Leukämie bei Hühnern. II. *Z. Für Hyg. Infekt.* 63, 231–272. <https://doi.org/10.1007/BF02227892>
- El-Wahab, E.W.A., Smyth, R.P., Mailler, E., Bernacchi, S., Vivet-Boudou, V., Hijnen, M., Jossinet, F., Mak, J., Paillart, J.-C., Marquet, R., 2014. Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat. Commun.* 5, 1–13. <https://doi.org/10.1038/ncomms5304>
- Engelman, A., Cherepanov, P., 2014. Retroviral Integrase Structure and DNA Recombination Mechanism. *Microbiol. Spectr.* 2, 1–22. <https://doi.org/10.1128/microbiolspec.MDNA3-0024-2014>
- Engelman, A., Mizuuchi, K., Craigie, R., 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 67, 1211–1221. [https://doi.org/10.1016/0092-8674\(91\)90297-c](https://doi.org/10.1016/0092-8674(91)90297-c)
- Esté, J.A., Cabrera, C., Blanco, J., Gutierrez, A., Bridger, G., Henson, G., Clotet, B., Schols, D., De Clercq, E., 1999. Shift of Clinical Human Immunodeficiency Virus Type 1 Isolates from X4 to R5 and Prevention of Emergence of the Syncytium-Inducing Phenotype by Blockade of CXCR4. *J. Virol.* 73, 5577–5585.
- Farjot, G., Sergeant, A., Mikaélian, I., 1999. A New Nucleoporin-like Protein Interacts with Both HIV-1 Rev Nuclear Export Signal and CRM-1. *J. Biol. Chem.* 274, 17309–17317. <https://doi.org/10.1074/jbc.274.24.17309>
- Ferrer, M., Clerté, C., Chamontin, C., Basyuk, E., Lainé, S., Hottin, J., Bertrand, E., Margeat, E., Mougél, M., 2016. Imaging HIV-1 RNA dimerization in cells by multicolor super-resolution and fluctuation microscopies. *Nucleic Acids Res.* 44, 7922–7934. <https://doi.org/10.1093/nar/gkw511>

- Fisher, R.D., Chung, H.-Y., Zhai, Q., Robinson, H., Sundquist, W.I., Hill, C.P., 2007. Structural and Biochemical Studies of ALIX/AIP1 and Its Role in Retrovirus Budding. *Cell* 128, 841–852. <https://doi.org/10.1016/j.cell.2007.01.035>
- Forouzanfar, F., Ali, S., Wallet, C., De Rovere, M., Ducloy, C., El Mekdad, H., El Maassarani, M., Aït-Ammar, A., Van Assche, J., Boutant, E., Daouad, F., Margottin-Goguet, F., Moog, C., Van Lint, C., Schwartz, C., Rohr, O., 2019. HIV-1 Vpr mediates the depletion of the cellular repressor CTIP2 to counteract viral gene silencing. *Sci. Rep.* 9, 13154. <https://doi.org/10.1038/s41598-019-48689-x>
- Francis, A.C., Marin, M., Singh, P.K., Achuthan, V., Prellberg, M.J., Palermino-Rowland, K., Lan, S., Tedbury, P.R., Sarafianos, S.G., Engelman, A.N., Melikyan, G.B., 2020. HIV-1 replication complexes accumulate in nuclear speckles and integrate into speckle-associated genomic domains. *Nat. Commun.* 11, 3505. <https://doi.org/10.1038/s41467-020-17256-8>
- Freed, E.O., 2015. HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.* 13, 484–496.
- Fritz, J.V., Dujardin, D., Godet, J., Didier, P., De Mey, J., Darlix, J.-L., Mély, Y., de Rocquigny, H., 2010. HIV-1 Vpr Oligomerization but Not That of Gag Directs the Interaction between Vpr and Gag. *J. Virol.* 84, 1585–1596. <https://doi.org/10.1128/JVI.01691-09>
- Fujii, K., Munshi, U.M., Ablan, S.D., Demirov, D.G., Soheilian, F., Nagashima, K., Stephen, A.G., Fisher, R.J., Freed, E.O., 2009. Functional Role of Alix in Hiv-1 Replication. *Virology* 391, 284–292. <https://doi.org/10.1016/j.virol.2009.06.016>
- Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M., Hahn, B.H., 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes*. *Nature* 397, 436–441. <https://doi.org/10.1038/17130>
- Garrus, J.E., von Schwedler, U.K., Pornillos, O.W., Morham, S.G., Zavitz, K.H., Wang, H.E., Wettstein, D.A., Stray, K.M., Côté, M., Rich, R.L., Myszka, D.G., Sundquist, W.I., 2001. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* 107, 55–65. [https://doi.org/10.1016/s0092-8674\(01\)00506-2](https://doi.org/10.1016/s0092-8674(01)00506-2)
- Geijtenbeek, T.B., Kwon, D.S., Torensma, R., van Vliet, S.J., van Duynhoven, G.C., Middel, J., Cornelissen, I.L., Nottet, H.S., KewalRamani, V.N., Littman, D.R., Figdor, C.G., van Kooyk, Y., 2000. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* 100, 587–597. [https://doi.org/10.1016/s0092-8674\(00\)80694-7](https://doi.org/10.1016/s0092-8674(00)80694-7)
- Gilmer, O., Mailler, E., Paillart, J.-C., Mouhand, A., Tisné, C., Mak, J., Smyth, R.P., Marquet, R., Vivet-Boudou, V., 2021. Structural maturation of the HIV-1 RNA 5' untranslated region by Pr55Gag and its maturation products. *RNA Biol.* 19, 191–205. <https://doi.org/10.1080/15476286.2021.2021677>
- Giudice, R.D., Paracini, N., Laursen, T., Blanchet, C., Roosen-Runge, F., Cárdenas, M., 2022. Expanding the Toolbox for Bicelle-Forming Surfactant-Lipid Mixtures. *Molecules* 27, 7628. <https://doi.org/10.3390/molecules27217628>
- Godet, J., Mély, Y., 2010. Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol.* 7, 687–699. <https://doi.org/10.4161/rna.7.6.13616>
- Göttlinger, H.G., Dorfman, T., Sodroski, J.G., Haseltine, W.A., 1991. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc. Natl. Acad. Sci. U. S. A.* 88, 3195–3199. <https://doi.org/10.1073/pnas.88.8.3195>
- Gupta, A., Verma, A., Kashyap, M., Gautam, P., 2020. ART in Prevention of Mother-to-Child Transmission of HIV. *J. Obstet. Gynaecol. India* 70, 18–22. <https://doi.org/10.1007/s13224-019-01263-x>
- Harrison, G.P., Miele, G., Hunter, E., Lever, A.M.L., 1998. Functional Analysis of the Core Human Immunodeficiency Virus Type 1 Packaging Signal in a Permissive Cell Line. *J. Virol.* 72, 5886–5896.
- Hebben, M., Brants, J., Birck, C., Samama, J.-P., Wasyluk, B., Spehner, D., Pradeau, K., Domi, A., Moss, B., Schultz, P., Drillien, R., 2007. High level protein expression in mammalian cells using a safe viral vector: modified vaccinia virus Ankara. *Protein Expr. Purif.* 56, 269–278. <https://doi.org/10.1016/j.pep.2007.08.003>
- Hemelaar, J., Gouws, E., Ghys, P.D., Osmanov, S., 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS Lond. Engl.* 20, W13–23. <https://doi.org/10.1097/01.aids.0000247564.73009.bc>
- Hemonnot, B., Cartier, C., Gay, B., Rebuffat, S., Bardy, M., Devaux, C., Boyer, V., Briant, L., 2004. The host cell MAP kinase ERK-2 regulates viral assembly and release by phosphorylating the p6gag protein of HIV-1. *J. Biol. Chem.* 279, 32426–32434. <https://doi.org/10.1074/jbc.M313137200>
- Hendrix, J., Baumgärtel, V., Schrimpf, W., Ivanchenko, S., Digman, M.A., Gratton, E., Kräusslich, H.-G., Müller, B., Lamb, D.C., 2015. Live-cell observation of cytosolic HIV-1 assembly onset reveals RNA-interacting Gag oligomers. *J. Cell Biol.* 210, 629–646. <https://doi.org/10.1083/jcb.201504006>
- Hladik, F., McElrath, M.J., 2008. Setting the stage – HIV host invasion. *Nat. Rev. Immunol.* 8, 447–457. <https://doi.org/10.1038/nri2302>

- Hogue, I., Llewellyn, G., Ono, A., 2012. Dynamic Association between HIV-1 Gag and Membrane Domains. *Mol. Biol. Int.* 2012, 979765. <https://doi.org/10.1155/2012/979765>
- Houzet, L., Paillart, J.C., Smagulova, F., Maurel, S., Morichaud, Z., Marquet, R., Mougel, M., 2007. HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. *Nucleic Acids Res.* 35, 2695–2704. <https://doi.org/10.1093/nar/gkm153>
- Hu, W.-S., Hughes, S.H., 2012. HIV-1 Reverse Transcription. *Cold Spring Harb. Perspect. Med.* 2, a006882. <https://doi.org/10.1101/cshperspect.a006882>
- Hu, W.S., Temin, H.M., 1990. Retroviral recombination and reverse transcription. *Science* 250, 1227–1233. <https://doi.org/10.1126/science.1700865>
- Hurt, C.B., Nelson, J.A.E., Hightown-Weidman, L.B., Miller, W.C., 2017. Selecting an HIV Test: A Narrative Review for Clinicians and Researchers. *Sex. Transm. Dis.* 44, 739–746. <https://doi.org/10.1097/OLQ.0000000000000719>
- Hymes, Kenneth B., Greene, Jeffrey B., Marcus, A., William, Daniel C., Cheung, T., Prose, Neil S., Ballard, H., Laubenstein, Linda J., 1981. Kaposi's Sarcoma In Homosexual Men—A Report Of Eight Cases. *The Lancet*, 2, Issue 8247 318, 598–600. [https://doi.org/10.1016/S0140-6736\(81\)92740-9](https://doi.org/10.1016/S0140-6736(81)92740-9)
- Ivancjenko, S., Godinez, W., Lampe, M., Kräusslich, H.-G., Eils, R., Rohr, K., Bräuchle, C., Müller, B., Lamb, D., C., 2009. Dynamics of HIV-1 assembly and release. *PLoS Pathog.* 5, 11 : e1000652 <https://doi.org/10.1371/journal.ppat.1000652>
- Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J., Varmus, H.E., 1988. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* 331, 280–283. <https://doi.org/10.1038/331280a0>
- Jacobo-Molina, A., Ding, J., Nanni, R.G., Clark, A.D., Lu, X., Tantillo, C., Williams, R.L., Kamer, G., Ferris, A.L., Clark, P., 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. U. S. A.* 90, 6320–6324. <https://doi.org/10.1073/pnas.90.13.6320>
- Jacqué, J.M., Mann, A., Enslin, H., Sharova, N., Brichacek, B., Davis, R.J., Stevenson, M., 1998. Modulation of HIV-1 infectivity by MAPK, a virion-associated kinase. *EMBO J.* 17, 2607–2618. <https://doi.org/10.1093/emboj/17.9.2607>
- Jouvenet, N., Bieniasz, P.D., Simon, S.M., 2008. Imaging the biogenesis of individual HIV-1 virions in live cells. *Nature* 454, 236–240. <https://doi.org/10.1038/nature06998>
- Jouvenet, N., Neil, S.J.D., Bess, C., Johnson, M.C., Virgen, C.A., Simon, S.M., Bieniasz, P.D., 2006. Plasma Membrane Is the Site of Productive HIV-1 Particle Assembly. *PLoS Biol.* 4, e435. <https://doi.org/10.1371/journal.pbio.0040435>
- Jouvenet, N., Simon, S.M., Bieniasz, P.D., 2009. Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc. Natl. Acad. Sci.* 106, 19114–19119. <https://doi.org/10.1073/pnas.0907364106>
- Kalland, K.H., Szilvay, A.M., Brokstad, K.A., Saetrevik, W., Haukenes, G., 1994. The human immunodeficiency virus type 1 Rev protein shuttles between the cytoplasm and nuclear compartments. *Mol. Cell. Biol.* 14, 7436–7444.
- Karn, J., Stoltzfus, C.M., 2012. Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression. *Cold Spring Harb. Perspect. Med.* 2, a006916. <https://doi.org/10.1101/cshperspect.a006916>
- Katoh, K., Shibata, H., Suzuki, H., Nara, A., Ishidoh, K., Kominami, E., Yoshimori, T., Maki, M., 2003. The ALG-2-interacting Protein Alix Associates with CHMP4b, a Human Homologue of Yeast Snf7 That Is Involved in Multivesicular Body Sorting\*. *J. Biol. Chem.* 278, 39104–39113. <https://doi.org/10.1074/jbc.M301604200>
- Kaye, J.F., Lever, A.M.L., 1998. Nonreciprocal Packaging of Human Immunodeficiency Virus Type 1 and Type 2 RNA: a Possible Role for the p2 Domain of Gag in RNA Encapsidation. *J. Virol.* 72, 5877–5885.
- Keane, S.C., Van, V., Frank, H.M., Sciandra, C.A., McCowin, S., Santos, J., Heng, X., Summers, M.F., 2016. NMR detection of intermolecular interaction sites in the dimeric 5'-leader of the HIV-1 genome. *Proc. Natl. Acad. Sci. U. S. A.* 113, 13033–13038. <https://doi.org/10.1073/pnas.1614785113>
- Keele, B.F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., Bibollet-Ruche, F., Chen, Y., Wain, L.V., Liegeois, F., Loul, S., Ngole, E.M., Bienvenue, Y., Delaporte, E., Brookfield, J.F.Y., Sharp, P.M., Shaw, G.M., Peeters, M., Hahn, B.H., 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* 313, 523–526. <https://doi.org/10.1126/science.1126531>
- Kenyon, J.C., Prestwood, L.J., Le Grice, S.F.J., Lever, A.M.L., 2013. In-gel probing of individual RNA conformers within a mixed population reveals a dimerization structural switch in the HIV-1 leader. *Nucleic Acids Res.* 41, e174. <https://doi.org/10.1093/nar/gkt690>
- Kharytonchyk, S., Monti, S., Smaldino, P.J., Van, V., Bolden, N.C., Brown, J.D., Russo, E., Swanson, C., Shuey, A., Telesnitsky, A., Summers, M.F., 2016. Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proc. Natl. Acad. Sci.* 113, 13378–13383. <https://doi.org/10.1073/pnas.1616627113>

- Killian, M.S., Fujimura, S.H., Hecht, F.M., Levy, J.A., 2006. Similar changes in plasmacytoid dendritic cell and CD4 T-cell counts during primary HIV-1 infection and treatment. *AIDS Lond. Engl.* 20, 1247–1252. <https://doi.org/10.1097/01.aids.0000232231.34253.bd>
- Kim, K., Dauphin, A., Komurlu, S., McCauley, S.M., Yurkovetskiy, L.A., Carbone, C., Diehl, W.E., Strambio-De-Castillia, C., Campbell, E.M., Luban, J., 2019. Cyclophilin A protects HIV-1 from restriction by human TRIM5 $\alpha$ . *Nat. Microbiol.* 4, 2044–2051. <https://doi.org/10.1038/s41564-019-0592-5>
- Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J.-C., Montagnier, L., 1984. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* 312, 767–768. <https://doi.org/10.1038/312767a0>
- Kleiman, L., 2002. tRNALys3: The Primer tRNA for Reverse Transcription in HIV-1. *IUBMB Life* 53, 107–114. <https://doi.org/10.1080/15216540211469>
- Klingler, J., Anton, H., Réal, E., Zeiger, M., Moog, C., Mély, Y., Boutant, E., 2020a. How HIV-1 Gag Manipulates Its Host Cell Proteins: A Focus on Interactors of the Nucleocapsid Domain. *Viruses* 12, E888. <https://doi.org/10.3390/v12080888>
- Kogan, M., Rappaport, J., 2011. HIV-1 accessory protein Vpr: relevance in the pathogenesis of HIV and potential for therapeutic intervention. *Retrovirology* 8, 25. <https://doi.org/10.1186/1742-4690-8-25>
- Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., Steitz, T.A., 1992. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 256, 1783–1790. <https://doi.org/10.1126/science.1377403>
- Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., Detours, V., 2001. Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.* 58, 19–42. <https://doi.org/10.1093/bmb/58.1.19>
- Kudoh, A., Takahama, S., Sawasaki, T., Ode, H., Yokoyama, M., Okayama, A., Ishikawa, A., Miyakawa, K., Matsunaga, S., Kimura, H., Sugiura, W., Sato, H., Hirano, H., Ohno, S., Yamamoto, N., Ryo, A., 2014. The phosphorylation of HIV-1 Gag by atypical protein kinase C facilitates viral infectivity by promoting Vpr incorporation into virions. *Retrovirology* 11, 9. <https://doi.org/10.1186/1742-4690-11-9>
- Kuritzkes, D.R., 2009. HIV-1 entry inhibitors: an overview. *Curr. Opin. HIV AIDS* 4, 82–87. <https://doi.org/10.1097/COH.0b013e328322402e>
- Kutluay, S.B., Bieniasz, P.D., 2010. Analysis of the Initiating Events in HIV-1 Particle Assembly and Genome Packaging. *PLOS Pathog.* 6, e1001200. <https://doi.org/10.1371/journal.ppat.1001200>
- Kutluay, S.B., Zang, T., Blanco-Melo, D., Powell, C., Jannain, D., Errando, M., Bieniasz, P.D., 2014. Global changes in the RNA binding specificity of HIV-1 Gag regulate virion genesis. *Cell* 159, 1096–1109. <https://doi.org/10.1016/j.cell.2014.09.057>
- Lahouassa, H., Daddacha, W., Hofmann, H., Ayinde, D., Logue, E.C., Dragin, L., Bloch, N., Maudet, C., Bertrand, M., Gramberg, T., Pancino, G., Priet, S., Canard, B., Laguette, N., Benkirane, M., Transy, C., Landau, N.R., Kim, B., Margottin-Goguet, F., 2012. SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat. Immunol.* 13, 223–228. <https://doi.org/10.1038/ni.2236>
- Le Rouzic, E., Benichou, S., 2005. The Vpr protein from HIV-1: distinct roles along the viral life cycle. *Retrovirology* 2, 11. <https://doi.org/10.1186/1742-4690-2-11>
- Lee, S.-K., Potempa, M., Kolli, M., Özen, A., Schiffer, C.A., Swanstrom, R., 2012. Context surrounding processing sites is crucial in determining cleavage rate of a subset of processing sites in HIV-1 Gag and Gag-Pro-Pol polyprotein precursors by viral protease. *J. Biol. Chem.* 287, 13279–13290. <https://doi.org/10.1074/jbc.M112.339374>
- Lehmann, M., Milev, M.P., Abrahamyan, L., Yao, X.-J., Pante, N., Mouland, A.J., 2009. Intracellular Transport of Human Immunodeficiency Virus Type 1 Genomic RNA and Viral Production Are Dependent on Dynein Motor Function and Late Endosome Positioning. *J. Biol. Chem.* 284, 14572–14585. <https://doi.org/10.1074/jbc.M808531200>
- Lever, A., Gottlinger, H., Haseltine, W., Sodroski, J., 1989. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J. Virol.* 63, 4085–4087. <https://doi.org/10.1128/JVI.63.9.4085-4087.1989>
- Levitz, T.S., Weckener, M., Fong, I., Naismith, J.H., Drennan, C.L., Brignole, E.J., Clare, D.K., Darrow, M.C., 2022. Approaches to Using the Chameleon: Robust, Automated, Fast-Plunge cryoEM Specimen Preparation. *Front. Mol. Biosci.* 9.
- Levy, J.A., 1993. Pathogenesis of human immunodeficiency virus infection. *Microbiol. Rev.* 57, 183–289. <https://doi.org/10.1128/mr.57.1.183-289.1993>
- Levy, N., Eiler, S., Pradeau-Aubreton, K., Maillot, B., Stricher, F., Ruff, M., 2016. Production of unstable proteins through the formation of stable core complexes. *Nat. Commun.* 7, 10932. <https://doi.org/10.1038/ncomms10932>
- Li, Z., Li, W., Lu, M., Bess, J., Chao, C.W., Gorman, J., Terry, D.S., Zhang, B., Zhou, T., Blanchard, S.C., Kwong, P.D., Lifson, J.D., Mothes, W., Liu, J., 2020. Subnanometer structures of HIV-1 envelope trimers on aldrithiol-2-inactivated virus particles. *Nat. Struct. Mol. Biol.* 27, 726–734. <https://doi.org/10.1038/s41594-020-0452-2>

- Lin, K., Yan, Q., Mitchell, A., Funk, N., Lu, C., Xiao, H., 2020. A simple method for non-denaturing purification of biotin-tagged proteins through competitive elution with free biotin. *BioTechniques* 68, 41–44. <https://doi.org/10.2144/btn-2019-0088>
- Lingappa, J.R., Dooher, J.E., Newman, M.A., Kiser, P.K., Klein, K.C., 2006. Basic Residues in the Nucleocapsid Domain of Gag Are Required for Interaction of HIV-1 Gag with ABCE1 (HP68), a Cellular Protein Important for HIV-1 Capsid Assembly\*. *J. Biol. Chem.* 281, 3773–3784. <https://doi.org/10.1074/jbc.M507255200>
- Lorber, B., 2018. Analytical light scattering methods in molecular and structural biology: Experimental aspects and results. <https://doi.org/10.48550/arXiv.1810.00611>
- Lu, K., Heng, X., Garyu, L., Monti, S., Garcia, E.L., Kharytonchyk, S., Dorjsuren, B., Kulandaivel, G., Jones, S., Hiremath, A., Divakaruni, S.S., LaCotti, C., Barton, S., Tummlillo, D., Hosis, A., Edme, K., Albrecht, S., Telesnitsky, A., Summers, M.F., 2011. NMR Detection of Structures in the HIV-1 5'-Leader RNA that Regulate Genome Packaging. *Science* 334, 242–245. <https://doi.org/10.1126/science.1210460>
- Luban, J., Goff, S.P., 1994. Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. *J. Virol.* 68, 3784–3793. <https://doi.org/10.1128/JVI.68.6.3784-3793.1994>
- Luban, J., Goff, S.P., 1991. Binding of human immunodeficiency virus type 1 (HIV-1) RNA to recombinant HIV-1 gag polyprotein. *J. Virol.* 65, 3203–3212.
- Lubow, J., Collins, K.L., 2020. Vpr Is a VIP: HIV Vpr and Infected Macrophages Promote Viral Pathogenesis. *Viruses* 12, 809. <https://doi.org/10.3390/v12080809>
- Lusic, M., Siliciano, R.F., 2017. Nuclear landscape of HIV-1 infection and integration. *Nat. Rev. Microbiol.* 15, 69–82. <https://doi.org/10.1038/nrmicro.2016.162>
- Maertens, G.N., Engelman, A.N., Cherepanov, P., 2022. Structure and function of retroviral integrase. *Nat. Rev. Microbiol.* 20, 20–34. <https://doi.org/10.1038/s41579-021-00586-9>
- Mailler, E., Bernacchi, S., Marquet, R., Paillart, J.-C., Vivet-Boudou, V., Smyth, R.P., 2016. The Life-Cycle of the HIV-1 Gag–RNA Complex. *Viruses* 8, 248. <https://doi.org/10.3390/v8090248>
- Malim, M.H., Bieniasz, P.D., 2012. HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harb. Perspect. Med.* 2, a006940. <https://doi.org/10.1101/cshperspect.a006940>
- Malim, M.H., Böhnlein, S., Hauber, J., Cullen, B.R., 1989. Functional dissection of the HIV-1 Rev trans-activator--derivation of a trans-dominant repressor of Rev function. *Cell* 58, 205–214. [https://doi.org/10.1016/0092-8674\(89\)90416-9](https://doi.org/10.1016/0092-8674(89)90416-9)
- Marquet, R., Guerrero, S., Bernacchi, S., Pernot, S., Batisse, J., Paillart, J.-C., 2012. Un facteur de transcription se fait complice du VIH-1 pour détruire les défenses cellulaires. *médecine/sciences* 28, 356–358. <https://doi.org/10.1051/medsci/2012284007>
- Martin-Serrano, J., Perez-Caballero, D., Bieniasz, P.D., 2004. Context-Dependent Effects of L Domains and Ubiquitination on Viral Budding. *J. Virol.* 78, 5554–5563. <https://doi.org/10.1128/JVI.78.11.5554-5563.2004>
- Martin-Serrano, J., Yaravoy, A., Perez-Caballero, D., Bieniasz, P.D., 2003. Divergent retroviral late-budding domains recruit vacuolar protein sorting factors by using alternative adaptor proteins. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12414–12419. <https://doi.org/10.1073/pnas.2133846100>
- Martin-Serrano, J., Zang, T., Bieniasz, P.D., 2001. HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. *Nat. Med.* 7, 1313–1319. <https://doi.org/10.1038/nm1201-1313>
- Matreyek, K.A., Wang, W., Serrao, E., Singh, P.K., Levin, H.L., Engelman, A., 2014. Host and viral determinants for MxB restriction of HIV-1 infection. *Retrovirology* 11, 90. <https://doi.org/10.1186/s12977-014-0090-z>
- Mattei, S., Flemming, A., Anders-Össwein, M., Kräusslich, H.-G., Briggs, J.A.G., Müller, B., 2015. RNA and Nucleocapsid Are Dispensable for Mature HIV-1 Capsid Assembly. *J. Virol.* 89, 9739–9747. <https://doi.org/10.1128/JVI.00750-15>
- McDowell, G.S., Philpott, A., 2013. Non-canonical ubiquitylation: Mechanisms and consequences. *Int. J. Biochem. Cell Biol.* 45, 1833–1842. <https://doi.org/10.1016/j.biocel.2013.05.026>
- McKinstry, W.J., Hijnen, M., Tanwar, H.S., Sparrow, L.G., Nagarajan, S., Pham, S.T., Mak, J., 2014. Expression and purification of soluble recombinant full length HIV-1 Pr55(Gag) protein in *Escherichia coli*. *Protein Expr. Purif.* 100, 10–18. <https://doi.org/10.1016/j.pep.2014.04.013>
- Meissner, M.E., Talledge, N., Mansky, L.M., 2022. Molecular Biology and Diversification of Human Retroviruses. *Front. Virol.* 2.
- Mendonça, L., Sun, D., Ning, J., Liu, J., Kotecha, A., Olek, M., Frosio, T., Fu, X., Himes, B.A., Kleinpeter, A.B., Freed, E.O., Zhou, J., Aiken, C., Zhang, P., 2021. CryoET structures of immature HIV Gag reveal six-helix bundle. *Commun. Biol.* 4, 1–9. <https://doi.org/10.1038/s42003-021-01999-1>
- Menéndez-Arias, L., 2008. Mechanisms of resistance to nucleoside analogue inhibitors of HIV-1 reverse transcriptase. *Virus Res.* 134, 124–146. <https://doi.org/10.1016/j.virusres.2007.12.015>

- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682. <https://doi.org/10.1038/s41592-022-01488-1>
- Misono, T., 2019. Dynamic Light Scattering (DLS), in: Abe, M. (Ed.), *Measurement Techniques and Practices of Colloid and Interface Phenomena*. Springer, Singapore, pp. 65–69. [https://doi.org/10.1007/978-981-13-5931-6\\_10](https://doi.org/10.1007/978-981-13-5931-6_10)
- Mitra, S., 2009. Using analytical ultracentrifugation (AUC) to measure global conformational changes accompanying equilibrium tertiary folding of RNA molecules. *Methods Enzymol.* 469, 209–236. [https://doi.org/10.1016/S0076-6879\(09\)69010-8](https://doi.org/10.1016/S0076-6879(09)69010-8)
- Moore, M.D., Nikolaitchik, O.A., Chen, J., Hammarskjöld, M.-L., Rekosh, D., Hu, W.-S., 2009. Probing the HIV-1 Genomic RNA Trafficking Pathway and Dimerization by Genetic Recombination and Single Virion Analyses. *PLoS Pathog.* 5, e1000627. <https://doi.org/10.1371/journal.ppat.1000627>
- Morellet, N., Bouaziz, S., Petitjean, P., Roques, B.P., 2003. NMR Structure of the HIV-1 Regulatory Protein VPR. *J. Mol. Biol.* 327, 215–227. [https://doi.org/10.1016/S0022-2836\(03\)00060-3](https://doi.org/10.1016/S0022-2836(03)00060-3)
- Morellet, N., Roques, B.P., Bouaziz, S., 2009. Structure-Function Relationship of Vpr: Biological Implications. *Curr. HIV Res.* 7, 184–210.
- Mouhand, A., Belfetmi, A., Catala, M., Larue, V., Zargarian, L., Brachet, F., Gorelick, R.J., Van Heijenoort, C., Mirambeau, G., Barraud, P., Mauffret, O., Tisné, C., 2018. Modulation of the HIV nucleocapsid dynamics finely tunes its RNA-binding properties during virion genesis. *Nucleic Acids Res.* 46, 9699–9710. <https://doi.org/10.1093/nar/gky612>
- Mouhand, A., Pasi, M., Catala, M., Zargarian, L., Belfetmi, A., Barraud, P., Mauffret, O., Tisné, C., 2020. Overview of the Nucleic-Acid Binding Properties of the HIV-1 Nucleocapsid Protein in Its Different Maturation States. *Viruses* 12, 1109. <https://doi.org/10.3390/v12101109>
- Mouland, A.J., Mercier, J., Luo, M., Bernier, L., DesGroseillers, L., Cohen, E.A., 2000. The double-stranded RNA-binding protein Staufen is incorporated in human immunodeficiency virus type 1: evidence for a role in genomic RNA encapsidation. *J. Virol.* 74, 5441–5451. <https://doi.org/10.1128/jvi.74.12.5441-5451.2000>
- Mourez, T., Simon, F., Plantier, J.-C., 2013. Non-M variants of human immunodeficiency virus type 1. *Clin. Microbiol. Rev.* 26, 448–461. <https://doi.org/10.1128/CMR.00012-13>
- Müller, B., Tessmer, U., Schubert, U., Kräusslich, H.G., 2000. Human immunodeficiency virus type 1 Vpr protein is incorporated into the virion in significantly smaller amounts than gag and is phosphorylated in infected cells. *J. Virol.* 74, 9727–9731. <https://doi.org/10.1128/jvi.74.20.9727-9731.2000>
- Münch, J., Rajan, D., Schindler, M., Specht, A., Rücker, E., Novembre, F.J., Nerrienet, E., Müller-Trutwin, M.C., Peeters, M., Hahn, B.H., Kirchhoff, F., 2007. Nef-Mediated Enhancement of Virion Infectivity and Stimulation of Viral Replication Are Fundamental Properties of Primate Lentiviruses. *J. Virol.* 81, 13852–13864. <https://doi.org/10.1128/JVI.00904-07>
- Munshi, U.M., Kim, J., Nagashima, K., Hurley, J.H., Freed, E.O., 2007. An Alix Fragment Potently Inhibits HIV-1 Budding: Characterization Of Binding To Retroviral YPX Late Domains. *J. Biol. Chem.* 282, 3847–3855. <https://doi.org/10.1074/jbc.M607489200>
- Neil, S., Bieniasz, P., 2009. Human Immunodeficiency Virus, Restriction Factors, and Interferon. *J. Interferon Cytokine Res.* 29, 569–580. <https://doi.org/10.1089/jir.2009.0077>
- Ocwieja, K.E., Sherrill-Mix, S., Mukherjee, R., Custers-Allen, R., David, P., Brown, M., Wang, S., Link, D.R., Olson, J., Travers, K., Schadt, E., Bushman, F.D., 2012. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res.* 40, 10345–10355. <https://doi.org/10.1093/nar/gks753>
- Olety, B., Ono, A., 2014. Roles played by acidic lipids in HIV-1 Gag membrane binding. *Virus Res.* 193, 108–115. <https://doi.org/10.1016/j.virusres.2014.06.015>
- Olson, E.D., Musier-Forsyth, K., 2019. Retroviral Gag protein–RNA interactions: Implications for specific genomic RNA packaging and virion assembly. *Semin. Cell Dev. Biol., SI: Human dendritic cells* 86, 129–139. <https://doi.org/10.1016/j.semcdb.2018.03.015>
- Ott, D.E., Coren, L.V., Shatzer, T., 2009. The Nucleocapsid Region of Human Immunodeficiency Virus Type 1 Gag Assists in the Coordination of Assembly and Gag Processing: Role for RNA-Gag Binding in the Early Stages of Assembly. *J. Virol.* 83, 7718–7727. <https://doi.org/10.1128/JVI.00099-09>
- Paillart, J.-C., Dettenhofer, M., Yu, X.-F., Ehresmann, C., Ehresmann, B., Marquet, R., 2004. First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J. Biol. Chem.* 279, 48397–48403. <https://doi.org/10.1074/jbc.M408294200>
- Paillart, J.-C., Göttlinger, H.G., 1999. Opposing Effects of Human Immunodeficiency Virus Type 1 Matrix Mutations Support a Myristyl Switch Model of Gag Membrane Targeting. *J. Virol.* 73, 2604–2612.
- Paillart, J.-C., Skripkin, E., Ehresmann, B., Ehresmann, C., Marquet, R., 2002. In Vitro Evidence for a Long Range Pseudoknot in the 5′-Untranslated and Matrix Coding Regions of HIV-1 Genomic RNA. *J. Biol. Chem.* 5995–6004.

- Paillart, J.C., Skripkin, E., Ehresmann, B., Ehresmann, C., Marquet, R., 1996. A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci. U. S. A.* 93, 5572–5577.
- Peeters, M., Gueye, A., Mboup, S., Bibollet-Ruche, F., Ekaza, E., Mulanga, C., Ouedrago, R., Gandji, R., Mpele, P., Dibanga, G., Koumare, B., Saidou, M., Esu-Williams, E., Lombart, J.P., Badombena, W., Luo, N., Vanden Haesevelde, M., Delaporte, E., 1997. Geographical distribution of HIV-1 group O viruses in Africa. *AIDS Lond. Engl.* 11, 493–498. <https://doi.org/10.1097/00002030-199704000-00013>
- Peeters, M., Liegeois, F., Torimiro, N., Bourgeois, A., Mpoudi, E., Vergne, L., Saman, E., Delaporte, E., Saragosti, S., 1999. Characterization of a Highly Replicative Intergroup M/O Human Immunodeficiency Virus Type 1 Recombinant Isolated from a Cameroonian Patient. *J. Virol.* 73, 7368–7375.
- Perlman, M., Resh, M.D., 2006. Identification of an intracellular trafficking and assembly pathway for HIV-1 gag. *Traffic Cph. Den.* 7, 731–745. <https://doi.org/10.1111/j.1398-9219.2006.00428.x>
- Pettit, S.C., Lindquist, J.N., Kaplan, A.H., Swanstrom, R., 2005. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology* 2, 66. <https://doi.org/10.1186/1742-4690-2-66>
- Phanuphak, N., Gulick, R.M., 2020. HIV treatment and prevention 2019: current standards of care. *Curr. Opin. HIV AIDS* 15, 4–12. <https://doi.org/10.1097/>
- Pires, R., Hartlieb, B., Signor, L., Schoehn, G., Lata, S., Roessle, M., Moriscot, C., Popov, S., Hinz, A., Jamin, M., Boyer, V., Sadoul, R., Forest, E., Svergun, D.I., Göttlinger, H.G., Weissenhorn, W., 2009. A crescent shaped ALIX dimer targets ESCRT-III CHMP4 filaments. *Struct. Lond. Engl.* 1993 17, 843–856. <https://doi.org/10.1016/j.str.2009.04.007>
- Plantier, J.-C., Leoz, M., Dickerson, J.E., De Oliveira, F., Cordonnier, F., Lemée, V., Damond, F., Robertson, D.L., Simon, F., 2009. A new human immunodeficiency virus derived from gorillas. *Nat. Med.* 15, 871–872. <https://doi.org/10.1038/nm.2016>
- Poiesz, B.J., Russetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D., Gallo, R.C., 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* 77, 7415–7419. <https://doi.org/10.1073/pnas.77.12.7415>
- Poljak, L., Batson, S.M., Ficheux, D., Roques, B.P., Darlix, J.-L., Käs, E., 2003. Analysis of NCp7-dependent activation of HIV-1 cDNA integration and its conservation among retroviral nucleocapsid proteins. *J. Mol. Biol.* 329, 411–421. [https://doi.org/10.1016/s0022-2836\(03\)00472-8](https://doi.org/10.1016/s0022-2836(03)00472-8)
- Pornillos, O., Alam, S.L., Davis, D.R., Sundquist, W.I., 2002. Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein. *Nat. Struct. Biol.* 9, 812–817. <https://doi.org/10.1038/nsb856>
- Pornillos, O., Ganser-Pornillos, B.K., 2019. Maturation of retroviruses. *Curr. Opin. Virol.* 36, 47–55. <https://doi.org/10.1016/j.coviro.2019.05.004>
- Pornillos, O., Ganser-Pornillos, B.K., Yeager, M., 2011. Atomic level modeling of the HIV capsid. *Nature* 469, 424–427. <https://doi.org/10.1038/nature09640>
- Poulsen, N.N., Andersen, N.Z., Østergaard, J., Zhuang, G., Petersen, N.J., Jensen, H., 2015. Flow induced dispersion analysis rapidly quantifies proteins in human plasma samples. *Analyst* 140, 4365–4369. <https://doi.org/10.1039/C5AN00697J>
- Radestock, B., Morales, I., Rahman, S.A., Radau, S., Glass, B., Zahedi, R.P., Müller, B., Kräusslich, H.-G., 2013. Comprehensive Mutational Analysis Reveals p6Gag Phosphorylation To Be Dispensable for HIV-1 Morphogenesis and Replication. *J. Virol.* 87, 724–734. <https://doi.org/10.1128/JVI.02162-12>
- Raja, A., Venturi, M., Kwong, P., Sodroski, J., 2003. CD4 binding site antibodies inhibit human immunodeficiency virus gp120 envelope glycoprotein interaction with CCR5. *J. Virol.* 77, 713–718. <https://doi.org/10.1128/jvi.77.1.713-718.2003>
- Reed, J.C., Molter, B., Geary, C.D., McNevin, J., McElrath, J., Giri, S., Klein, K.C., Lingappa, J.R., 2012. HIV-1 Gag co-opts a cellular complex containing DDX6, a helicase that facilitates capsid assembly. *J. Cell Biol.* 198, 439–456. <https://doi.org/10.1083/jcb.201111012>
- Rein, A., Datta, S.A.K., Jones, C.P., Musier-Forsyth, K., 2011. Diverse interactions of retroviral Gag proteins with RNAs. *Trends Biochem. Sci.* 36, 373–380. <https://doi.org/10.1016/j.tibs.2011.04.001>
- Rein, A., Henderson, L.E., Levin, J.G., 1998. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem. Sci.* 23, 297–301. [https://doi.org/10.1016/S0968-0004\(98\)01256-0](https://doi.org/10.1016/S0968-0004(98)01256-0)
- Rensen, E., Mueller, F., Scoca, V., Parmar, J.J., Souque, P., Zimmer, C., Di Nunzio, F., 2021. Clustering and reverse transcription of HIV-1 genomes in nuclear niches of macrophages. *EMBO J.* 40, e105247. <https://doi.org/10.15252/embj.2020105247>
- Resnick, R., Omer, C.A., Faras, A.J., 1984. Involvement of retrovirus reverse transcriptase-associated RNase H in the initiation of strong-stop (+) DNA synthesis and the generation of the long terminal repeat. *J. Virol.* 51, 813–821.

- Retureau, R., Oguey, C., Mauffret, O., Hartmann, B., 2019. Structural Explorations of NCp7–Nucleic Acid Complexes Give Keys to Decipher the Binding Process. *J. Mol. Biol.* 431, 1966–1980. <https://doi.org/10.1016/j.jmb.2019.03.002>
- Rhim, H., Park, J., Morrow, C.D., 1991. Deletions in the tRNA(Lys) primer-binding site of human immunodeficiency virus type 1 identify essential regions for reverse transcription. *J. Virol.* 65, 4555–4564. <https://doi.org/10.1128/JVI.65.9.4555-4564.1991>
- Robertson, D.L., Anderson, J.P., Bradac, J.A., Carr, J.K., Foley, B., Funkhouser, R.K., Gao, F., Hahn, B.H., Kalish, M.L., Kuiken, C., Learn, G.H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P.M., Wolinsky, S., Korber, B., 2000. HIV-1 nomenclature proposal. *Science* 288, 55–56.
- Rodi, P.M., Bocco Gianello, M.D., Gennaro, A.M., 2014. Insights About CHAPS Aggregation Obtained by Spin Label EPR Spectroscopy. *Appl. Magn. Reson.* 45, 1319–1332. <https://doi.org/10.1007/s00723-014-0604-9>
- Roepstorff, P., Fohlman, J., 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* 11, 601. <https://doi.org/10.1002/bms.1200111109>
- Roques, P., Robertson, D.L., Souquière, S., Apetrei, C., Nerrienet, E., Barré-Sinoussi, F., Müller-Trutwin, M., Simon, F., 2004. Phylogenetic characteristics of three new HIV-1 N strains and implications for the origin of group N. *AIDS Lond. Engl.* 18, 1371–1381. <https://doi.org/10.1097/01.aids.0000125990.86904.28>
- Rose, K.M., Hirsch, V.M., Bouamr, F., 2020. Budding of a Retrovirus: Some Assemblies Required. *Viruses* 12, 1188. <https://doi.org/10.3390/v12101188>
- Rous, P., 1911. A sarcoma of the fowl transmissible by an agent separable from the tumor cells. *J. Exp. Med.* 13, 397–411. <https://doi.org/10.1084/jem.13.4.397>
- Rousseau, C.M., Learn, G.H., Bhattacharya, T., Nickle, D.C., Heckerman, D., Chetty, S., Brander, C., Goulder, P.J.R., Walker, B.D., Kiepiela, P., Korber, B.T., Mullins, J.I., 2007. Extensive Intrasubtype Recombination in South African Human Immunodeficiency Virus Type 1 Subtype C Infections. *J. Virol.* 81, 4492–4500. <https://doi.org/10.1128/JVI.02050-06>
- Rowland-Jones, S.L., Whittle, H.C., 2007. Out of Africa: what can we learn from HIV-2 about protective immunity to HIV-1? *Nat. Immunol.* 8, 329–331. <https://doi.org/10.1038/ni0407-329>
- Roy, B.B., Russell, R.S., Turner, D., Liang, C., 2006. The T12I mutation within the SP1 region of Gag restricts packaging of spliced viral RNA into human immunodeficiency virus type 1 with mutated RNA packaging signals and mutated nucleocapsid sequence. *Virology* 344, 304–314. <https://doi.org/10.1016/j.virol.2005.09.011>
- Saad, J.S., Loeliger, E., Luncsford, P., Liriano, M., Tai, J., Kim, A., Miller, J., Joshi, A., Freed, E.O., Summers, M.F., 2007. Point mutations in the HIV-1 matrix protein turn off the myristyl switch. *J. Mol. Biol.* 366, 574–585. <https://doi.org/10.1016/j.jmb.2006.11.068>
- Saad, J.S., Miller, J., Tai, J., Kim, A., Ghanam, R.H., Summers, M.F., 2006. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. *Proc. Natl. Acad. Sci. U. S. A.* 103, 11364–11369. <https://doi.org/10.1073/pnas.0602818103>
- Saag, M.S., Gandhi, R.T., Hoy, J.F., Landovitz, R.J., Thompson, M.A., Sax, P.E., Smith, D.M., Benson, C.A., Buchbinder, S.P., Del Rio, C., Eron, J.J., Fätkenheuer, G., Günthard, H.F., Molina, J.-M., Jacobsen, D.M., Volberding, P.A., 2020. Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults: 2020 Recommendations of the International Antiviral Society-USA Panel. *JAMA* 324, 1651–1669. <https://doi.org/10.1001/jama.2020.17025>
- Sabbah, E.N., Druillennec, S., Morellet, N., Bouaziz, S., Kroemer, G., Roques, B. p., 2006. Interaction between the HIV-1 Protein Vpr and the Adenine Nucleotide Translocator. *Chem. Biol. Drug Des.* 67, 145–154. <https://doi.org/10.1111/j.1747-0285.2006.00340.x>
- Salgado, G.F., Marquant, R., Vogel, A., Alves, I.D., Feller, S.E., Morellet, N., Bouaziz, S., 2009a. Structural studies of HIV-1 Gag p6ct and its interaction with Vpr determined by solution nuclear magnetic resonance. *Biochemistry* 48, 2355–2367. <https://doi.org/10.1021/bi801794v>
- Salgado, G.F., Vogel, A., Marquant, R., Feller, S.E., Bouaziz, S., Alves, I.D., 2009b. The Role of Membranes in the Organization of HIV-1 Gag p6 and Vpr: p6 Shows High Affinity for Membrane Bilayers Which Substantially Increases the Interaction between p6 and Vpr. *J. Med. Chem.* 52, 7157–7162. <https://doi.org/10.1021/jm901106t>
- Saphire, A.C., Bobardt, M.D., Zhang, Z., David, G., Gallay, P.A., 2001. Syndecans serve as attachment receptors for human immunodeficiency virus type 1 on macrophages. *J. Virol.* 75, 9187–9200. <https://doi.org/10.1128/JVI.75.19.9187-9200.2001>
- Scatchard, G., 1949. The Attractions of Proteins for Small Molecules and Ions. *Ann. N. Y. Acad. Sci.* 51, 660–672. <https://doi.org/10.1111/j.1749-6632.1949.tb27297.x>

- Schindler, M., Würfl, S., Benaroch, P., Greenough, T.C., Daniels, R., Easterbrook, P., Brenner, M., Münch, J., Kirchhoff, F., 2003. Down-Modulation of Mature Major Histocompatibility Complex Class II and Up-Regulation of Invariant Chain Cell Surface Expression Are Well-Conserved Functions of Human and Simian Immunodeficiency Virus nef Alleles. *J. Virol.* 77, 10548–10556. <https://doi.org/10.1128/JVI.77.19.10548-10556.2003>
- Schmidt, M., Dikic, I., Bögl, O., 2005. Src phosphorylation of Alix/AIP1 modulates its interaction with binding partners and antagonizes its activities. *J. Biol. Chem.* 280, 3414–3425. <https://doi.org/10.1074/jbc.M409839200>
- Seissler, T., Marquet, R., Paillart, J.-C., 2017. Hijacking of the Ubiquitin/Proteasome Pathway by the HIV Auxiliary Proteins. *Viruses* 9, E322. <https://doi.org/10.3390/v9110322>
- Selyutina, A., Persaud, M., Lee, K., KewalRamani, V., Diaz-Griffero, F., 2020. Nuclear Import of the HIV-1 Core Precedes Reverse Transcription and Uncoating. *Cell Rep.* 32, 108201. <https://doi.org/10.1016/j.celrep.2020.108201>
- Serrao, E., Thys, W., Demeulemeester, J., Al-Mawsawi, L.Q., Christ, F., Debyser, Z., Neamati, N., 2012. A symmetric region of the HIV-1 integrase dimerization interface is essential for viral replication. *PLoS One* 7, e45177. <https://doi.org/10.1371/journal.pone.0045177>
- Sette, P., Jadwin, J.A., Dussupt, V., Bello, N.F., Bouamr, F., 2010. The ESCRT-associated protein Alix recruits the ubiquitin ligase Nedd4-1 to facilitate HIV-1 release through the LYPXnL L domain motif. *J. Virol.* 84, 8181–8192. <https://doi.org/10.1128/JVI.00634-10>
- Sette, P., Nagashima, K., Piper, R.C., Bouamr, F., 2013. Ubiquitin conjugation to Gag is essential for ESCRT-mediated HIV-1 budding. *Retrovirology* 10, 79. <https://doi.org/10.1186/1742-4690-10-79>
- Shan, L., Deng, K., Gao, H., Xing, S., Capoferri, A.A., Durand, C.M., Rabi, S.A., Laird, G.M., Kim, M., Hosmane, N.N., Yang, H.-C., Zhang, H., Margolick, J.B., Li, L., Cai, W., Ke, R., Flavell, R.A., Siliciano, J.D., Siliciano, R.F., 2017. Transcriptional Reprogramming during Effector-to-Memory Transition Renders CD4+ T Cells Permissive for Latent HIV-1 Infection. *Immunity* 47, 766–775.e3. <https://doi.org/10.1016/j.immuni.2017.09.014>
- Sharp, P.M., Hahn, B.H., 2011. Origins of HIV and the AIDS pandemic. *Cold Spring Harb. Perspect. Med.* 1, a006841. <https://doi.org/10.1101/cshperspect.a006841>
- Shaw, G.M., Hunter, E., 2012. HIV Transmission. *Cold Spring Harb. Perspect. Med.* 2, a006965. <https://doi.org/10.1101/cshperspect.a006965>
- Simon, F., Maucière, P., Roques, P., Loussert-Ajaka, I., Müller-Trutwin, M.C., Saragosti, S., Georges-Courbot, M.C., Barré-Sinoussi, F., Brun-Vézinet, F., 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* 4, 1032–1037. <https://doi.org/10.1038/2017>
- Simon, V., Ho, D.D., Karim, Q.A., 2006. HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. *Lancet* 368, 489–504. [https://doi.org/10.1016/S0140-6736\(06\)69157-5](https://doi.org/10.1016/S0140-6736(06)69157-5)
- Singh, S.P., Lai, D., Cartas, M., Serio, D., Murali, R., Kalyanaraman, V.S., Srinivasan, A., 2000. Epitope-Tagging Approach to Determine the Stoichiometry of the Structural and Nonstructural Proteins in the Virus Particles: Amount of Vpr in Relation to Gag in HIV-1. *Virology* 268, 364–371. <https://doi.org/10.1006/viro.2000.0191>
- Skrupnik, E., Paillart, J.C., Marquet, R., Ehresmann, B., Ehresmann, C., 1994. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 91, 4945–4949.
- Smith, R.D., Loo, J.A., Edmonds, C.G., Barinaga, C.J., Udseth, H.R., 1990. New developments in biochemical mass spectrometry: electrospray ionization. *Anal. Chem.* 62, 882–899. <https://doi.org/10.1021/ac00208a002>
- Smyth, R.P., Despons, L., Huili, G., Bernacchi, S., Hijnen, M., Mak, J., Jossinet, F., Weixi, L., Paillart, J.-C., von Kleist, M., Marquet, R., 2015. Mutational interference mapping experiment (MIME) for studying RNA structure and function. *Nat. Methods* 12, 866–872. <https://doi.org/10.1038/nmeth.3490>
- Song, R., Kafaie, J., Laughrea, M., 2008. Role of the 5' TAR stem-loop and the U5-AUG duplex in dimerization of HIV-1 genomic RNA. *Biochemistry* 47, 3283–3293. <https://doi.org/10.1021/bi7023173>
- Spinner, C.D., Boesecke, C., Zink, A., Jessen, H., Stellbrink, H.-J., Rockstroh, J.K., Esser, S., 2016. HIV pre-exposure prophylaxis (PrEP): a review of current knowledge of oral systemic HIV PrEP in humans. *Infection* 44, 151–158. <https://doi.org/10.1007/s15010-015-0850-2>
- Starcich, B.R., Hahn, B.H., Shaw, G.M., McNeely, P.D., Modrow, S., Wolf, H., Parks, E.S., Parks, W.P., Josephs, S.F., Gallo, R.C., 1986. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* 45, 637–648. [https://doi.org/10.1016/0092-8674\(86\)90778-6](https://doi.org/10.1016/0092-8674(86)90778-6)
- Strack, B., Calistri, A., Accola, M.A., Palù, G., Göttlinger, H.G., 2000. A role for ubiquitin ligase recruitment in retrovirus release. *Proc. Natl. Acad. Sci. U. S. A.* 97, 13063–13068.

- Strack, B., Calistri, A., Craig, S., Popova, E., Göttlinger, H.G., 2003. AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell* 114, 689–699. [https://doi.org/10.1016/s0092-8674\(03\)00653-6](https://doi.org/10.1016/s0092-8674(03)00653-6)
- Stupfler, B., Verriez, C., Gallois-Montbrun, S., Marquet, R., Paillart, J.-C., 2021. Degradation-Independent Inhibition of APOBEC3G by the HIV-1 Vif Protein. *Viruses* 13, 617. <https://doi.org/10.3390/v13040617>
- Sundquist, W.I., Kräusslich, H.-G., 2012. HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harb. Perspect. Med.* 2, 1–24. <https://doi.org/10.1101/cshperspect.a006924>
- Tada, T., Zhang, Y., Koyama, T., Tobiume, M., Tsunetsugu-Yokota, Y., Yamaoka, S., Fujita, H., Tokunaga, K., 2015. MARCH8 inhibits HIV-1 infection by reducing virion incorporation of envelope glycoproteins. *Nat. Med.* 21, 1502–1507. <https://doi.org/10.1038/nm.3956>
- Tanwar, H.S., Khoo, K.K., Garvey, M., Waddington, L., Leis, A., Hijnen, M., Velkov, T., Dumsday, G.J., McKinstry, W.J., Mak, J., 2017. The thermodynamics of Pr55Gag-RNA interaction regulate the assembly of HIV. *PLOS Pathog.* 13, e1006221. <https://doi.org/10.1371/journal.ppat.1006221>
- Tchounga, B., Ekouevi, D.K., Balestre, E., Dabis, F., 2016. Mortality and survival patterns of people living with HIV-2. *Curr. Opin. HIV AIDS* 11, 537–544. <https://doi.org/10.1097/COH.0000000000000299>
- Tedbury, P.R., Mercredi, P.Y., Gaines, C.R., Summers, M.F., Freed, E.O., 2015. Elucidating the mechanism by which compensatory mutations rescue an HIV-1 matrix mutant defective for gag membrane targeting and envelope glycoprotein incorporation. *J. Mol. Biol.* 427, 1413–1427. <https://doi.org/10.1016/j.jmb.2015.01.018>
- Temin, H.M., Mizutani, S., 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 226, 1211–1213. <https://doi.org/10.1038/2261211a0>
- Thornhill, D., Olety, B., Ono, A., 2019. Relationships between MA-RNA Binding in Cells and Suppression of HIV-1 Gag Mislocalization to Intracellular Membranes. *J. Virol.* 93, 23. <https://doi.org/10.1128/JVI.00756-19>
- Tisné, C., Guillièrre, F., Dardel, F., 2005. NMR-based identification of peptides that specifically recognize the d-arm of tRNA. *Biochimie* 87, 885–888. <https://doi.org/10.1016/j.biochi.2005.02.003>
- Trautz, B., Pierini, V., Wombacher, R., Stolp, B., Chase, A.J., Pizzato, M., Fackler, O.T., 2016. The Antagonism of HIV-1 Nef to SERINC5 Particle Infectivity Restriction Involves the Counteraction of Virion-Associated Pools of the Restriction Factor. *J. Virol.* 90, 10915–10927. <https://doi.org/10.1128/JVI.01246-16>
- Vallari, A., Holzmayer, V., Harris, B., Yamaguchi, J., Ngansop, C., Makamche, F., Mbanya, D., Kaptué, L., Ndembu, N., Gürtler, L., Devare, S., Brennan, C.A., 2011. Confirmation of Putative HIV-1 Group P in Cameroon. *J. Virol.* 85, 1403–1407. <https://doi.org/10.1128/JVI.02005-10>
- Van Damme, N., Guatelli, J., 2008. HIV-1 Vpu inhibits accumulation of the envelope glycoprotein within clathrin-coated, Gag-containing endosomes. *Cell. Microbiol.* 10, 1040–1057. <https://doi.org/10.1111/j.1462-5822.2007.01101.x>
- Velours, C., Zhou, J., Zecchin, P., He, N., Salameh, M., Golinelli-Cohen, M.-P., Golinelli-Pimpaneau, B., 2022. Determination of the Absolute Molar Mass of [Fe-S]-Containing Proteins Using Size Exclusion Chromatography-Multi-Angle Light Scattering (SEC-MALS). *Biomolecules* 12, 270. <https://doi.org/10.3390/biom12020270>
- Verriez, C., Marquet, R., Paillart, J.-C., Stupfler, B., 2020. APOBEC3s: history of an antiviral and mutagenic protein family. *Virologie* 24, 381–418. <https://doi.org/10.1684/vir.2020.0870>
- Visseaux, B., Damond, F., Matheron, S., Descamps, D., Charpentier, C., 2016. Hiv-2 molecular epidemiology. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* 46, 233–240. <https://doi.org/10.1016/j.meegid.2016.08.010>
- Votteler, J., Iavnilovitch, E., Fingrut, O., Shemesh, V., Taglicht, D., Erez, O., Sörgel, S., Walther, T., Bannert, N., Schubert, U., Reiss, Y., 2009. Exploring the functional interaction between POSH and ALIX and the relevance to HIV-1 release. *BMC Biochem.* 10, 12. <https://doi.org/10.1186/1471-2091-10-12>
- Wang, W., Naiyer, N., Mitra, M., Li, J., Williams, M.C., Rouzina, I., Gorelick, R.J., Wu, Z., Musier-Forsyth, K., 2014. Distinct nucleic acid interaction properties of HIV-1 nucleocapsid protein precursor NCp15 explain reduced viral infectivity. *Nucleic Acids Res.* 42, 7145–7159. <https://doi.org/10.1093/nar/gku335>
- Watanabe, S.M., Strickland, M., Tjandra, N., Carter, C.A., 2020. RNA Binding Suppresses Tsg101 Recognition of Ub-Modified Gag and Facilitates Recruitment to the Plasma Membrane. *Viruses* 12, 447. <https://doi.org/10.3390/v12040447>
- Webb, J.A., Jones, C.P., Parent, L.J., Rouzina, I., Musier-Forsyth, K., 2013. Distinct binding interactions of HIV-1 Gag to Psi and non-Psi RNAs: implications for viral genomic RNA packaging. *RNA N. Y. N.* 19, 1078–1088. <https://doi.org/10.1261/rna.038869.113>
- Welker, L., Paillart, J.-C., Bernacchi, S., 2021. Importance of Viral Late Domains in Budding and Release of Enveloped RNA Viruses. *Viruses* 13, 1559. <https://doi.org/10.3390/v13081559>

- Wen, Y., Feigenson, G.W., Vogt, V.M., Dick, R.A., 2020. Mechanisms of PI(4,5)P<sub>2</sub> Enrichment in HIV-1 Viral Membranes. *J. Mol. Biol.* 432, 5343–5364. <https://doi.org/10.1016/j.jmb.2020.07.018>
- Wilén, C.B., Tilton, J.C., Doms, R.W., 2012. HIV: Cell Binding and Entry. *Cold Spring Harb. Perspect. Med.* 2, a006866. <https://doi.org/10.1101/cshperspect.a006866>
- Wyatt, P. J., 1993. Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta.* 272, 1-40 [https://doi.org/10.1016/0003-2670\(93\)80373-5](https://doi.org/10.1016/0003-2670(93)80373-5)
- Yang, X., Liu, Y., Cui, W., Liu, M., Wang, W., 2022. Distinct Gag interaction properties of HIV-1 RNA 5' leader conformers reveal a mechanism for dimeric genome selection. *RNA* rna.079347.122. <https://doi.org/10.1261/rna.079347.122>
- Ye, L., Gribling-Burrer, A.-S., Bohn, P., Kibe, A., Börtlein, C., Ambi, U.B., Ahmad, S., Olguin-Nava, M., Smith, M., Caliskan, N., von Kleist, M., Smyth, R.P., 2022. Short- and long-range interactions in the HIV-1 5' UTR regulate genome dimerization and packaging. *Nat. Struct. Mol. Biol.* 29, 306–319. <https://doi.org/10.1038/s41594-022-00746-2>
- Yeo, J.Y., Koh, D.W.-S., Yap, P., Goh, G.-R., Gan, S.K.-E., 2020. Spontaneous Mutations in HIV-1 Gag, Protease, RT p66 in the First Replication Cycle and How They Appear: Insights from an In Vitro Assay on Mutation Rates and Types. *Int. J. Mol. Sci.* 22, E370. <https://doi.org/10.3390/ijms22010370>
- Yoshida, M., Miyoshi, I., Hinuma, Y., 1982. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. *Proc. Natl. Acad. Sci. U. S. A.* 79, 2031–2035. <https://doi.org/10.1073/pnas.79.6.2031>
- Zhang, J., Crumpacker, C., 2022. HIV UTR, LTR, and Epigenetic Immunity. *Viruses* 14, 1084. <https://doi.org/10.3390/v14051084>
- Zhou, Y., Ratner, L., 2000. Phosphorylation of human immunodeficiency virus type 1 Vpr regulates cell cycle arrest. *J. Virol.* 74, 6520–6527. <https://doi.org/10.1128/jvi.74.14.6520-6527.2000>
- Zhu, P., Liu, J., Bess, J., Chertova, E., Lifson, J.D., Grisé, H., Ofek, G.A., Taylor, K.A., Roux, K.H., 2006. Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* 441, 847–852. <https://doi.org/10.1038/nature04817>

# Annexes





Review

# Post-Translational Modifications of Retroviral HIV-1 Gag Precursors: An Overview of Their Biological Role

Charlotte Bussienne, Roland Marquet, Jean-Christophe Paillart and Serena Bernacchi \*

Architecture et Réactivité de l'ARN, UPR 9002, Université de Strasbourg, CNRS, IBMC, 2 allée Konrad Roentgen, CEDEX F-67084 Strasbourg, France; c.bussienne@ibmc-cnrs.unistra.fr (C.B.);

r.marquet@ibmc-cnrs.unistra.fr (R.M.); jc.paillart@ibmc-cnrs.unistra.fr (J.-C.P.)

\* Correspondence: s.bernacchi@ibmc-cnrs.unistra.fr; Tel.: +33-(0)3-8841-7040

**Abstract:** Protein post-translational modifications (PTMs) play key roles in eukaryotes since they finely regulate numerous mechanisms used to diversify the protein functions and to modulate their signaling networks. Besides, these chemical modifications also take part in the viral hijacking of the host, and also contribute to the cellular response to viral infections. All domains of the human immunodeficiency virus type 1 (HIV-1) Gag precursor of 55-kDa (Pr55<sup>Gag</sup>), which is the central actor for viral RNA specific recruitment and genome packaging, are post-translationally modified. In this review, we summarize the current knowledge about HIV-1 Pr55<sup>Gag</sup> PTMs such as myristoylation, phosphorylation, ubiquitination, sumoylation, methylation, and ISGylation in order to figure out how these modifications affect the precursor functions and viral replication. Indeed, in HIV-1, PTMs regulate the precursor trafficking between cell compartments and its anchoring at the plasma membrane, where viral assembly occurs. Interestingly, PTMs also allow Pr55<sup>Gag</sup> to hijack the cell machinery to achieve viral budding as they drive recognition between viral proteins or cellular components such as the ESCRT machinery. Finally, we will describe and compare PTMs of several other retroviral Gag proteins to give a global overview of their role in the retroviral life cycle.

**Keywords:** HIV-1; Pr55<sup>Gag</sup> precursor; post-translational modifications; retroviral Gag precursors; retroviral life cycle

**Citation:** Bussienne, C.; Marquet, R.; Paillart, J.-C.; Bernacchi, S.

Post-Translational Modifications of Retroviral HIV-1 Gag Precursors: An Overview of Their Biological Role. *Int. J. Mol. Sci.* **2021**, *22*, 2871. <https://doi.org/10.3390/ijms22062871>

Received: 25 January 2021

Accepted: 9 March 2021

Published: 11 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Post translational modifications (PTMs) introduce a vast diversity in proteome including addition of chemical groups, like phosphorylation, methylation, acetylation, redox-based modifications, or alternatively, addition of polypeptides like ubiquitination, sumoylation or ISGylation. PTMs thus play a key role in functional proteomic by regulating proteins activity, their localization, and the interaction with cellular or viral factors. Even though many proteins are modified shortly after translation, PTMs can also occur at different steps such as after protein folding or protein re-localization to influence their biological activity at those specific sites (for reviews see [1,2]). Besides, depending on the nature of the modification, they can also finely tune reversible processes. Consequently, analysis of PTMs can provide an invaluable insight into cellular functions.

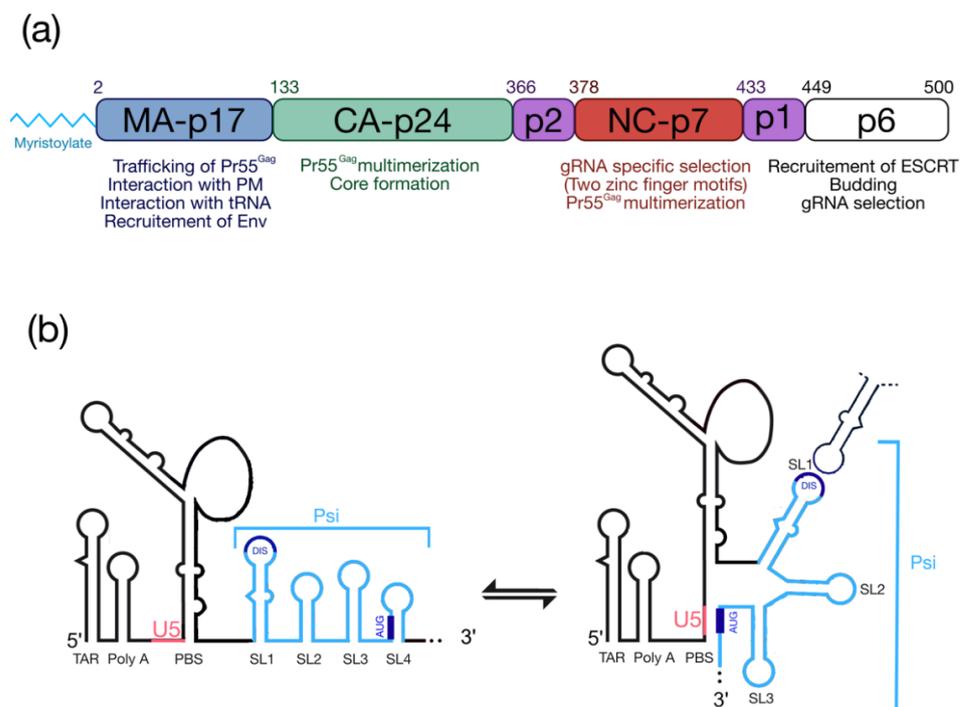
Viruses rely on the protein synthesis machinery of the host to support the production of viral progeny, and several cellular pathways are modulated by viruses to achieve the critical steps in viral replication. Hence, it is not surprising that viruses developed different strategies to either counteract or exploit PTMs of cellular factors, and that many viral proteins carry PTMs. Interestingly, PTMs are strongly involved in the regulation of different steps of the retrovirus viral cycle (for reviews see [3,4]). More specifically, in the HIV-1 (human immunodeficiency virus type 1) context, PTMs within the 55-kDa viral precursor, Pr55<sup>Gag</sup> (or Gag), were found to be necessary for regulating the last phase of the viral cycle, leading to the assembly of viral particles. Besides, several pieces of evidence

have shown that other retroviral Gag carry various PTMs regulate viral replication and pathogenesis. This review will summarize our current knowledge on PTMs observed in HIV-1 Pr55<sup>Gag</sup> and in other retroviral Gag proteins. Considering the role of the PMTs in the retroviral life cycle, the analysis of PTMs in retroviral Gag precursors could be particularly important for a deeper understanding of the molecular mechanisms driving retroviral replication. In a further step, this knowledge could contribute to the identification of new targets, and the design of new treatments against retroviral replication.

## 2. HIV-1 Pr55<sup>Gag</sup>

The HIV-1 Pr55<sup>Gag</sup> precursor (Figure 1a) plays a crucial for genomic RNA (gRNA) packaging, since it specifically selects the full-length gRNA amongst many other RNAs (cellular and spliced viral RNAs) and this process involves specific interactions between Pr55<sup>Gag</sup> and the highly structured 5' region of the gRNA [5,6], which contains the packaging signal (Psi) spanning SL (stem-loop)1 to SL4 in the 5'-end region of gRNA [7–9] (Figure 1b). In cells, the HIV-1 gRNA dimer in association with low-order Pr55<sup>Gag</sup> multimers [10–12] forms a viral ribonucleoprotein complex that traffics to the plasma membrane (PM) where the assembly of the viral particle occurs (for reviews see [13–15]). HIV-1 Pr55<sup>Gag</sup> is composed of four structural domains named matrix (MA), capsid (CA), nucleocapsid (NC), p6, and two spacer peptides (p2 and p1) (Figure 1a) [16] and each of them carry PTMs.

From the N-terminus, the 17 kDa MA domain that possesses a bipartite signal leads to Pr55<sup>Gag</sup> interaction with the PM. The first signal corresponds to the N-terminal myristoylated Glycine 2 (G2) (see § “HIV-1 Pr55<sup>Gag</sup> Myristoylation”), while the second one is constituted by a highly basic region (HBR) at the MA surface (for a review see [17]). MA was also found to interact with nucleic acids such as host tRNAs [18], and recent findings showed that MA-RNA binding ensures the specific interaction between Pr55<sup>Gag</sup> and the PM, by preventing nonspecific binding of Gag to intracellular membranes [19,20]. The CA is a 24 kDa domain that drives Pr55<sup>Gag</sup> multimerization and leads to formation of the viral core [21–23]. Next, NC is a 7 kDa domain, which is crucial for specific interaction with gRNA and for the incorporation of tRNA<sup>Lys3</sup>, which is the primer for reverse transcription. NC displays two zinc finger motifs (CCHC) that specifically interact with the Psi (Figure 1b) [24,25]. This domain also contributes to Pr55<sup>Gag</sup> multimerization thus promoting viral assembly [26–28]. At the C-terminal end of Pr55<sup>Gag</sup>, the unstructured p6 domain of 6 kDa is required for specific binding to the gRNA [29], and is involved in the recruitment of the ESCRT (Endosomal Sorting Complex Required for Transport) machinery that regulate viral particle budding. Finally, Pr55<sup>Gag</sup> codes for two spacer peptides, sp1 and sp2 (also named p2 and p1, respectively), regulating the kinetics of Pr55<sup>Gag</sup> maturation.

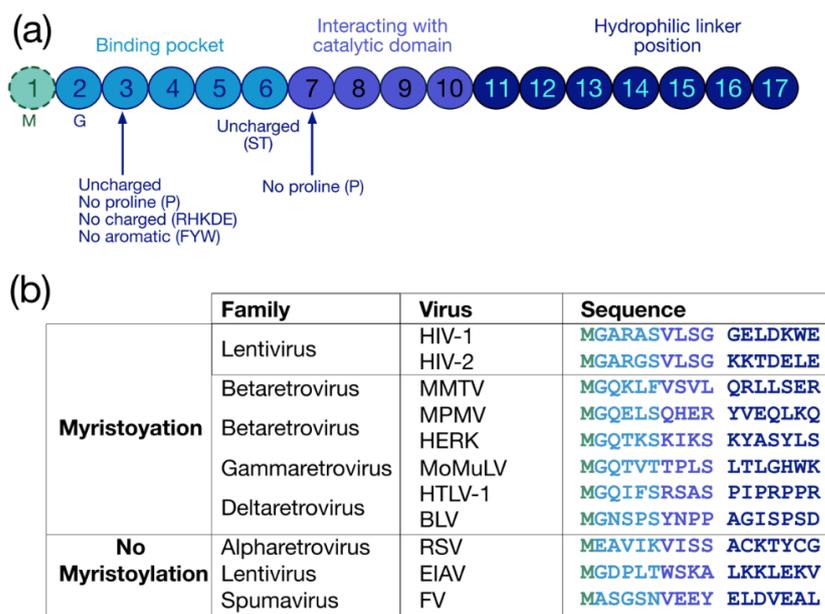


**Figure 1.** Pr55<sup>Gag</sup> and the 5'UTR of HIV-1 genomic RNA. (a) Functional domains of Pr55<sup>Gag</sup> and a short description of their roles. (b) Schematic representation of the secondary structure model of the 5'UTR (adapted from [29]). TAR: transactivation response element; Poly-A: 5'-copy of the polyadenylation signal; PBS: Primer Binding Site; DIS: Dimerization Initiation Site; Psi: packaging signal spanning SL1 to SL4; U5: unique region at the 5' end. The structure represents the U5-AUG conformation [5,6].

The next sections of this review will describe which PMTs are carried by the different Pr55<sup>Gag</sup> domains and what are their roles in the viral life cycle.

### 3. HIV-1 Pr55<sup>Gag</sup> Myristoylation

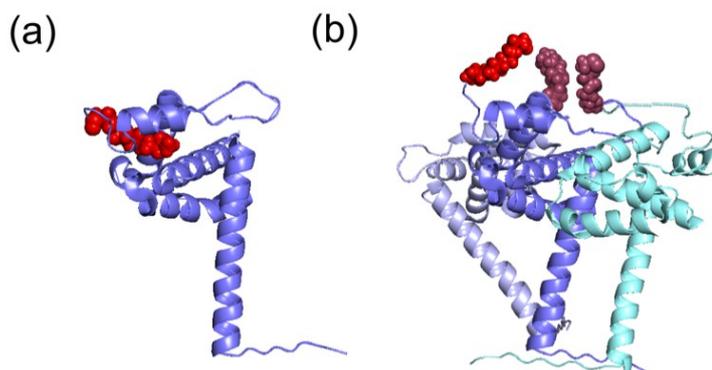
The myristoylation is a co-translational and irreversible modification consisting in the addition of a 14-carbon saturated fatty acid myristate to the protein via an amid bond by the N-myristoyl-transferase (NMT) (for reviews see [30–32]). The myristoylation can be achieved on an internal glycine (G) inside a consensus sequence recognized by NMTs, which is G-X2-X3-X4-(S/T/C)-X6 (Figure 2a). The G residue at the first position is necessary for this PMT, while at the second position there is preferentially an uncharged residue (except for proline (P)) or an aromatic amino acid. At the fifth position, uncharged residues are found, preferentially serine (S) and threonine (T) (for a review see [33]), while P is not accepted at the sixth position [34]. In sum, three regions finely regulate myristoylation: the binding pocket (positions from 1 to 6), the catalytic domain (positions from 7 to 10) and the hydrophilic linker (position from 11 to 17) [34,35] (Figure 2a).



**Figure 2.** Protein sequence required for myristoylation and sequences of retroviral myristoylated MA domains. (a) Pro-myristoylated consensus sequence underlying the three regions regulating myristoylation: the binding pocket (positions 1–6), the catalytic domain (positions 7–10) and the hydrophilic linker (positions 11–17) [34,35]. (b) Comparison of the first 17 residues of myristoylated MA domains in different retroviruses. Myristoylation is generally conserved in retroviruses such as lentivirus (HIV-1), betaretrovirus (Mason-Pfizer monkey virus (MPMV), mouse mammary tumor virus (MMTV), and human endogenous retrovirus type K (HERK)), gammaretrovirus (Moloney murine leukemia virus (MoMuLV) and murine leukemia virus (MLV)), and deltaretrovirus (human T-lymphotropic viruses (HTLV-1) and bovine leukemia virus (BLV)), but not in alpharetrovirus (Rous sarcoma virus (RSV)), some other lentivirus (equine infectious anemia virus (EIAV)), and in spumavirus (foamy virus (FV)).

Myristoylation is rather conserved in retroviruses (Figure 2b) (For reviews see [17,36] and [37,38]), and this PTM globally regulates the interaction of retroviral precursors with membranes and sub-membrane domains, such as lipid rafts. However, this modification is not sufficient by itself for membrane binding, and a distant polybasic domain is thus required to complete the optimal attachment of myristoylated proteins to the PM (for reviews see [39,40]). In HIV-1, this task is reached by the HBR spanning residues 17 to 31 of the MA domain, which contributes to a strengthening of the interaction with the PM thanks to electrostatic interactions with the negatively charged PI(4,5)P2 [41–43].

The myristoyl moiety can be exposed or sequestered in the hydrophobic pocket of the mature MA (Figure 3a,b) by the so-called myristoyl-conformational switch [37,39,42,44,45], which controls the exposure of myristoyl group for insertion into the PM, thus contributing to the prevention of aberrant interactions with intracellular membranes. The myristate exposure was found to be triggered by the interactions occurring between Pr55<sup>Gag</sup> and PI(4,5)P2 [41]. Besides, NMR studies demonstrated that myristate exposure is also regulated by the trimerization of the protein [37,42], and this would explain why the mature MA displays a lower affinity for membranes in comparison with the full-length precursor [37,41,42]. Indeed, several Pr55<sup>Gag</sup> domains, such as CA, p2 and NC, contribute to the self-association of the precursor and, as a consequence, to the myristate exposure (Figure 3b) [37]. Accordingly, mutational experiments on these domains inhibiting Pr55<sup>Gag</sup> multimerization, impair Pr55<sup>Gag</sup> binding to the membrane [37].



**Figure 3.** Different structural conformations of HIV-1 MA monomer or trimer. The tertiary structures of the MA domain in the different conformations of the switch look similar. (a) The MA domain in its monomeric conformation (in blue) displays a sequestered myristoyl group (in red) (PDB: 1UPH [36]). (b) Representation of the trimer of MA (in blue, light blue and cyan) and the corresponding exposed myristoyl groups (in red). This model was proposed according to which the myristoyl group is exposed in the multimeric form, thus allowing its interaction with PM (adapted from [37]).

#### 4. Gag Myristoylation in Other Retroviruses

The MA domains of retroviral Gag polyproteins display two main roles: they participate in genome incorporation, as several analyses recently pointed out, and they are implicated in membrane association. Interestingly, the majority of the retrovirus family displays a myristoylated MA domain (for reviews see [46–49]).

Among the different genera in which MA is myristoylated, the genus gammaretrovirus is composed by simple and oncogenic retroviruses. One representative virus of this family is MLV inducing leukemias or lymphomas in mice [50]. The MA domain of the primary form of MLV Gag, Pr65<sup>Gag</sup>, is myristoylated and contains a polybasic region in its globular domain that interacts electrostatically with PI(4,5)P<sub>2</sub> at the PM, similarly to HIV-1 [47]. Besides, MLV has the particularity of encoding an additional form of Gag, gPr80<sup>Gag</sup>, which is glycosylated, but not myristoylated, and this last one is involved in the Pr65<sup>Gag</sup> trafficking to the PM [51]. However, beside MLV budding at PM, intracellular budding events can also occur into multivesicular bodies (MVBs) or in intracellular compartments as late endosomes in which virus-like particles (VLPs) accumulated [52]. Then, the deltaretrovirus genus contains complex and oncogenic retroviruses, and consists of two different groups, the primate T-lymphotropic viruses (PTLVs) including HTLV-1 and non-primate species, such as BLV [53–55]. Similarly, to MLV, the assembly of those retroviruses can occur at the PM, as well as in intracellular compartments such as late endosomes, MVBs or similar compartments [56]. The myristoylation of MA and the presence of basic amino acids leads to membrane binding and is, in this case, a PI(4,5)P<sub>2</sub>-independent process [57–59]. Indeed, the HTLV-I viral precursor Pr53<sup>Gag</sup> is able to bind membranes by electrostatic interactions involving the zwitterionic phosphatidylcholines (PC) and the negatively charged phosphatidylserines (PS) contained in endocytic membranes [57,60]. Moreover, a model was proposed in which the HBR in the HIV-1 MA domain would bind RNA to prevent premature or non-specific binding to cellular membranes [19,20,61]. Interestingly, a similar regulation between MA and RNA was proposed for BLV [62]. Conversely, the lack of this RNA regulation in HTLV-1 could explain the binding of myristoylated MA to the cellular membranes of intracellular compartments [57]. Finally, betaretroviruses show many similarities with lentiviruses, including a myristoylated MA domain [63]. This genus is composed of two groups: the first one is represented by MMTV [64] and by MPMV; and the second one is represented by HERK [65]. Myristoylated-deficient HERK Gag was observed to localize in the nucleus [66]. Contrary to other lentiviruses, NMR structures of MA domains of MPMV [63] and MMTV [67] show that the

myristate group is hidden inside the MA in its oligomeric form. These differences suggest that betaretroviruses have developed different strategies to sequester the myristoyl group until the VLP is bound to the PM. At this site, a conformational change, leading to exposure of the myristate group would occur, similarly to other retroviral MAs that bind PM [67].

In contrast, some retroviral Gag precursors are not myristoylated. Indeed, alpharetroviruses represents simple and oncogenic retroviruses like RSV [68]. At the PM, RSV Pr76<sup>Gag</sup> interacts with charged lipids PI(4,5)P2 [38,69], and to ensure proper Pr76<sup>Gag</sup>-PM association, the lack of myristoylation is then counterbalanced by electrostatic interactions occurring between the inositol phosphates and a membrane binding domain (MBD), which is composed of basic residues forming a patch of clustered lysines (K) and arginines (R) on the MA surface ([49,68,70], for a review see [71]). Similarly, the MA domain of the lentivirus EIAV [72,73] is not myristoylated, but binds preferentially with phosphatidylinositol 3-phosphate (PI(3)P) with a higher affinity compared to PI(4,5)P2 [73,74]. Finally, foamy viruses (FV) as the PFV presents interesting differences compared to HIV-1 ([75], for a review see [76]). In particular, the FV Pr74<sup>Gag</sup> displays a limited number of PTMs compared to the other retroviruses, and strikingly, the FV MA domain contains neither the HBR nor a myristoylation modification. All those elements emphasize a different evolutionary history among retroviruses [75]. Indeed, in this case, viral Env proteins play a major role for viral budding, and the co-expression of Pr74<sup>Gag</sup> with Env is necessary for VLP production [77].

In sum, there are three main distinct strategies used by retroviruses to target membranes for budding. The first one requires the myristate exposure and a highly basic region (HBR) in the MA domain of retroviral precursors to interact with PM. The two others display dispensable myristoylation to achieve proper membrane binding since the hydrophobic interactions are in this case substituted by electrostatic ones produced by a basic domain in the MA, or alternatively by interactions between the precursor and viral elements such as Env proteins.

## 5. HIV-1 Pr55<sup>Gag</sup> Phosphorylation

Phosphorylation consists of the addition of a phosphate group to the side chain of amino acids. This PMT modifies the local electrostatic potential of proteins, induces conformational modifications, and affects the protein subcellular localization (for a review see [78], and [79,80]). Kinases, which are the enzymes that catalyze the transfer of phosphate group, have a role at multiple steps of HIV-1 viral, and the inhibition of cellular kinases interacting with HIV-1 at the nuclear level has been shown to affect the viral replication cycle [81]. Among HIV-1 viral proteins, which are phosphorylated, there is Pr55<sup>Gag</sup> (Table 1 and Figure 4). The MA domain is a substrate for the protein kinase C (PKC) [82], which catalyzes S and T phosphorylation. Several studies identified S111 in HIV-1 MA as the substrate for PKC [82]. Interestingly, substituting S111 with an alanine (A) led to decreased association of Pr55<sup>Gag</sup> with PM, even though MA was myristoylated. This suggests that PKC could also be involved in membrane binding by regulating the exposure of the myristoyl group [83,84].

Alpha-screen assays allowed us to screen for human kinases interaction with the HIV-1 precursor, and the p6 domain resulted to be a target for PKC. In a further step, mass spectrometry indicated the phosphorylation of S488 residue [85,86]. Its substitution with a hydrophobic aromatic residue such as phenylalanine (F), which can occur spontaneously during anti-retroviral treatments, was found to perturb CA-SP1 processing, virus morphogenesis, maturation and virion infectivity [87–89]. On the other hand, the substitution of S488 by another non phosphorylatable residue, such as asparagine (N), displayed no global impact on infectivity, thus suggesting that the production of non-mature viral particles would not be due to the lack of phosphorylation, but by the substitution itself [89]. Moreover, the phosphorylation of the p6 domain was found to be also important for the recruitment of the viral factor Vpr. As a consequence, the inhibition of PKC activity

reduced Vpr level in virions, and this affected HIV-1 infectivity [85]. The p6 domain is the main phosphorylated domain in Pr55<sup>Gag</sup> and can be phosphorylated at several positions [86,90]. Indeed, phosphoamino acid analysis [90] and mass spectrometry experiments [86] identified several phosphorylated amino acids (Table 1 and Figure 4), that were found to globally promote viral budding [91]. Moreover, electron microscopy analysis revealed that mutation T471A leads to immature viral particles incompletely separated from PM, and immunoblotting analysis showed an incomplete Pr55<sup>Gag</sup> proteolytic maturation [91]. In contrast, other findings showed no effects on assembly or on viral release when T471 was substituted with isoleucine (I) or N. Since none of these amino acids can be phosphorylated, it is possible that the observed differences were not due to phosphorylation itself. Furthermore, except for T456 located in the PTAP late domain, the other eleven positions that can be phosphorylated in the p6 domain present redundancy. Mutagenesis experiments confirmed that the modifications of those residues seem to be dispensable for viral release and infectivity [86].

Experiments using an inhibitor of cyclin-dependent kinases [92] showed that also a MAP kinase, the extracellular-signal-regulated kinase 2 (ERK2), is involved in p6 phosphorylation, and this factor can be incorporated into viral particles by interacting with the S148-P149 motif in CA and T471-P472 in p6 [91,93–96] (Table 1, Figure 4). HIV-1 particles without active ERK2 were found to be poorly infectious due to a defect in reverse transcription [93,95]. Interestingly, ERK2 phosphorylates other viral proteins including Rev [97], Nef [98], Vif [99,100], and mature MA [95]. Besides, the substitution of four highly conserved and major phospho-acceptor S residues in the mature MA (Table 1) with A was found to impair viral replication [95,101].

Finally, the tyrosine kinase Src can also be incorporated into HIV-1 particle [102], and it is involved in the phosphorylation of the tyrosine (Y) 132 in a minority of mature MA proteins. This PMT was shown to play a role in the early phases of HIV-1 replication as the proviral DNA nuclear import [103] and its deletion causes the enhancement of MA accumulation in the cytoplasm at the expense of PM. On the contrary, Src overexpression was found to promote the localization of Pr55<sup>Gag</sup> at the PM [102].

In sum, HIV-1 Pr55<sup>Gag</sup> is phosphorylated by at least three kinases, PKC, ERK-2 and Src. Interestingly, mutation of phosphorylated residues in the p6 domain revealed that this domain, in addition to MA, can act as membrane targeting domain of Gag [104]. However, phosphorylation positions in p6 mainly display redundancy, thus hindering the evaluation of the impact of each individual phosphorylated residue.

```

1  MGARASVLSG GELDKWEKIR LRPGGKKQYK LKHIVWASRE LERFAVNPGL
51  LETSEGCRQI LGQLQPSLQT GSEELRSLYN TIAVLYCVHQ RIDVKDTKEA
101 LDKIEEEQNK SKKKAQQA AA DTGNSQVSQ NYPIVQNLQG QMVHQAISPR
151 TLNAWVKVVE EKAFSPEVIP MFSALSEGAT PQDLNTMLNT VGGHQAAQM
201 LKETINEEAA EWDR LHPVHA GPIAPGQMR PRGSDIAGTT STLQEQIGWM
251 THNPPIPVGE IYKRWIILGL NKIVRMYSP SILDIRQGP EPFRDYVDRF
301 YKTLRAEQAS QEVKNWMTET LLVQANPDC KTILKALGPG ATLEEMMTAC
351 QGVGGPGHKA RVLAEAMSQV TNPATIMIQK GNFRNQRKTV KCFNCGKEGH
401 IAKNCRAPRK KGCWKCGKEG HQMKDCTERQ ANFLGKIWPS HKGRPGNFLO
451 SRPEPTAPPE ESFRFGEETT TPSQKQEPID KELYPLASLR SLFGSDPSSQ

```

■ Phosphorylation

**Figure 4.** Phosphorylated residues in HIV-1 Pr55<sup>Gag</sup>. The different colors represent the Pr55<sup>Gag</sup> domains, MA (blue), CA (green), spacer peptides p1 and p2 (purple), NC (red), and p6 (black). Phosphorylation positions are highlighted in yellow. TP (in p6) and SP (in CA) motifs involved in the ERK2 recruitment and incorporation into viral particle are indicated in bold [91,93–96].

**Table 1.** Summary of different roles of phosphorylated residues in HIV-1 Pr55<sup>Gag</sup>.

Domain	Residue	Enzyme	Observations and Associated (or Proposed) Roles	References
MA	S9	ERK2	Involved in the viral replication Phosphorylation of the mature form of MA	[95,101]
	S67			
	S72			
	S77			
MA	S111		PKC could be involved in membrane binding by regulating the exposure of the myristoyl group	[82–84]
	Y132	Src	In MA mature 1% of Y132 is phosphorylated Src overexpression was found to promote the localization of Pr55 <sup>Gag</sup> at the PM	[102,103]
CA	S148	ERK2	Belongs to S-P motif involved in recruitment of ERK-2	[93–96]
	T456		Belongs to the PTAP late domain Potential role in viral infectivity and assembly	[86]
	T470		Redundancy with T471, S473, S488, S491, and S499	[91]
p6	T471	ERK-2	Belongs to T-P motif involved in the recruitment of ERK-2	[91]
			Its substitution induces the accumulation of immature viral particles incompletely separated from PM	[91]
			Redundancy with T470, S473, S488, S491, and S499	[91]
			Effects on assembly or on viral release is not due to phosphorylation	[86]
	S473		Redundancy with T470, S471, S491, and S499	[91]
	S488	ERK2	Viral particles without active ERK2 were found to be poorly infectious due to a defect in reverse transcription	[93,95]
			Involved in the phosphorylation of other viral proteins: Rev, Nef, Vif, mature MA	[97–100]
The p6 domain of Pr55 <sup>Gag</sup> is a target for PKC The inhibition of PKC activity reduced Vpr level in virions			[85,86,90] [85,87,88]	
S488	PKC	Its mutation with F perturbs: - Viral morphology, maturation and infectivity	[87–89]	
		Effects on assembly or on viral release could be not due to phosphorylation	[86]	
		S491		Redundancy with T470, S471, S473, and S499
S499		Redundancy with T470, S471, S473, and S491	[91]	

## 6. Gag Phosphorylation in Other Retroviruses

Phosphorylation is a conserved modification in the retroviral family (Table 2). In alpharetroviruses, within the RSV MA domain, a small proportion of Y residues results in being phosphorylated [105], as well as S68 and S106 residues (Table 2). However, S68 seems to be transitionally phosphorylated, while S106 is the main phosphorylated signal [106]. Besides, MA phosphorylation could be involved in the recruitment of factors promoting NC phosphorylation [106,107]. In turn, phosphorylation of S529 in NC was found to be necessary for the specific interaction with gRNA [108], but no other notable effects on assembly, or on infectivity, were observed [106].

The deltaretroviruses, HTLV-1 MA is also a phosphoprotein, and S105, which is located close to the two late domains involved in viral release [109], PPPY [110] and PTAP [111], is phosphorylated by ERK-2. Similarly, to HIV-1, ERK-2 is incorporated into HTLV-1 particles, and phosphorylation of the MA domain was found to be involved in virus release and budding efficiency [110].

Interestingly, betaretroviruses such as MPMV encode a phosphoprotein pp24 within the Gag precursor, and its C-terminal cleavage produces the protein pp18 which contains proline-rich motifs (PPPY). Deletion assays indicated that the phosphorylated residue Y205 in pp18 is dispensable for capsid assembly, but is necessary for the viral release [112].

Immunoprecipitation experiments identified the presence of phosphoserines in pp18, [113,114] displaying a redundant character. Similarly, for spumaviruses such as FV, mapping of the p4 domain revealed that seven residues can be phosphorylated (Table 2), but a single substitution of those residues displayed no influence on viral replication [115]. In gammaretrovirus, the phosphorylation of the RNA binding phosphoprotein (p12) within the Gag precursor was found to be necessary for early events of viral life cycle and virion production [116,117]. Mutagenesis experiments identified two residues which can be phosphorylated (S192 and S209). In particular, S192 mainly contributes to p12 phosphorylation and its substitution by A impairs viral assembly and infectivity. However residual phosphorylation can also occur at other positions (Table 2) [117], thus suggesting a redundant character of these modified amino acids. Indeed, the single substitution of one of these residues induced different levels of phosphorylation in p12, displaying no overall effect on the viral cycle [117], even though these PTMs were proposed to modulate p12 early and late functions and p12 viral RNA-binding activity [117,118].

In conclusion, similarly to HIV-1, the kinases PKC and ERK-2 are the main drivers of retroviral Gag phosphorylation. Interestingly, ERK-2 can be incorporated into the viral particle of HTLV-1. Globally, these PTMs generally seem to play a role in viral particle release and in virus infectivity, even though the impact of the phosphorylation rate in retroviral proteins is complicated by the redundancy of phosphorylated positions.

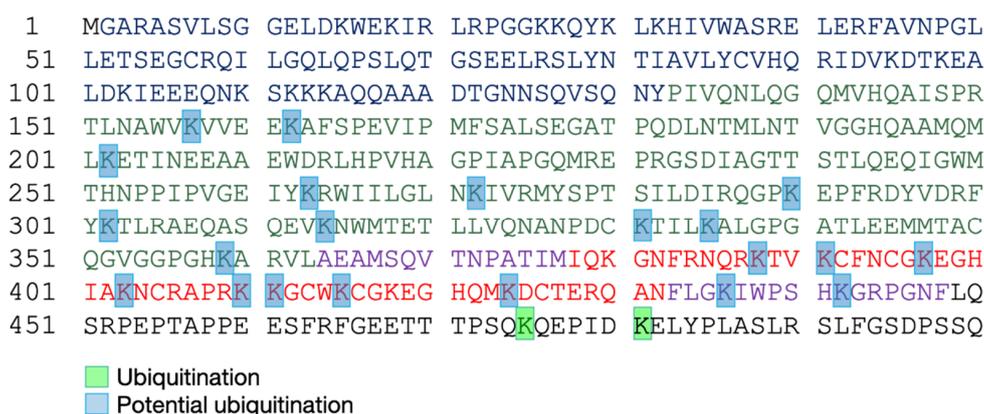
**Table 2.** Summary of phosphorylated positions in the different domains of retroviral Gag precursors.

Retrovirus	Protein	Residues	Enzyme	Observation and Associated (or Proposed) Roles	References
RSV	MA	Y15	PKC	No effect on the viral cycle	[105,106]
		Y46			
		S68			
	NC	S106	PKC	Major site of phosphorylation Involved in the recruitment of factors which promote NC phosphorylation	[105,106]
		Y155	PKC	No effect on the viral cycle	[105,106]
HTLV-1	MA	S529	ERK2	Role for the specific interaction with the gRNA	[106,107]
MoMuLV	p12	S105	ERK2	Close to late domains (PPPY et PTAP) Involved in viral release and budding efficiency	[110]
		S137		- Redundancy	[117]
		S148		- Modulation of early and late functions and the RNA-binding activity of p12	
		S150			
		S173			
S192	- S192 mainly contributes to p12 phosphorylation and its substitution by A impairs viral assembly and infectivity	[117]			
S209		]			
MPMV	p18	Y205		Belongs to proline-rich motif (PPPY) Necessary for the viral release	[112]
		S167	Redundancy	[113,114]	
		S176			
FV	p4	S211			
		S116	Redundancy	[115]	
		S119			
		S120			
S124					

## 7. HIV-1 Pr55<sup>Gag</sup> Ubiquitination

Another crucial PTM for retroviral infectivity is ubiquitination. This PTM consists of intracellular protein modification by adding one or more ubiquitin (Ub) protein(s) (for a review see [119]). Ub is a 76-amino acid polypeptide, which has a conserved structure [120]. The Ub sequence contains seven K residues that can be used for subsequent Ub linkage leading to polyubiquitination (for a review see [121]), even if the two most common polyubiquitination chains consist in the formation of Ub chain connected to residues K48 or K63 of Ub. Monoubiquitylation corresponds to a signal for DNA repair, and vesicle sorting or signal transduction, while polyubiquitinated proteins are often targeted to the 26S proteasome for degradation, or alternatively involved in regulation of the endocytosis of ESCRT-dependent cargo proteins into Multi Vesicular Bodies (MVB) (for a review see [122]) and DNA damage response [123]. Ubiquitination can be reversed by deubiquitinating enzymes (DUB) [124].

HIV-1 Pr55<sup>Gag</sup> is ubiquitinated in its domains at different levels (Table 3 and Figure 5). Indeed, MA, CA, and NC are monoubiquitinated, while p2 can be mono or bi-ubiquitinated [125]. The cumulative mutations of ubiquitin acceptor sites were observed to cause generally budding defects, even if the substitution of K residues by R in CA (Table 3) revealed very limited effect on viral release, showing that these ubiquitination sites are likely redundant [126]. Besides, it was observed that the level of Pr55<sup>Gag</sup> ubiquitination increases in cellula when a full-length HIV-1 molecular clone is expressed in comparison to a Pr55<sup>Gag</sup> expression plasmid, suggesting a role of other viral proteins in Pr55<sup>Gag</sup> ubiquitination [125]. Globally, the ubiquitination of Pr55<sup>Gag</sup> was found to be involved in the viral release and, during HIV-1 assembly, viral particles incorporate free Ub proteins corresponding to about 10% of the Pr55<sup>Gag</sup> level, and around 2–5% of ubiquitinated Pr55<sup>Gag</sup> are mono-ubiquitinated [125,127–130]. When the level of free Ub in cells is reduced by proteasomal inhibition, the number of free Ub in viral particles and the number of mono-ubiquitinated residues in the p6 domain of Pr55<sup>Gag</sup> also decreased [125,127,131]. However free Ub incorporation into viral particles seems to be independent from the global Pr55<sup>Gag</sup> ubiquitination state [132], and the ubiquitination level in virions increased upon overexpression of free Ub [133]. Furthermore, ubiquitination seems to take place at the PM, and interestingly the level of Pr55<sup>Gag</sup> mono-ubiquitination was found to be directly correlated with ability of the precursor to bind the PM [134].



**Figure 5.** Ubiquitinated residues in HIV-1 Pr55<sup>Gag</sup>. The domains of Pr55<sup>Gag</sup> are represented by different colors (see Figure 4). Experimentally identified ubiquitinated positions are highlighted in light green. Potential ubiquitinated positions are highlighted in light blue.

The C-terminal p6 domain is the most ubiquitinated domain in Pr55<sup>Gag</sup> [125], and K475 and K481 are the major targets. Even if these mono-ubiquitinated residues are neither directly involved in virus release, nor in infectivity, they were found to be necessary

to promote the overall ubiquitination of Pr55<sup>Gag</sup> [132]. Besides, the mutation of the highly conserved and phosphorylated S488 residue in p6 domain with F (S488F), which can occur spontaneously during anti-retroviral treatments, has not only an impact on virus morphogenesis, maturation and virion infectivity (Table1) [87–89], but it can also induce conformational changes in p6, resulting in an enhanced interaction of Pr55<sup>Gag</sup> with the PM. This would lead to the polyubiquitination of the precursor and consequently to its proteasomal degradation [104].

The p6 domain is known to be involved in the recruitment of host factors, such as Tsg101 (Tumor susceptibility gene 101) and ALIX (ALG-2 interacting protein X), and ubiquitination of those factors strongly promote viral budding [135]. In this frame, fusion experiments in which the p6 domain was coupled with Ub showed that the affinity of Tsg101 for p6 in this case results in being strengthened [136], and the ubiquitination of Pr55<sup>Gag</sup> can increase Tsg101 recruitment [137]. Besides, Tsg101 displays an N-terminal Ub E2 variant (UEV) domain that shows homology with E2 Ub ligases, and that can specifically bind Ub proteins, as well the PTAP late domain in Pr55<sup>Gag</sup> [136]. During assembly, the interaction of Pr55<sup>Gag</sup> with the PM promotes the intermolecular interaction between Tsg101 and the PTAP domain in Pr55<sup>Gag</sup> [137]. In this conformation, the di-ubiquitinated K63 of Tsg101 was found to interact with p6, with the consequence of impairing the potential polyubiquitination of the precursor at PM [137].

Finally, the ESCRT-III-associated ALIX protein is also ubiquitinated [128] and specifically interacts with the E3 ubiquitin-protein ligase NEDD4 that can bind the proline-rich retroviral domain PPPY. The interaction between NEDD4 and the retroviral precursor leads to the recruitment of the ESCRT-III complex, including the eukaryotic sucrose non-fermenting protein 7 (Snf7), and the vacuolar protein sorting-associated proteins Vps 2, Vps20 and Vps24, and Vps4 in order to promote retroviral release [138–141]. Since in HIV-1 the PPPY domain is absent, ALIX recruits directly NEDD4 to facilitate this step [128,129].

**Table 3.** Summary of ubiquitinations in HIV-1 Pr55<sup>Gag</sup> proteins.

Domain	Residues	Observation and Associated (or Proposed) Roles	References
MA		Mono-ubiquitination	[125,126]
CA	K157	Mono-ubiquitination Observed redundancy	[125,126]
	K162		
	K202		
	K263		
	K272		
	K290		
	K302		
	K314		
	K331		
	K335		
NC	K359	Mono-ubiquitination	[125,126]
	K388		
	K391		
	K397		
	K403		
	K410		
	K411		
K415			
p2	K424	Mono or di-ubiquitination	[125,126]
	K436		
p6	K442	Mono or di-ubiquitination. Most ubiquitinated domain in Pr55 <sup>Gag</sup>	[125,126]
	K475		

	Involved in global Pr55 <sup>Gag</sup> ubiquitination	
	Major targets for mono-ubiquitination	
K481	No effect on virus release and infectivity	[125,132]
	Involved in the global Pr55 <sup>Gag</sup> ubiquitination	
	Conformational changes: formation of a hydrophobic patch in $\alpha$ -helix at the C-terminus of p6	
S488F	Leads to strong interaction of Pr55 <sup>Gag</sup> with the PM	[104]
	This structure promotes L48 linked polyubiquitination	

## 8. Gag Ubiquitination in Other Retroviruses

The role of ubiquitination in the retroviral cycle is not yet fully elucidated. Some retroviruses display a functional contribution of Ub modifications in virus release such as MLV, MPMV or RSV, and for those viruses, it was shown that, similarly to what observed for HIV-1 [131], the inhibition of proteasome not only induces a reduction of the level of free Ubs in the cytoplasm, but also impairs the release of the viral particles (Table 4) [127,132,133,142]. In addition, fusion experiments between RSV Pr76<sup>Gag</sup> and Ub, or overexpression of Ub, displayed an increase in viral particle release [142], thus supporting the idea that ubiquitination of retroviral precursors is crucial for viral budding [133]. However, for other retroviruses such as MMTV or HTLV-1, to date it was not possible to identify a precise role of ubiquitination [127] (Table 4). Besides, the inhibition of the proteasome did not impair the budding of EIAV [127,143]. On the other hand, similarly to HIV-1, EIAV particles contain free Ubs corresponding to 10–15% of Gag proteins. Likewise, the C-terminal p9 domain is mono-ubiquitinated and contains a YPDL late domain which is involved in the recruitment of the ESCRT machinery [143,144]. Moreover, p9 also contains an Ub-like motif (NVKEKD) that may contribute to virus release, thus suggesting alternative release pathways for EIAV even if Ub quantity is low [133,144].

MMTV Gag is monoubiquitinated in its p8 domain, in CA, and is potentially di-ubiquitinated in NC [127]. In comparison with other retroviruses, MMTV does not contain late domains such as PPPY and PTAP, but an alternative PSAP late domain was found in CA, although its functional role was not yet elucidated. Besides, YXXL motifs, which also represent alternative late domains, were identified in MA and in pp21 viral factor. Importantly, since Gag ubiquitination seem to take place mostly in regions close to the late domains [133], it is possible that the presence of these alternative late domains in EIAV and MMTV precursors promote virus release.

In HTLV-1, more than 40% of MA are ubiquitinated [111,145,146], and MA can be mono- and di-ubiquitinated [146]. Furthermore, mutagenesis experiments identified K74 in Pr53<sup>Gag</sup> as the main substrate for ubiquitination [146]. Indeed, substitution experiments in which the K74 is replaced by an R resulted in a decreased release of infectious particles [146]. The ubiquitination of K74 could also play a role in the recruitment of NEDD4 [146], which is also involved, through its interaction with the PPPY late domains, in the release of other retroviruses such as MMPV [112], avian sarcoma virus (ASV) [147], and MLV [132,148].

In RSV particles, more than 100 free Ubs were found [142,149]. However, contrary to other retroviruses, RSV displays free Ubs exclusively into mature viruses [149]. Since Pr76<sup>Gag</sup> mono-ubiquitination was found to be necessary for budding and to recruit the ESCRT machinery [142,149], it is thus possible that the presence of free Ubs could be the result of a host-encoded and encapsidated deubiquitinating enzyme (DUB) [124]. Interestingly, this process was also observed to occur during budding or cells lysis [124]. Finally, Gag precursors from spumavirus encode a very limited number of K residue. This observation suggests that Gag of spumavirus could not be a favorable substrate for the ubiquitination machinery [150].

**Table 4.** Summary of ubiquitinations in the different domains of retroviral Gag proteins.

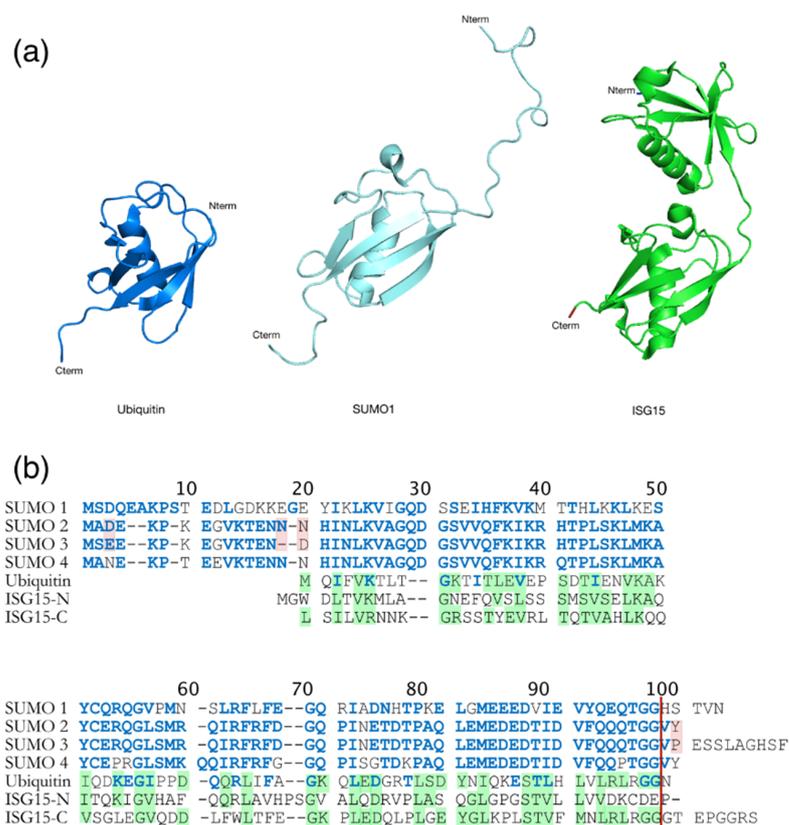
Retrovirus	Domain	Residues	Associated (or Proposed) Roles	References
HIV			About 100 free Ubs are incorporated into viral particles 2–5% mono-ubiquitinated	[130,132–134]
			Pr55 <sup>Gag</sup> ubiquitination promotes the virus release K475 and K481 in p6 domain are major targets for ubiquitinations Pr55 <sup>Gag</sup> ubiquitination is correlated with the ability of the precursor to bind the PM	[132,134]
MLV	p12		Increases viral release and infectivity	[127]
			PPPY late domain is involved in the recruitment of NEDD4	[148]
HTLV-1	MA		Ubiquitination of this domain has a crucial role in release	[111,145,146]
			40% of MA are ubiquitinated MA can be mono- or di-ubiquitinated	[111,145,146]
		K74	Substrate for Pr53 <sup>Gag</sup> ubiquitination	[146]
MPMV			PPPY late domain is involved in the recruitment of NEDD4	[112]
RSV			- Mono-ubiquitination is crucial for viral release - Ubiquitylation is required for the recruitment of ESCRT machinery and for the budding	[133,143,149]
			- Contains free Ubs into mature particles - Pr76 <sup>Gag</sup> mono-ubiquitination is necessary for budding and to recruit the ESCRT machinery	[142,149]
EIAV	p9		10–15% of the molar level of the Gag protein of free Ub	[143]
			Proteasome inhibition: does not impair the release	[143]
			Ub-like motif (NVKEKD) Mono-ubiquitinated domain Contains YPDL late domain	[143,144]
MMTV	MA (p10) pp21 p8 CA (p27) NC (p14)		Proteasome inhibition: does not decrease the release	[143]
			YXXL Late domain	[127]
			YXXL Late domain	[127]
			Mono-ubiquitinated	[127]
			Mono-ubiquitinated	[127]
			PSAP domain	[127]
PFV				[127]
			Encodes for a very restricted number of K residues	[150]

### 9. HIV-1 Pr55<sup>Gag</sup> Sumoylation

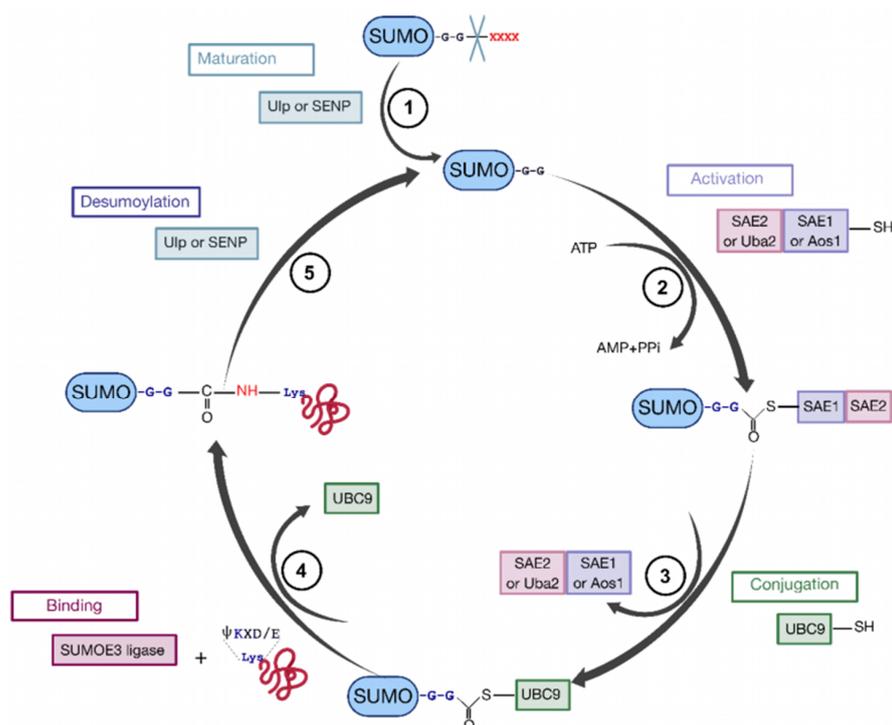
Another modification important for retroviral infectivity is sumoylation, which is a reversible PTM and consists of intracellular protein modification by a covalently attached small Ub related modifier (SUMO) protein to a K substrate (for reviews see [151,152]). Even though SUMO is structurally comparable to Ub (Figure 6a), it presents many differences in amino acids sequence (only 18% of homology) (Figure 6b) [152]. This PTM is usually involved in the maintenance of genomic integrity, with a role in repair of damaged DNA, and in the regulation of transcription and in gene expression. Like ubiquitination, sumoylation is involved in intracellular signal transduction and can regulate biological processes such as apoptosis, immune response, and carcinogenesis. Besides, sumoylation controls protein localization and it can induce protein conformational changes. SUMOs are highly conserved in eukaryotes, and four SUMO isoforms (SUMO-1 to SUMO-4) are present in mammals [152–154] (Figure 6b). Similarly, to Ub, the C-terminus region of SUMO-1 is linked to  $\epsilon$ -amino groups of K residues in the target protein [155,156]. SUMO-1 was interestingly found to counterbalance the effect of ubiquitination [157]. SUMO-2 and SUMO-3 are mainly involved in the cellular response to environmental stresses [156] and display very similar sequences with more than 95% identity [151,152,156]. For this

reason, they are often named SUMO-2/3. Finally, SUMO-4 is less well known, and its mRNA had been found in few organs such as kidney, spleen, and lymph nodes [152].

To sumoylate a protein, different successive biochemical reactions are required [152,158–160] (Figure 7). Generally, the consensus sequence for K sumoylation is  $\psi$ KXD/E ( $\psi$  stands for a hydrophobic residue). Nevertheless, targets with non-consensus acceptor sites have also been identified [151,152].



**Figure 6.** Comparison between Ub and Ub-like proteins: SUMO and ISG15. (a) Structural comparison between Ub (heavy blue, PDB: 1A5R), SUMO-1 (light blue, PDB: 2QHO), and ISG15 (green, PDB: 3PHX). They contain a typical  $\beta\beta\alpha\beta\alpha\beta$  fold, even if SUMO-1 has long unstructured N-terminal domain which is absent in Ub. ISG15 is composed with two Ub-like domains in N- (TSG15N) and C- (TSG15C) terminus. (b) Amino acid sequence alignments of Ub, the four SUMO homologs and ISG15 from humans. Identities and similarities are indicated between Ub and SUMO (blue residues into Ub sequence) and between Ub and ISG15 (shaded green residues in Ub and ISG15). Differences between SUMO-2 and 3 are highlighted in pink. The red vertical line represents the GG end free after the maturation step required for sumoylation. The amino acid sequence homology between SUMO and Ub is 18% [152], and 30% between Ub and ISG15 [161].



**Figure 7.** The cycle of sumoylation. This modification is catalyzed by different enzymes and consists in ligation of SUMO protein to K residues of protein substrates. (1) SUMO is matured by Ub-like specific protease 1 (Ulp1) or human sentrin-specific protease 1 (SENP1). This proteolytic cleavage exposes the C-terminal GG motif required for the activation step. (2) SUMO is activated by a heterodimer composed with SAE1/SAE2 (Aos1/Uba2) to form the SUMO (E1/E2)-activating enzyme. Heterodimer is bound via a thioester bond between the C-terminal G residue of SUMO and the catalytic C of SAE2. (3) SUMO is transferred to the catalytic C of SUMO-conjugating enzyme E2 (or Ubc9) by a transesterification reaction. (4) SUMO is bound to the target protein by Ubc9 in association with SUMO E3 ligase. Ubc9 forms an amide bond between the SUMO C-terminus and ε-amino groups of the acceptor L residues in the target protein. (5) These reactions are reversible by means of the Ulp or SENP proteases.

The p6 domain of HIV-1 Pr55<sup>Gag</sup> is sumoylated by SUMO-1, which covalently links K475 in the consensus sequence ( $\psi$ KXE: QKQE) (Table 5). The K475R substitution partially inhibits binding of the precursor to the SUMO-conjugating enzyme E2 (Ubc9) [162], suggesting that more than one Pr55<sup>Gag</sup> domain could be involved in the recruitment of Ubc9 [162,163]. It was proposed that SUMO-Ubc9 could be involved in intracellular trafficking of Pr55<sup>Gag</sup> [164]. Indeed, after translation, the first trafficking complex intermediate observed in the perinuclear region is composed of Pr55<sup>Gag</sup>, kinesin family motor 4 (KIF4), Ubc9, and SUMO-1 [164]. In contrast, other studies suggested that the recruitment of Ubc9 would be required for the late stages of viral replication, thus participating to Env incorporation into viral particles [163]. Moreover, the overexpression of SUMO-1 was observed to globally decrease viral infectivity, and sumoylation could be then involved in the negative regulation of viral replication [162]. Interestingly, sumoylation and ubiquitination co-regulate each other [165], and sumoylation and mono-ubiquitination of p6 were both found to occur on K475. It is thus possible that SUMO-1 interaction with p6 protects Pr55<sup>Gag</sup> from proteasomal degradation [162]. Overproduction of SUMO-1 should have no direct effect on viral assembly, but if sumoylation competes with ubiquitination, subsequent decrease of Tsg101 recruitment could produce a negative effect on budding [162].

### 10. Gag Sumoylation in Other Retroviruses

As for HIV-1, other retroviruses are sumoylated (Table 5); however, the impact of sumoylation is not yet fully elucidated. The Ubc9 factor was found to interact with MLV and MPMV Gag proteins [166,167]. Similarly, to HIV-1, in MPMV this factor was suggested to be involved in the trafficking of Pr78<sup>Gag</sup> to the PM [167]. Besides, the CA domain of MoMuLV Gag was shown to interact not only with Ubc9 [168], but also with PIASy, a SUMO E3 ligase [168]. These interactions, leading to CA sumoylation during the early stages of the viral life cycle after reverse transcription, might have a role in viral replication [168]. Single K substitutions have generally no effect on the viral cycle, suggesting redundancy between sumoylable positions. On the other hand, the modification of K218 with an R residue was found to reduce the overall viral replication, without affecting the overall SUMO-1 rate on Gag [168].

In the EIAV p9 domain of Gag, K465 is the main target for sumoylation, and mutational experiments showed that this PTM is involved in the regulation of viral replication and infectivity [144,169]. Moreover, sumoylation of K465 seems to regulate the sumoylation of other K residues in different domains of the precursor, such as the MA, the CA and NC (Table 5) [169]. However, a specific role of all those PTMs in viral replication remains to be clarified. Finally, K244 in CA of RSV Pr76<sup>Gag</sup> was found to be sumoylated, and its substitution with a non-sumoylable R residue (K244R) displayed decreased viral infectivity [142].

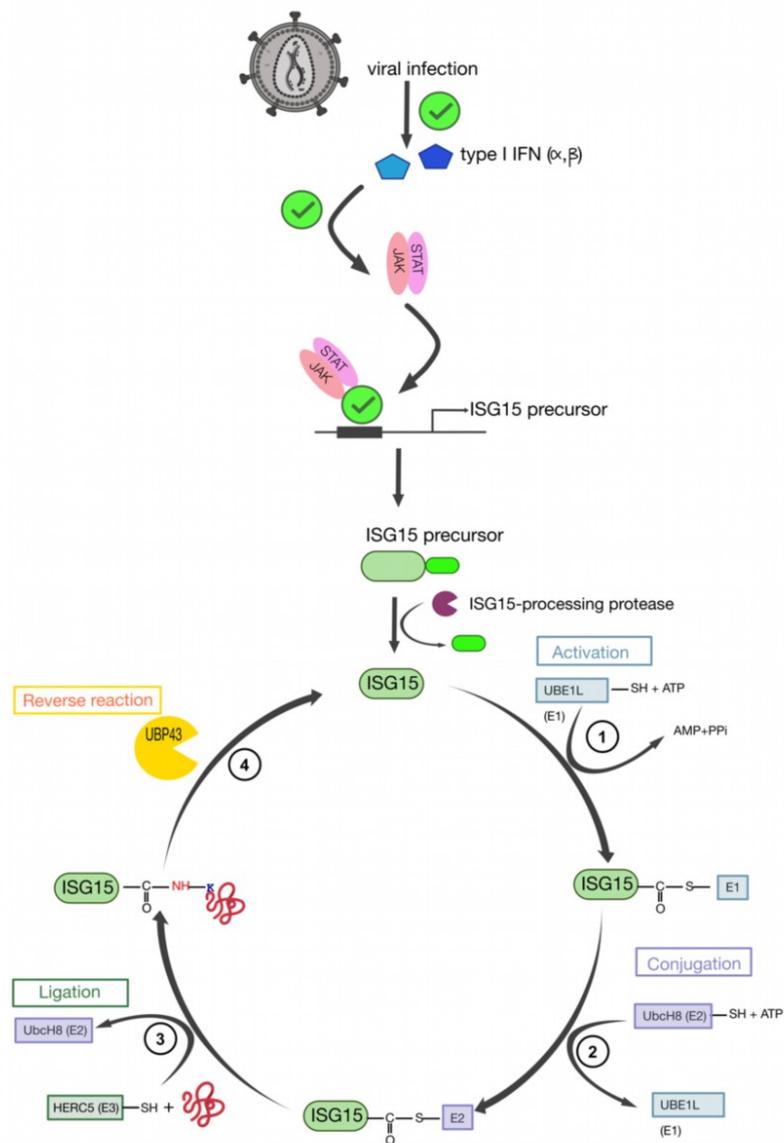
Similarly, to ubiquitination, the exact role of sumoylation is still a matter of debate and might be different among retroviruses. Moreover, sumoylation is still very difficult to detect, and thus further technological advances will be required to better identify and characterize this PTM.

**Table 5.** Summary of sumoylated positions in retroviral Gag proteins.

Retrovirus	Domains	Residues	Associated (or Proposed) Roles	References
HIV-1	p6		Sumoylation and ubiquitination co-regulate each other	[165]
			More than one domain should be involved in Ubc9 recruitment	[162,163]
			SUMO-Ubc9 could be involved in intracellular trafficking of Pr55 <sup>Gag</sup> - in the early phase: perinuclear region or - in the late stages of replication: potential role in Env incorporation	[163,164]
		K 475	Sumoylation could be then involved in the negative regulation of viral replication	[162]
			Belongs to QKQE consensus sequence Sumoylation and mono-ubiquitination of p6 can both occur on K475	[162]
MoMuLV	CA		CA domain of MLV Gag interacts with Ubc9 and with PIASy	[168]
MPMV			Recruitment of Ubc9 involved in the active transport of MPMV Pr78 <sup>Gag</sup> to the PM	[167]
RSV	CA	K 244	Its substitution with non sumoylable R reduces the overall viral infectivity	[142]
EIAV	MA	K 13	Targets of sumoylation	[144,169]
		K 86		
		K 107		
	CA	K 282		
		K 297		
	NC	K 368		
		K 373		
		K 388		
		K 420		
		K 423		
p9	K 465	Constitutes the main target for sumoylation	[144,169]	

### 11. Retroviral Gag Protein ISGylation

Besides sumoylation, there exists another Ub-like protein, which is the interferon stimulated gene 15 (ISG15) (for reviews see [161,170]). ISG15 was identified in mammals and corresponds to a 17 kDa protein induced by type I IFN ( $\alpha$  and  $\beta$ ) ([171,172], for a review see [173]) (Figure 8) that contributes to regulation of the cell cycle, and plays a role in stress response, signaling transduction, and immune response. The IFN response starts with the binding of type I IFN to cell specific receptors, leading to the activation of the Janus kinase (JAK), the signal transducers and the activators of transcription (STAT) signaling pathway (for a review see [174]), which stimulate the transcription of several hundreds of ISGs, including ISG15. This last one presents a sequence homology with Ub, as it contains two Ub-like domains (Figure 6b). The cycle of ISGylation is comparable with the one of Ub (Figure 8). Indeed, three distinct biochemical reactions leads the binding between ISG15 and the target protein. This reaction is reversible and the Ub specific peptidase 18 (USP18), also called UBP43, is involved in the reverse reaction, thus acting as an ISG15 deconjugating enzyme [175].

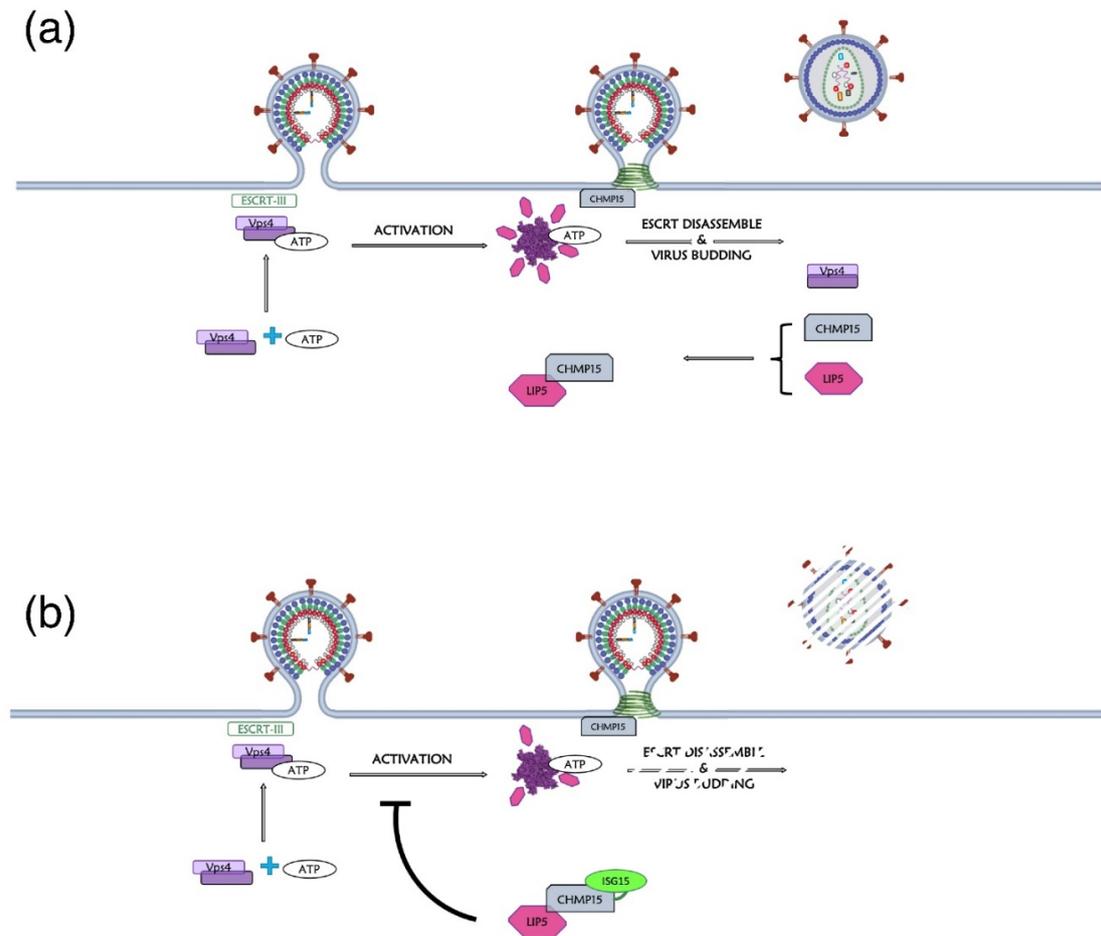


**Figure 8.** The cycle of ISG15. Viral infection induces the expression of type I IFN. These molecules activate the JAK/STAT signaling pathway, which is responsible for the activation of the ISG15

promoter (for reviews see [161,172]). The ISG15 is matured by ISG15-specific proteases which cleave the C-terminal extension from ISG15 precursor. (1) The mature ISG15 is activated by UBE1L (E1). It corresponds to the formation of a thioester bond between ISG15 and E1. (2) ISG15 linked to UBE1L is transferred to UbcH8 (E2). (3) Finally, E2 recruits an E3 ligase such as HERC5, transferring the activated Ub from the E2 to the K substrate (ligation reaction). (4) The reaction can be reversed via UBP43. Indeed, it cleaves ISG15 molecules that are conjugated to the substrate proteins via isopeptide bonds (adapted from [161]).

Interestingly, recent studies on HIV-1 gave first information about the role of ISG15 in viral replication. Indeed, suppression of IFN produced by dendritic cell (pDC2) induces the rapid progression of viral infection [176], thus displaying an antiviral role for ISG15. Moreover, *in vitro* studies showed that ISG15 would inhibit not only the early [177,178] but also the late steps [179,180] of viral replication. The co-transfection with plasmids expressing ISG15 and HIV-1, inhibited the release of viruses, while no impact was seen on HIV-1 proteins production [181]. Moreover, the overexpression of ISG15 and of the activating enzyme UBE1 was observed to impair HIV-1 replication [181]. Interestingly, in these assays, the ubiquitination of Pr55<sup>Gag</sup> and Tsg101 also resulted affected, with the consequent abolition of the interactions between the p6 late domain and Tsg101. This impairs the assembly of viral particles and their release, and EM assays showed that immature viruses accumulate at the PM [180,181]. Furthermore, the production of viral particles was also found to be impaired by the E3 ligase HERC5. Interestingly, this inhibition was not found to alter the trafficking of HIV-1 Gag to the PM, but the budding at the PM [180]. Since HERC5 is described to restrict also MLV Gag particle production, it results that HERC5 and more generally, the response ISG15, can be considered as a restrictive factor against retrovirus [180].

In general, ISG15 may affect many other RNA viruses and other retroviruses such as the avian sarcoma leukosis virus (ASLV) (for review see [182]). Similar to HIV-1, ISG15 inhibits the release of ASLV, and the ubiquitination of ASLV Gag precursor. In this context the E3 ubiquitin-protein ligase, NEDD4 was found to maintain its ability to bind the late motif in ASLV Gag [179]. Thus, ISG15 does not seem to prevent directly the interaction between NEDD4 and ASLV Gag, but likely it interferes with the Ub ligase activity of NEDD4, which inhibits ubiquitination [183] even if the precise mechanism remains unclear. Alternatively, it was proposed that the impaired budding of ASLV and HIV-1 could be due to the ISGylation of the ESCRT-III component Charged Multivesicular body Protein 5 (CHMP5) (Figure 9a,b). Indeed, during retroviral budding, the ESCRT-III complex polymerizes at PM in interaction with Pr55<sup>Gag</sup>. This complex then recruits the inactive dimer form of the ATPase Vps4, which requires it to recruit ATP and its coactivator protein, Vps-associated protein LIP5, to achieve its activated double hexameric-ring structure. This leads to the disassembly of the ESCRT-III complex, thus promoting the viral budding (Figure 9a). According to the proposed model, ISGylation of CHMP5 was found to impair Vps4 binding to LIP5, and thus Vps4 would remain in its inactivated conformation while ESCRT-III complexes would be trapped at the PM, thus blocking virus budding [179] (Figure 9b).



**Figure 9.** Model of the impact of ISG15 on Vps4 function during retroviral budding. (a) Normal assembly and budding phase during the retroviral cycle. Vps4 activity depends on its oligomeric state. In its dimeric form, Vps4 is cytosolic and inactive. During retrovirus assembly at PM, upon polymerization of the ESCRT-III complex with the p6 domain of HIV-1 Pr55<sup>Gag</sup>, ATP-bound Vps4 is recruited at the PM. At this step, Vps4 interacts with the coactivator protein LIP5, which is bound to CHMP5, and achieves its double hexameric-ring structure. Then, ATP hydrolysis by the Vps4-LIP5 oligomer releases the ESCRT-III complexes from PM and the dissociation of the ESCRT complex coincides with the membrane fission event that releases retrovirus particles. (b) ISG15 inhibits the budding phase. When CHMP5 is ISGylated, this prevents the interaction between Vps4 and LIP5 by excluding LIP5. In the absence of the Vps4-LIP5 complex, the ESCRT-III complex remains trapped at the PM and the viral release is thus impaired (adapted from [179]).

Similar mechanisms can occur in the context of other retroviruses [179].

## 12. Post-Translational Methylation of Retroviral Gag Proteins

Finally, retroviral Gag precursors are also subjected to methylation. This covalent PTM is reversible [184]. It consists of the transfer of a methyl group from a donor, the S-adenosylmethionine (SAM), to a target K residue, and this reaction is catalyzed by the K-methyltransferases (for reviews see [185,186]). The same residue can be mono-, di- or trimethylated, thus conferring a signature which can be specifically recognized by transacting factors named “readers”, whose recruitment can promote signaling pathways, regulation of protein–protein interactions, transcription, T-cells activation [187] and subcellular localization [188]. Interestingly, immunoblotting of the CA domain of HIV-1 in presence of AdOx, an inhibitor of methylation, showed an increase of mature CA, suggesting that methylation of HIV-1 Pr55<sup>Gag</sup> could affect proteolytic maturation and likely the viral assembly [189]. The basic region of the NC domain of HIV-1 Pr55<sup>Gag</sup> is also methylated,

and this modification was proposed to be involved in the subnuclear localization of the precursor [190]. This same PMT on NC would also decrease the rate of tRNA<sup>Lys3</sup> annealing to the PBS region in gRNA, thus inducing defects in the reverse transcription step [191]. Similarly, the R540 residue in the NC domain of PFV Pr74<sup>Gag</sup> was found to be methylated, and this modification seems to be required for the subnuclear localization of the precursor [192].

### 13. Conclusions

PTMs create a vast diversity in proteins and thus regulate their functions. Globally, PTMs play a role in many processes such as cell signaling, and protein–protein and protein–RNA interactions. Besides, PTMs are crucial for the life cycle of many viruses and the characterization of viral PTMs would provide a better understanding of the mechanisms of viral processes. HIV-1 Pr55<sup>Gag</sup>, as many other retroviral Gag precursors, displays several PTMs in its different domains (Figure 10). These PTMs include myristoylation, phosphorylation, ubiquitination, sumoylation, and methylation. All these PTMs can have either antagonistic or cooperative roles, thus allowing fine regulation of the viral cycle. However, up to now, the role of many of these modifications is not fully elucidated and further investigations will be required to better understand their contributions in the viral life cycle. One of the main challenges to study PTMs carried by proteins consists of the development of refined proteomic technologies, allowing the specific detection and characterization of the modifications. The improved knowledge of those regulations would be useful in the future to identify new targets for antiretroviral treatments.



**Figure 10.** Summary of post-translational modifications of HIV-1 Pr55<sup>Gag</sup> residues. The domains of Pr55<sup>Gag</sup> are represented by different colors (see Figure 4). Experimentally identified modified residues are highlighted: myristoylation (pink), phosphorylation (yellow), ubiquitination (light blue), potential ubiquitinations (light green), and sumoylation (black).

**Author Contributions:** S.B. supervised the project; C.B. wrote the manuscript with contributions from S.B., J.-C.P. and R.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Agence Nationale de Recherche sur le SIDA et les hépatites virales (ANRS), SIDACTION and IdEx (Initiative d’Excellence, Université de Strasbourg, France). CB is supported by a fellowship from the French ministry of Research and Higher Education.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** No applicable.

**Data Availability Statement:** All figures and discussed literature are available in the main text of this review.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Audagnotto, M.; Peraro, M.D. Protein post-translational modifications: In silico prediction tools and molecular modeling. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 307–319, doi:10.1016/j.csbj.2017.03.004.
2. Spoel, S.H. Orchestrating the proteome with post-translational modifications. *J. Exp. Bot.* **2018**, *69*, 4499–4503, doi:10.1093/jxb/ery295.
3. Biard-Piechaczyk, M.; Borel, S.; Espert, L.; De Bettignies, G.; Coux, O. HIV-1, ubiquitin and ubiquitin-like proteins: The dialectic interactions of a virus with a sophisticated network of post-translational modifications. *Biol. Cell* **2012**, *104*, 165–187, doi:10.1111/boc.201100112.
4. Chen, L.; Keppler, O.T.; Schölz, C. Post-translational Modification-Based Regulation of HIV Replication. *Front. Microbiol.* **2018**, *9*, 2131, doi:10.3389/fmicb.2018.02131.
5. Lu, K.; Heng, X.; Garyu, L.; Monti, S.; Garcia, E.L.; Kharytonchyk, S.; Dorjsuren, B.; Kulandaivel, G.; Jones, S.; Hiremath, A.; et al. NMR Detection of Structures in the HIV-15'-Leader RNA That Regulate Genome Packaging. *Science* **2011**, *334*, 242–245, doi:10.1126/science.1210460.
6. Wilkinson, A.K.; Gorelick, R.J.; Vasa, S.M.; Guex, N.; Rein, A.; Mathews, D.H.; Giddings, M.C.; Weeks, K.M. High-Throughput SHAPE Analysis Reveals Structures in HIV-1 Genomic RNA Strongly Conserved across Distinct Biological States. *PLoS Biol.* **2008**, *6*, 96, doi:10.1371/journal.pbio.0060096.
7. El-Wahab, E.W.A.; Smyth, R.P.; Mailler, E.; Bernacchi, S.; Vivet-Boudou, V.; Hijnen, M.; Jossinet, F.; Mak, J.; Paillart, J.-C.; Marquet, R. Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat. Commun.* **2014**, *5*, 4304, doi:10.1038/ncomms5304.
8. Smyth, R.P.; Despons, L.; Huili, G.; Bernacchi, S.; Hijnen, M.; Mak, J.; Jossinet, F.; Weixi, L.; Paillart, J.-C.; Von Kleist, M.; et al. Mutational interference mapping experiment (MIME) for studying RNA structure and function. *Nat. Methods* **2015**, *12*, 866–872, doi:10.1038/nmeth.3490.
9. Bernacchi, S.; El-Wahab, E.W.A.; Dubois, N.; Hijnen, M.; Smyth, R.P.; Mak, J.; Marquet, R.; Paillart, J.-C. HIV-1 Pr55Gag binds genomic and spliced RNAs with different affinity and stoichiometry. *RNA Biol.* **2017**, *14*, 90–103, doi:10.1080/15476286.2016.1256533.
10. Ferrer, M.; Clerté, C.; Chamontin, C.; Basyuk, E.; Lainé, S.; Hottin, J.; Bertrand, E.; Margeat, E.; Mougél, M. Imaging HIV-1 RNA dimerization in cells by multicolor super-resolution and fluctuation microscopies. *Nucleic Acids Res.* **2016**, *44*, 7922–7934, doi:10.1093/nar/gkw511.
11. Jouvenet, N.; Simon, S.M.; Bieniasz, P.D. Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19114–19119, doi:10.1073/pnas.0907364106.
12. Kutluay, S.B.; Bieniasz, P.D. Analysis of the Initiating Events in HIV-1 Particle Assembly and Genome Packaging. *PLoS Pathog.* **2010**, *6*, e1001200, doi:10.1371/journal.ppat.1001200.
13. Mailler, E.; Bernacchi, S.; Marquet, R.; Paillart, J.-C.; Vivet-Boudou, V.; Smyth, R.P. The Life-Cycle of the HIV-1 Gag–RNA Complex. *Viruses* **2016**, *8*, 248, doi:10.3390/v8090248.
14. Bieniasz, P.; Telesnitsky, A. Multiple, Switchable Protein:RNA Interactions Regulate Human Immunodeficiency Virus Type 1 Assembly. *Annu. Rev. Virol.* **2018**, *5*, 165–183, doi:10.1146/annurev-virology-092917-043448.
15. Comas-Garcia, M.; Davis, S.R.; Rein, A. On the Selective Packaging of Genomic RNA by HIV-1. *Viruses* **2016**, *8*, 246, doi:10.3390/v8090246.
16. Bell, N.M.; Lever, A.M. HIV Gag polyprotein: Processing and early viral particle assembly. *Trends Microbiol.* **2013**, *21*, 136–144, doi:10.1016/j.tim.2012.11.006.
17. Olety, B.; Ono, A. Roles played by acidic lipids in HIV-1 Gag membrane binding. *Virus Res.* **2014**, *193*, 108–115, doi:10.1016/j.virusres.2014.06.015.
18. Kutluay, S.B.; Zang, T.; Blanco-Melo, D.; Powell, C.; Jannain, D.; Errando, M.; Bieniasz, P.D. Global Changes in the RNA Binding Specificity of HIV-1 Gag Regulate Virion Genesis. *Cell* **2014**, *159*, 1096–1109, doi:10.1016/j.cell.2014.09.057.
19. Gaines, C.R.; Tkacik, E.; Rivera-Oven, A.; Somani, P.; Achimovich, A.; Alabi, T.; Zhu, A.; Getachew, N.; Yang, A.L.; McDonough, M.; et al. HIV-1 Matrix Protein Interactions with tRNA: Implications for Membrane Targeting. *J. Mol. Biol.* **2018**, *430*, 2113–2127, doi:10.1016/j.jmb.2018.04.042.
20. Thornhill, D.; Olety, B.; Ono, A. Relationships between MA-RNA Binding in Cells and Suppression of HIV-1 Gag Mislocalization to Intracellular Membranes. *J. Virol.* **2019**, *93*, 23, doi:10.1128/jvi.00756-19.
21. Zhao, G.; Perilla, J.R.; Yufenyuy, E.L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A.M.; Schulten, K.; Aiken, C.; et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nat. Cell Biol.* **2013**, *497*, 643–646, doi:10.1038/nature12162.
22. Schur, F.K.M.; Hagen, W.J.H.; Rumlová, M.; Ruml, T.; Müller, B.; Kräusslich, H.-G.; Briggs, J.A.G. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nat. Cell Biol.* **2015**, *517*, 505–508, doi:10.1038/nature13838.

23. Pornillos, O.; Ganser-Pornillos, B.K.; Kelly, B.N.; Hua, Y.; Whitby, F.G.; Stout, C.D.; Sundquist, W.I.; Hill, C.P.; Yeager, M. X-Ray Structures of the Hexameric Building Block of the HIV Capsid. *Cell* **2009**, *137*, 1282–1292, doi:10.1016/j.cell.2009.04.063.
24. Dannull, J.; Surovoy, A.; Jung, G.; Moelling, K. Specific binding of HIV-1 nucleocapsid protein to PSI RNA in vitro requires N-terminal zinc finger and flanking basic amino acid residues. *EMBO J.* **1994**, *13*, 1525–1533, doi:10.1002/j.1460-2075.1994.tb06414.x.
25. Webb, J.A.; Jones, C.P.; Parent, L.J.; Rouzina, I.; Musier-Forsyth, K. Distinct binding interactions of HIV-1 Gag to Psi and non-Psi RNAs: Implications for viral genomic RNA packaging. *RNA* **2013**, *19*, 1078–1088, doi:10.1261/rna.038869.113.
26. Cimarelli, A.; Sandin, S.; Höglund, S.; Luban, J. Basic Residues in Human Immunodeficiency Virus Type 1 Nucleocapsid Promote Virion Assembly via Interaction with RNA. *J. Virol.* **2000**, *74*, 3046–3057, doi:10.1128/jvi.74.7.3046-3057.2000.
27. Ott, D.E.; Coren, L.V.; Shatzer, T. The Nucleocapsid Region of Human Immunodeficiency Virus Type 1 Gag Assists in the Coordination of Assembly and Gag Processing: Role for RNA-Gag Binding in the Early Stages of Assembly. *J. Virol.* **2009**, *83*, 7718–7727, doi:10.1128/jvi.00099-09.
28. El Meshri, S.E.; Dujardin, D.; Godet, J.; Richert, L.; Boudier, C.; Darlix, J.L.; Didier, P.; Mély, Y.; De Rocquigny, H. Role of the Nucleocapsid Domain in HIV-1 Gag Oligomerization and Trafficking to the Plasma Membrane: A Fluorescence Lifetime Imaging Microscopy Investigation. *J. Mol. Biol.* **2015**, *427*, 1480–1494, doi:10.1016/j.jmb.2015.01.015.
29. Dubois, N.; Khoo, K.K.; Ghossein, S.; Seissler, T.; Wolff, P.; McKinstry, W.J.; Mak, J.; Paillart, J.-C.; Marquet, R.; Bernacchi, S. The C-terminal p6 domain of the HIV-1 Pr55Gag precursor is required for specific binding to the genomic RNA. *RNA Biol.* **2018**, *15*, 923–936, doi:10.1080/15476286.2018.1481696.
30. Farazi, T.A.; Waksman, G.; Gordon, J.I. The Biology and Enzymology of Protein N-Myristoylation. *J. Biol. Chem.* **2001**, *276*, 39501–39504, doi:10.1074/jbc.r100042200.
31. Martin, D.D.; Beauchamp, E.; Berthiaume, L.G. Post-translational myristoylation: Fat matters in cellular life and death. *Biochimie* **2011**, *93*, 18–31, doi:10.1016/j.biochi.2010.10.018.
32. Dyda, F.; Klein, D.C.; Hickman, A.B. GCN5-Related N-Acetyltransferases: A Structural Overview. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 81–103, doi:10.1146/annurev.biophys.29.1.81.
33. Johnson, D.R.; Bhatnagar, R.S.; Knoll, L.J.; I. Gordon, J. Genetic and Biochemical Studies of Protein N-Myristoylation. *Annu. Rev. Biochem.* **1994**, *63*, 869–914, doi:10.1146/annurev.bi.63.070194.004253.
34. Cao, W.; Sumikoshi, K.; Nakamura, S.; Terada, T.; Shimizu, K. Prediction of N-myristoylation modification of proteins by SVM. *Bioinformatics* **2011**, *6*, 204–206, doi:10.6026/97320630006204.
35. Maurer-Stroh, S.; Eisenhaber, B.; Eisenhaber, F. N-terminal N-myristoylation of proteins: Refinement of the sequence motif and its taxon-specific differences 1 Edited by J. Thornton. *J. Mol. Biol.* **2002**, *317*, 523–540, doi:10.1006/jmbi.2002.5425.
36. Tang, C.; Loeliger, E.; Luncsford, P.; Kinde, I.; Beckett, D.; Summers, M.F. Entropic switch regulates myristate exposure in the HIV-1 matrix protein. *Proc. Natl. Acad. Sci. USA* **2003**, *101*, 517–522, doi:10.1073/pnas.0305665101.
37. Hogue, I.B.; Llewellyn, G.N.; Ono, A. Dynamic Association between HIV-1 Gag and Membrane Domains. *Mol. Biol. Int.* **2012**, *2012*, 979765, doi:10.1155/2012/979765.
38. Inlora, J.; Collins, D.R.; Trubin, M.E.; Chung, J.Y.J.; Ono, A. Membrane Binding and Subcellular Localization of Retroviral Gag Proteins Are Differentially Regulated by MA Interactions with Phosphatidylinositol-(4,5)-Bisphosphate and RNA. *mBio* **2014**, *5*, e02202-14, doi:10.1128/mbio.02202-14.
39. Resh, M.D. A myristoyl switch regulates membrane binding of HIV-1 Gag. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 417–418, doi:10.1073/pnas.0308043101.
40. Ono, A. Relationships between plasma membrane microdomains and HIV-1 assembly. *Biol. Cell* **2010**, *102*, 335–350, doi:10.1042/bc20090165.
41. Saad, J.S.; Miller, J.; Tai, J.; Kim, A.; Ghanam, R.H.; Summers, M.F. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11364–11369, doi:10.1073/pnas.0602818103.
42. Saad, J.S.; Loeliger, E.; Luncsford, P.; Liriano, M.; Tai, J.; Kim, A.; Miller, J.; Joshi, A.; Freed, E.O.; Summers, M.F. Point Mutations in the HIV-1 Matrix Protein Turn Off the Myristyl Switch. *J. Mol. Biol.* **2007**, *366*, 574–585, doi:10.1016/j.jmb.2006.11.068.
43. Dalton, A.K.; Ako-Adjei, D.; Murray, P.S.; Murray, D.; Vogt, V.M. Electrostatic Interactions Drive Membrane Association of the Human Immunodeficiency Virus Type 1 Gag MA Domain. *J. Virol.* **2007**, *81*, 6434–6445, doi:10.1128/jvi.02757-06.
44. Paillart, J.-C.; Göttlinger, H.G. Opposing Effects of Human Immunodeficiency Virus Type 1 Matrix Mutations Support a Myristyl Switch Model of Gag Membrane Targeting. *J. Virol.* **1999**, *73*, 2604–2612, doi:10.1128/jvi.73.4.2604-2612.1999.
45. Zhou, W.; Resh, M.D. Differential membrane binding of the human immunodeficiency virus type 1 matrix protein. *J. Virol.* **1996**, *70*, 8540–8548, doi:10.1128/jvi.70.12.8540-8548.1996.
46. Murray, P.S.; Li, Z.; Wang, J.; Tang, C.L.; Honig, B.; Murray, D. Retroviral Matrix Domains Share Electrostatic Homology: Models for Membrane Binding Function throughout the Viral Life Cycle. *Structure* **2005**, *13*, 1521–1531, doi:10.1016/j.str.2005.07.010.
47. Hamard-Peron, E.; Muriaux, D. Retroviral matrix and lipids, the intimate interaction. *Retrovirology* **2011**, *8*, 15, doi:10.1186/1742-4690-8-15.
48. Parent, L.J.; Gudleski, N. Beyond Plasma Membrane Targeting: Role of the MA domain of Gag in Retroviral Genome Encapsulation. *J. Mol. Biol.* **2011**, *410*, 553–564, doi:10.1016/j.jmb.2011.04.072.
49. Murray, D.; Ben-Tal, N.; Honig, B.; McLaughlin, S. Electrostatic interaction of myristoylated proteins with membranes: Simple physics, complicated biology. *Structure* **1997**, *5*, 985–989, doi:10.1016/s0969-2126(97)00251-7.
50. Kozak, C.A. Origins of the Endogenous and Infectious Laboratory Mouse Gammaretroviruses. *Viruses* **2014**, *7*, 1–26, doi:10.3390/v7010001.

51. Nitta, T.; Kuznetsov, Y.; McPherson, A.; Fan, H. Murine leukemia virus glycosylated Gag (gPr80gag) facilitates interferon-sensitive virus release through lipid rafts. *Proc. Natl. Acad. Sci. USA* **2009**, *107*, 1190–1195, doi:10.1073/pnas.0908660107.
52. Houzet, L.; Gay, B.; Morichaud, Z.; Briant, L.; Mougel, M. Intracellular assembly and budding of the Murine Leukemia Virus in infected cells. *Retrovirology* **2006**, *3*, 12, doi:10.1186/1742-4690-3-12.
53. Hron, T.; Elleder, D.; Gifford, R.J. Deltaretroviruses have circulated since at least the Paleogene and infected a broad range of mammalian species. *Retrovirology* **2019**, *16*, 1–7, doi:10.1186/s12977-019-0495-9.
54. Egessain, A.; Ecassar, O. Epidemiological Aspects and World Distribution of HTLV-1 Infection. *Front. Microbiol.* **2012**, *3*, 388, doi:10.3389/fmicb.2012.00388.
55. Barez, P.-Y.; De Brogniez, A.; Carpentier, A.; Gazon, H.; Gillet, N.; Gutiérrez, G.; Hamaidia, M.; Jacques, J.-R.; Perike, S.; Sriramareddy, S.N.; et al. Recent Advances in BLV Research. *Viruses* **2015**, *7*, 6080–6088, doi:10.3390/v7112929.
56. Blot, V.; Perugi, F.; Gay, B.; Prévost, M.-C.; Briant, L.; Tangy, F.; Abriel, H.; Staub, O.; Dokh elar, M.-C.; Pique, C. Nedd4.1-mediated ubiquitination and subsequent recruitment of Tsg101 ensure HTLV-1 Gag trafficking towards the multivesicular body pathway prior to virus budding. *J. Cell Sci.* **2004**, *117*, 2357–2367, doi:10.1242/jcs.01095.
57. Inlora, J.; Chukkapalli, V.; Derse, D.; Ono, A. Gag Localization and Virus-Like Particle Release Mediated by the Matrix Domain of Human T-Lymphotropic Virus Type 1 Gag Are Less Dependent on Phosphatidylinositol-(4,5)-Bisphosphate than Those Mediated by the Matrix Domain of HIV-1 Gag. *J. Virol.* **2011**, *85*, 3802–3810, doi:10.1128/jvi.02383-10.
58. Le Blanc, I.; Rosenberg, A.R.; Dokh elar, M.-C. Multiple Functions for the Basic Amino Acids of the Human T-Cell Leukemia Virus Type 1 Matrix Protein in Viral Transmission. *J. Virol.* **1999**, *73*, 1860–1867, doi:10.1128/jvi.73.3.1860-1867.1999.
59. Fogarty, K.H.; Zhang, W.; Grigsby, I.F.; Johnson, J.L.; Chen, Y.; Mueller, J.D.; Mansky, L.M. New Insights into HTLV-1 Particle Structure, Assembly, and Gag-Gag Interactions in Living Cells. *Viruses* **2011**, *3*, 770–793, doi:10.3390/v3060770.
60. Yeung, T.; Gilbert, G.E.; Shi, J.; Silvius, J.; Kapus, A.; Grinstein, S. Membrane Phosphatidylserine Regulates Surface Charge and Protein Localization. *Science* **2008**, *319*, 210–213, doi:10.1126/science.1152066.
61. Chukkapalli, V.; Oh, S.J.; Ono, A. Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1600–1605, doi:10.1073/pnas.0908661107.
62. Wang, H.; Norris, K.M.; Mansky, L.M. Involvement of the Matrix and Nucleocapsid Domains of the Bovine Leukemia Virus Gag Polyprotein Precursor in Viral RNA Packaging. *J. Virol.* **2003**, *77*, 9431–9438, doi:10.1128/jvi.77.17.9431-9438.2003.
63. Prchal, J.; Srb, P.; Hunter, E.; Ruml, T.; Hrabal, R. The Structure of Myristoylated Mason-Pfizer Monkey Virus Matrix Protein and the Role of Phosphatidylinositol-(4,5)-Bisphosphate in Its Membrane Binding. *J. Mol. Biol.* **2012**, *423*, 427–438, doi:10.1016/j.jmb.2012.07.021.
64. Ross, S.R. Mouse Mammary Tumor Virus Molecular Biology and Oncogenesis. *Viruses* **2010**, *2*, 2000–2012, doi:10.3390/v2092000.
65. Hayward, A.J.; Tachedjian, M.; Cui, J.; Field, H.; Holmes, E.C.; Wang, L.-F.; Tachedjian, G. Identification of diverse full-length endogenous betaretroviruses in megabats and microbats. *Retrovirology* **2013**, *10*, 35, doi:10.1186/1742-4690-10-35.
66. Monde, K.; Contreras-Galindo, R.; Kaplan, M.H.; Markovitz, D.M.; Ono, A. Human Endogenous Retrovirus K Gag Coassembles with HIV-1 Gag and Reduces the Release Efficiency and Infectivity of HIV-1. *J. Virol.* **2012**, *86*, 11194–11208, doi:10.1128/jvi.00301-12.
67. Dole al, M.; Z abransk y, A.; Dost al, J.; Van ek, O.; Brynda, J.; Lep s ik, M.; Hadravov a, R.; Pichov a, I. Myristoylation drives dimerization of matrix protein from mouse mammary tumor virus. *Retrovirology* **2016**, *13*, 1–15, doi:10.1186/s12977-015-0235-8.
68. Dalton, A.K.; Murray, P.S.; Murray, D.; Vogt, V.M. Biochemical Characterization of Rous Sarcoma Virus MA Protein Interaction with Membranes. *J. Virol.* **2005**, *79*, 6227–6238, doi:10.1128/jvi.79.10.6227-6238.2005.
69. Nadaraia-Hoke, S.; Bann, D.V.; Lochmann, T.L.; Gudleski-O'Regan, N.; Parent, L.J. Alterations in the MA and NC Domains Modulate Phosphoinositide-Dependent Plasma Membrane Localization of the Rous Sarcoma Virus Gag Protein. *J. Virol.* **2013**, *87*, 3609–3615, doi:10.1128/jvi.03059-12.
70. Callahan, E.M.; Wills, J.W. Repositioning Basic Residues in the M Domain of the Rous Sarcoma Virus Gag Protein. *J. Virol.* **2000**, *74*, 11222–11229, doi:10.1128/jvi.74.23.11222-11229.2000.
71. Dick, R.A.; Vogt, V.M. Membrane interaction of retroviral Gag proteins. *Front. Microbiol.* **2014**, *5*, 187, doi:10.3389/fmicb.2014.00187.
72. Hatanaka, H.; Iourin, O.; Rao, Z.; Fry, E.; Kingsman, A.; Stuart, D.I. Structure of Equine Infectious Anemia Virus Matrix Protein. *J. Virol.* **2002**, *76*, 1876–1883, doi:10.1128/jvi.76.4.1876-1883.2002.
73. Zhang, Z.; Ma, J.; Zhang, X.; Su, C.; Yao, Q.-C.; Wang, X. Equine Infectious Anemia Virus Gag Assembly and Export Are Directed by Matrix Protein throughtrans-Golgi Networks and Cellular Vesicles. *J. Virol.* **2015**, *90*, 1824–1838, doi:10.1128/jvi.02814-15.
74. Fernandes, F.; Chen, K.; Ehrlich, L.S.; Jin, J.; Chen, M.H.; Medina, G.N.; Symons, M.; Montelaro, R.; Donaldson, J.; Tjandra, N.; et al. Phosphoinositides Direct Equine Infectious Anemia Virus Gag Trafficking and Release. *Traffic* **2010**, *12*, 438–451, doi:10.1111/j.1600-0854.2010.01153.x.
75. Goldstone, D.C.; Flower, T.G.; Ball, N.J.; Sanz-Ramos, M.; Yap, M.W.; Ogrodowicz, R.W.; Stanke, N.; Reh, J.; Lindemann, D.; Stoye, J.P.; et al. A Unique Spumavirus Gag N-terminal Domain with Functional Properties of Orthoretroviral Matrix and Capsid. *PLoS Pathog.* **2013**, *9*, e1003376, doi:10.1371/journal.ppat.1003376.
76. Heneine, W.; Schweizer, M.; Sandstrom, P.; Folks, T. Human Infection with Foamy Viruses. *Curr. Top. Microbiol. Immunol.* **2003**, *277*, 181–196, doi:10.1007/978-3-642-55701-9\_8.

77. Shaw, K.L.; Lindemann, D.; Mulligan, M.J.; Goepfert, P.A. Foamy Virus Envelope Glycoprotein Is Sufficient for Particle Budding and Release. *J. Virol.* **2003**, *77*, 2338–2348, doi:10.1128/jvi.77.4.2338-2348.2003.
78. Nishi, H.; Shaytan, A.; Panchenko, A.R. Physicochemical mechanisms of protein regulation by phosphorylation. *Front. Genet.* **2014**, *5*, 270, doi:10.3389/fgene.2014.00270.
79. Mandell, D.J.; Chorny, I.; Groban, E.S.; Wong, S.E.; Levine, E.; Rapp, A.C.S.; Jacobson, M.P. Strengths of Hydrogen Bonds Involving Phosphorylated Amino Acid Side Chains. *J. Am. Chem. Soc.* **2007**, *129*, 820–827, doi:10.1021/ja063019w.
80. Burnett, G.; Kennedy, E.P. The Enzymatic Phosphorylation of Proteins. *J. Biol. Chem.* **1954**, *211*, 969–980, doi:10.1016/s0021-9258(18)71184-8.
81. Francis, A.C.; Di Primio, C.; Allouch, A.; Cereseto, A. Role of phosphorylation in the nuclear biology of HIV-1. *Curr. Med. Chem.* **2011**, *18*, 2904–2912, doi:10.2174/092986711796150478.
82. Burnette, B.; Yu, G.; Felsted, R. Phosphorylation of HIV-1 gag proteins by protein kinase C. *J. Biol. Chem.* **1993**, *268*, 8698–8703, doi:10.1016/s0021-9258(18)52931-8.
83. Yu, G.; Shen, F.S.; Sturch, S.; Aquino, A.; Glazer, R.I.; Felsted, R.L. Regulation of HIV-1 gag Protein Subcellular Targeting by Protein Kinase C. *J. Biol. Chem.* **1995**, *270*, 4792–4796, doi:10.1074/jbc.270.9.4792.
84. Kräusslich, H.G. *Morphogenesis and Maturation of Retroviruses*—Google Livres, 1st ed.; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1996.
85. Kudoh, A.; Takahama, S.; Sawasaki, T.; Ode, H.; Yokoyama, M.; Okayama, A.; Ishikawa, A.; Miyakawa, K.; Matsunaga, S.; Kimura, H.; et al. The phosphorylation of HIV-1 Gag by atypical protein kinase C facilitates viral infectivity by promoting Vpr incorporation into virions. *Retrovirology* **2014**, *11*, 9, doi:10.1186/1742-4690-11-9.
86. Radestock, B.; Morales, I.; Rahman, S.A.; Radau, S.; Glass, B.; Zahedi, R.P.; Müller, B.; Kräusslich, H.-G. Comprehensive Mutational Analysis Reveals p6Gag Phosphorylation to Be Dispensable for HIV-1 Morphogenesis and Replication. *J. Virol.* **2012**, *87*, 724–734, doi:10.1128/jvi.02162-12.
87. Votteler, J.; Neumann, L.; Hahn, S.; Hahn, F.; Rauch, P.; Schmidt, K.; Studtrucker, N.; Solbak, S.M.; Fossen, T.; Henklein, P.; et al. Highly conserved serine residue 40 in HIV-1 p6 regulates capsid processing and virus core assembly. *Retrovirology* **2011**, *8*, 11, doi:10.1186/1742-4690-8-11.
88. Watanabe, S.M.; Chen, M.-H.; Khan, M.; Ehrlich, L.; Kemal, K.S.; Weiser, B.; Shi, B.; Chen, C.; Powell, M.; Anastos, K.; et al. The S40 residue in HIV-1 Gag p6 impacts local and distal budding determinants, revealing additional late domain activities. *Retrovirology* **2013**, *10*, 143, doi:10.1186/1742-4690-10-143.
89. Radestock, B.; Burk, R.; Müller, B.; Kräusslich, H.-G. Re-visiting the functional Relevance of the highly conserved Serine 40 Residue within HIV-1 p6Gag. *Retrovirology* **2014**, *11*, 1–5, doi:10.1186/s12977-014-0114-8.
90. Müller, B.; Patschinsky, T.; Kräusslich, H.-G. The Late-Domain-Containing Protein p6 Is the Predominant Phosphoprotein of Human Immunodeficiency Virus Type 1 Particles. *J. Virol.* **2002**, *76*, 1015–1024, doi:10.1128/jvi.76.3.1015-1024.2002.
91. Hemonnot, B.; Cartier, C.; Gay, B.; Rebuffat, S.; Bardy, M.; Devaux, C.; Boyer, V.; Briant, L. The Host Cell MAP Kinase ERK-2 Regulates Viral Assembly and Release by Phosphorylating the p6 Protein of HIV-1. *J. Biol. Chem.* **2004**, *279*, 32426–32434, doi:10.1074/jbc.m313137200.
92. Wang, Z.; Canagarajah, B.J.; Boehm, J.C.; Kassisà, S.; Cobb, M.H.; Young, P.R.; Abdel-Meguid, S.; Adams, J.L.; Goldsmith, E.J. Structural basis of inhibitor selectivity in MAP kinases. *Structure* **1998**, *6*, 1117–1128, doi:10.1016/s0969-2126(98)00113-0.
93. Cartier, C.; Deckert, M.; Grangeasse, C.; Trauger, R.; Jensen, F.; Bernard, A.; Cozzzone, A.; Desgranges, C.; Boyer, V. Association of ERK2 mitogen-activated protein kinase with human immunodeficiency virus particles. *J. Virol.* **1997**, *71*, 4832–4837, doi:10.1128/jvi.71.6.4832-4837.1997.
94. Cartier, C.; Hemonnot, B.; Gay, B.; Bardy, M.; Sanchiz, C.; Devaux, C.; Briant, L. Active cAMP-dependent Protein Kinase Incorporated within Highly Purified HIV-1 Particles Is Required for Viral Infectivity and Interacts with Viral Capsid Protein. *J. Biol. Chem.* **2003**, *278*, 35211–35219, doi:10.1074/jbc.m301257200.
95. Jacqué, J.; Mann, A.; Enslin, H.; Sharova, N.; Brichtacek, B.; Davis, R.J.; Stevenson, M. Modulation of HIV-1 infectivity by MAPK, a virion-associated kinase. *EMBO J.* **1998**, *17*, 2607–2618, doi:10.1093/emboj/17.9.2607.
96. Dochi, T.; Nakano, T.; Inoue, M.; Takamune, N.; Shoji, S.; Sano, K.; Misumi, S. Phosphorylation of human immunodeficiency virus type 1 capsid protein at serine 16, required for peptidyl-prolyl isomerase-dependent uncoating, is mediated by virion-incorporated extracellular signal-regulated kinase 2. *J. Gen. Virol.* **2014**, *95*, 1156–1166, doi:10.1099/vir.0.060053-0.
97. Yang, X.; Gabuzda, D. Regulation of Human Immunodeficiency Virus Type 1 Infectivity by the ERK Mitogen-Activated Protein Kinase Signaling Pathway. *J. Virol.* **1999**, *73*, 3460–3466, doi:10.1128/jvi.73.4.3460-3466.1999.
98. Greenway, A.; Azad, A.; Mills, J.; McPhee, D. Human immunodeficiency virus type 1 Nef binds directly to Lck and mitogen-activated protein kinase, inhibiting kinase activity. *J. Virol.* **1996**, *70*, 6701–6708, doi:10.1128/jvi.70.10.6701-6708.1996.
99. Yang, X.; Goncalves, J.; Gabuzda, D. Phosphorylation of Vif and Its Role in HIV-1 Replication. *J. Biol. Chem.* **1996**, *271*, 10121–10129, doi:10.1074/jbc.271.17.10121.
100. Yang, X.; Gabuzda, D. Mitogen-activated Protein Kinase Phosphorylates and Regulates the HIV-1 Vif Protein. *J. Biol. Chem.* **1998**, *273*, 29879–29887, doi:10.1074/jbc.273.45.29879.
101. Kaushik, R.; Ratner, L. Role of Human Immunodeficiency Virus Type 1 Matrix Phosphorylation in an Early Postentry Step of Virus Replication. *J. Virol.* **2004**, *78*, 2319–2326, doi:10.1128/jvi.78.5.2319-2326.2004.
102. Strasner, A.B.; Natarajan, M.; Doman, T.; Key, D.; August, A.; Henderson, A.J. The Src Kinase Lck Facilitates Assembly of HIV-1 at the Plasma Membrane. *J. Immunol.* **2008**, *181*, 3706–3713, doi:10.4049/jimmunol.181.5.3706.

103. Gallay, P.; Swingler, S.; Song, J.; Bushman, F.; Trono, D. HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. *Cell* **1995**, *83*, 569–576, doi:10.1016/0092-8674(95)90097-7.
104. Hahn, F.; Setz, C.; Friedrich, M.; Rauch, P.; Solbak, S.M.; Frøystein, N. Åge; Henklein, P.; Votteler, J.; Fossen, T.; Schubert, U. Mutation of the Highly Conserved Ser-40 of the HIV-1 p6 Gag Protein to Phe Causes the Formation of a Hydrophobic Patch, Enhances Membrane Association, and Polyubiquitination of Gag. *Viruses* **2014**, *6*, 3738–3765, doi:10.3390/v6103738.
105. Leis, J.; Phillips, N.; Fu, X.; Tuazon, P.T.; Traugh, J.A. Phosphorylation of avian retrovirus matrix protein by Ca<sup>2+</sup>/phospholipid-dependent protein kinase. *JBIC J. Biol. Inorg. Chem.* **1989**, *179*, 415–422, doi:10.1111/j.1432-1033.1989.tb14569.x.
106. Nelle, T.D.; Verderame, M.F.; Leis, J.; Wills, J.W. The Major Site of Phosphorylation within the Rous Sarcoma Virus MA Protein Is Not Required for Replication. *J. Virol.* **1998**, *72*, 1103–1107, doi:10.1128/jvi.72.2.1103-1107.1998.
107. Pepinsky, R.B.; A. Papayannopoulos, I.; Campbell, S.; Vogt, V.M. Analysis of Rous sarcoma virus Gag protein by mass spectrometry indicates trimming by host exopeptidase. *J. Virol.* **1996**, *70*, 3313–3318, doi:10.1128/jvi.70.5.3313-3318.1996.
108. Fu, X.; Tuazon, P.T.; Traugh, J.A.; Leis, J. Site-Directed Mutagenesis of the Avian Retrovirus Nucleocapsid Protein, Pp12, at Serine 40, the Primary Site of Phosphorylation in Vivo. *J. Biol. Chem.* **1988**, *268*, 2134–2139.
109. Wang, H.; Machesky, N.J.; Mansky, L.M. Both the PPPY and PTAP Motifs Are Involved in Human T-Cell Leukemia Virus Type 1 Particle Release. *J. Virol.* **2004**, *78*, 1503–1512, doi:10.1128/jvi.78.3.1503-1512.2004.
110. Hemonnot, B.; Molle, D.; Bardy, M.; Gay, B.; Laune, D.; Devaux, C.; Briant, L. Phosphorylation of the HTLV-1 matrix L-domain-containing protein by virus-associated ERK-2 kinase. *Virology* **2006**, *349*, 430–439, doi:10.1016/j.virol.2006.02.043.
111. Bouamr, F.; Melillo, J.A.; Wang, M.Q.; Nagashima, K.; Santos, M.D.L.; Rein, A.; Goff, S.P. PPPYEPTAP Motif Is the Late Domain of Human T-Cell Leukemia Virus Type 1 Gag and Mediates Its Functional Interaction with Cellular Proteins Neddd4 and Tsg101. *J. Virol.* **2003**, *77*, 11882–11895, doi:10.1128/jvi.77.22.11882-11895.2003.
112. Yasuda, J.; Hunter, E. A Proline-Rich Motif (PPPY) in the Gag Polyprotein of Mason-Pfizer Monkey Virus Plays a Maturation-Independent Role in Virion Release. *J. Virol.* **1998**, *72*, 4095–4103, doi:10.1128/jvi.72.5.4095-4103.1998.
113. Henderson, E.L.; Sowder, R.; Smythers, G.; E. Benveniste, R.; Oroszlan, S. Purification and N-terminal amino acid sequence comparisons of structural proteins from retrovirus-D/Washington and Mason-Pfizer monkey virus. *J. Virol.* **1985**, *55*, 778–787, doi:10.1128/jvi.55.3.778-787.1985.
114. Uckert, W.; Wunderlich, V.; Fiebach, H.; Hertling, I.; Stein, U.; Kraft, R.; Desrosiers, R. Biochemical and immunological characterization of structural proteins from retrovirus-D/New England and comparison to Mason-Pfizer monkey virus and permanent human fibroblast virus. *Arch. Virol.* **1987**, *94*, 267–282, doi:10.1007/bf01310719.
115. Enssle, J.; Fischer, N.; Moebes, A.; Mauer, B.; Smola, U.; Rethwilm, A. Carboxy-terminal cleavage of the human foamy virus Gag precursor molecule is an essential step in the viral life cycle. *J. Virol.* **1997**, *71*, 7312–7317, doi:10.1128/jvi.71.10.7312-7317.1997.
116. Yuan, B.; Li, X.; Goff, S.P. Mutations altering the moloney murine leukemia virus p12 Gag protein affect virion production and early events of the virus life cycle. *EMBO J.* **1999**, *18*, 4700–4710, doi:10.1093/emboj/18.17.4700.
117. Yueh, A.; Goff, S.P. Phosphorylated Serine Residues and an Arginine-Rich Domain of the Moloney Murine Leukemia Virus p12 Protein Are Required for Early Events of Viral Infection. *J. Virol.* **2003**, *77*, 1820–1829, doi:10.1128/jvi.77.3.1820-1829.2003.
118. Brzezinski, J.D.; Felkner, R.; Modi, A.; Liu, M.; Roth, M.J. Phosphorylation Requirement of Murine Leukemia Virus p12. *J. Virol.* **2016**, *90*, 11208–11219, doi:10.1128/jvi.01178-16.
119. Callis, J. The Ubiquitination Machinery of the Ubiquitin System. *Arab. Book* **2014**, *12*, e0174, doi:10.1199/tab.0174.
120. Vijay-Kumar, S.; Bugg, C.; Wilkinson, K.; Vierstra, R.; Hatfield, P.; Cook, W. Comparison of the three-dimensional structures of human, yeast, and oat ubiquitin. *J. Biol. Chem.* **1987**, *262*, 6396–6399, doi:10.1016/s0021-9258(18)45583-4.
121. Komander, D. The emerging complexity of protein ubiquitination. *Biochem. Soc. Trans.* **2009**, *37*, 937–953, doi:10.1042/bst0370937.
122. Shields, S.B.; Piper, R.C. How Ubiquitin Functions with ESCRTs. *Traffic* **2011**, *12*, 1306–1317, doi:10.1111/j.1600-0854.2011.01242.x.
123. Ohtake, F.; Saeki, Y.; Ishido, S.; Kanno, J.; Tanaka, K. The K48-K63 Branched Ubiquitin Chain Regulates NF-κB Signaling. *Mol. Cell* **2016**, *64*, 251–266, doi:10.1016/j.molcel.2016.09.014.
124. Suresh, B.; Lee, J.; Kim, K.-S.; Ramakrishna, S. The Importance of Ubiquitination and Deubiquitination in Cellular Reprogramming. *Stem Cells Int.* **2016**, *2016*, 1–14, doi:10.1155/2016/6705927.
125. Gottwein, E.; Kräusslich, H.-G. Analysis of Human Immunodeficiency Virus Type 1 Gag Ubiquitination. *J. Virol.* **2005**, *79*, 9134–9144, doi:10.1128/jvi.79.14.9134-9144.2005.
126. Gottwein, E.; Jäger, S.; Habermann, A.; Kräusslich, H.-G. Cumulative Mutations of Ubiquitin Acceptor Sites in Human Immunodeficiency Virus Type 1 Gag Cause a Late Budding Defect. *J. Virol.* **2006**, *80*, 6267–6275, doi:10.1128/jvi.02177-05.
127. Ott, D.E.; Coren, L.V.; Ii, R.C.S.; Adams, J.; Schubert, U. Retroviruses Have Differing Requirements for Proteasome Function in the Budding Process. *J. Virol.* **2003**, *77*, 3384–3393, doi:10.1128/jvi.77.6.3384-3393.2003.
128. Sette, P.; Jadwin, J.A.; Dussupt, V.; Bello, N.F.; Bouamr, F. The ESCRT-Associated Protein Alix Recruits the Ubiquitin Ligase Neddd4-1 To Facilitate HIV-1 Release through the LYPXnL L Domain Motif. *J. Virol.* **2010**, *84*, 8181–8192, doi:10.1128/jvi.00634-10.
129. Sette, P.; Nagashima, K.; Piper, R.C.; Bouamr, F. Ubiquitin conjugation to Gag is essential for ESCRT-mediated HIV-1 budding. *Retrovirology* **2013**, *10*, 79, doi:10.1186/1742-4690-10-79.

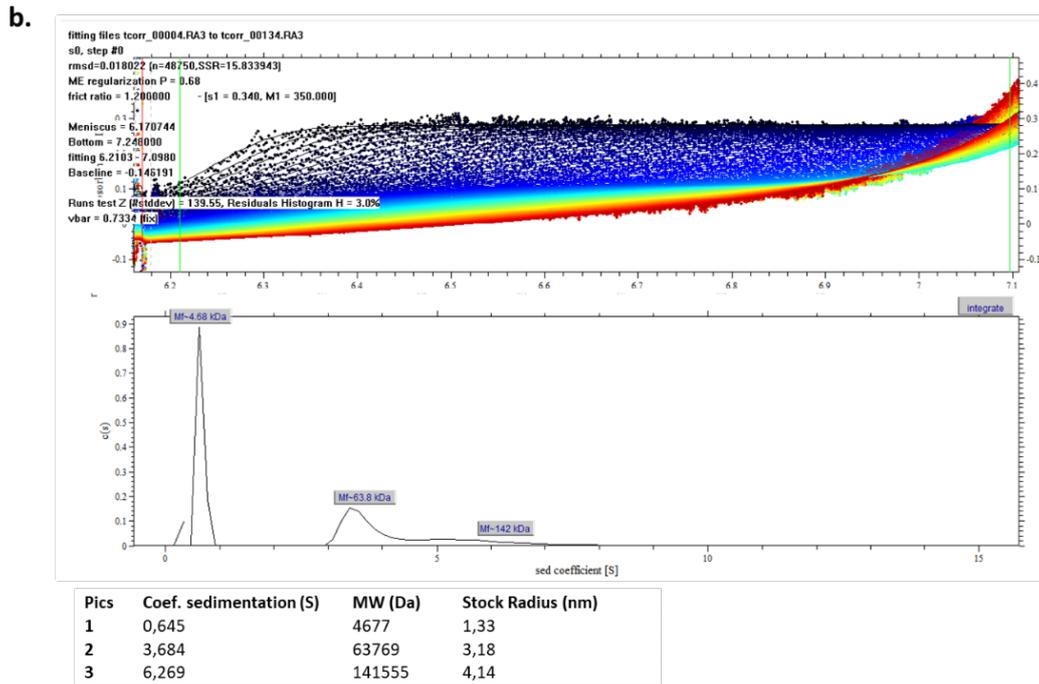
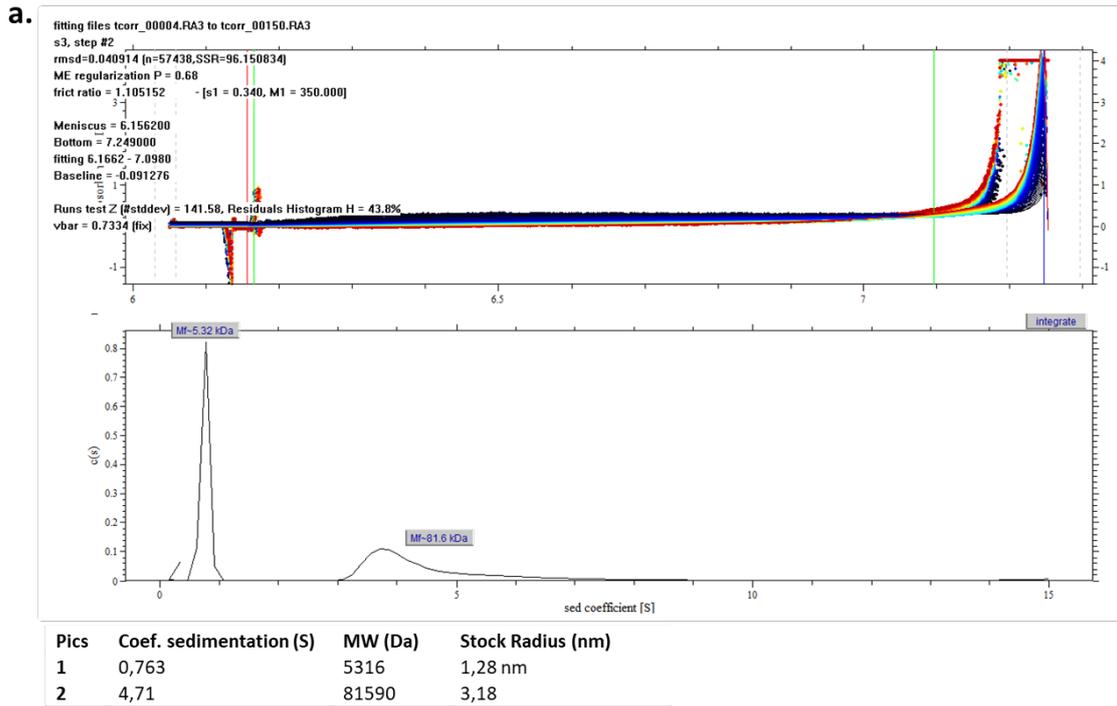
130. Ott, D.E.; Coren, L.V.; Copeland, T.D.; Kane, B.P.; Johnson, D.G.; Sowder, R.C.; Yoshinaka, Y.; Oroszlan, S.; Arthur, L.O.; Henderson, L.E. Ubiquitin Is Covalently Attached to the p6Gag Proteins of Human Immunodeficiency Virus Type 1 and Simian Immunodeficiency Virus and to the p12Gag Protein of Moloney Murine Leukemia Virus. *J. Virol.* **1998**, *72*, 2962–2968, doi:10.1128/jvi.72.4.2962-2968.1998.
131. Schubert, U.; Ott, D.E.; Chertova, E.N.; Welker, R.; Tessmer, U.; Princiotta, M.F.; Bennink, J.R.; Kräusslich, H.-G.; Yewdell, J.W. Proteasome inhibition interferes with Gag polyprotein processing, release, and maturation of HIV-1 and HIV-2. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13057–13062, doi:10.1073/pnas.97.24.13057.
132. Ott, D.E.; Coren, L.V.; Chertova, E.N.; Gagliardi, T.D.; Schubert, U. Ubiquitination of HIV-1 and MuLV Gag. *Virology* **2000**, *278*, 111–121, doi:10.1006/viro.2000.0648.
133. Patnaik, A.; Chau, V.; Wills, J.W. Ubiquitin is part of the retrovirus budding machinery. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13069–13074, doi:10.1073/pnas.97.24.13069.
134. Jäger, S.; Gottwein, E.; Kräusslich, H.-G. Ubiquitination of Human Immunodeficiency Virus Type 1 Gag Is Highly Dependent on Gag Membrane Association. *J. Virol.* **2007**, *81*, 9193–9201, doi:10.1128/jvi.00044-07.
135. Von Schwedler, U.K.; Stuchell, M.; Müller, B.; Ward, D.M.; Chung, H.Y.; Morita, E.; Wang, H.E.; Davis, T.; He, G.P.; Cimbara, D.M.; et al. The protein network of HIV budding. *Cell* **2003**, *114*, 701–713.
136. Pornillos, O.; Alam, S.L.; Rich, R.L.; Myszk, D.G.; Davis, D.R.; Sundquist, W.I. Structure and functional interactions of the Tsg101 UEV domain. *EMBO J.* **2002**, *21*, 2397–2406, doi:10.1093/emboj/21.10.2397.
137. Watanabe, S.M.; Strickland, M.; Tjandra, N.; Carter, C.A. RNA Binding Suppresses Tsg101 Recognition of Ub-Modified Gag and Facilitates Recruitment to the Plasma Membrane. *Viruses* **2020**, *12*, 447, doi:10.3390/v12040447.
138. Garrus, J.E.; von Schwedler, U.K.; Pornillos, O.W.; Morham, S.G.; Zavitz, K.H.; Wang, H.E.; Wettstein, D.A.; Stray, K.M.; Côté, M.; Rich, R.L.; et al. Tsg101 and the Vacuolar Protein Sorting Pathway Are Essential for HIV-1 Budding. *Cell* **2001**, *107*, 55–65, doi:10.1016/s0092-8674(01)00506-2.
139. Staub, O.; Dho, S.; Henry, P.; Correa, J.; Ishikawa, T.; McGlade, J.; Rotin, D. WW domains of Nedd4 bind to the proline-rich PY motifs in the epithelial Na<sup>+</sup> channel deleted in Liddle's syndrome. *EMBO J.* **1996**, *15*, 2371–2380, doi:10.1002/j.1460-2075.1996.tb00593.x.
140. Babst, M.; Katzmann, D.J.; Estepa-Sabal, E.J.; Meerloo, T.; Emr, S.D. Escrt-III. *Dev. Cell* **2002**, *3*, 271–282, doi:10.1016/s1534-5807(02)00220-4.
141. Dunn, R.; Klos, D.A.; Adler, A.S.; Hicke, L. The C2 domain of the Rsp5 ubiquitin ligase binds membrane phosphoinositides and directs ubiquitination of endosomal cargo. *J. Cell Biol.* **2004**, *165*, 135–144, doi:10.1083/jcb.200309026.
142. Spidel, J.L.; Craven, R.C.; Wilson, C.B.; Patnaik, A.; Wang, H.; Mansky, L.M.; Wills, J.W. Lysines Close to the Rous Sarcoma Virus Late Domain Critical for Budding. *J. Virol.* **2004**, *78*, 10606–10616, doi:10.1128/jvi.78.19.10606-10616.2004.
143. Ott, D.E.; Coren, L.V.; Sowder, R.C.; Adams, J.; Nagashima, K.; Schubert, U. Equine Infectious Anemia Virus and the Ubiquitin-Proteasome System. *J. Virol.* **2002**, *76*, 3038–3044, doi:10.1128/jvi.76.6.3038-3044.2002.
144. Chen, C.; Li, F.; Montelaro, R.C. Functional Roles of Equine Infectious Anemia Virus Gag p9 in Viral Budding and Infection. *J. Virol.* **2001**, *75*, 9762–9770, doi:10.1128/jvi.75.20.9762-9770.2001.
145. Heidecker, G.; Lloyd, P.A.; Fox, K.; Nagashima, K.; Derse, D. Late Assembly Motifs of Human T-Cell Leukemia Virus Type 1 and Their Relative Roles in Particle Release. *J. Virol.* **2004**, *78*, 6636–6648, doi:10.1128/jvi.78.12.6636-6648.2004.
146. Heidecker, G.; Lloyd, P.A.; Soheilian, F.; Nagashima, K.; Derse, D. The Role of WWP1-Gag Interaction and Gag Ubiquitination in Assembly and Release of Human T-Cell Leukemia Virus Type 1. *J. Virol.* **2007**, *81*, 9769–9777, doi:10.1128/jvi.00642-07.
147. Medina, G.; Pincetic, A.; Ehrlich, L.S.; Zhang, Y.; Tang, Y.; Leis, J.; Carter, C.A. Tsg101 can replace Nedd4 function in ASV Gag release but not membrane targeting. *Virology* **2008**, *377*, 30–38, doi:10.1016/j.viro.2008.04.024.
148. Bartusch, C.; Prange, R. ESCRT Requirements for Murine Leukemia Virus Release. *Viruses* **2016**, *8*, 103, doi:10.3390/v8040103.
149. Vana, M.L.; Tang, Y.; Chen, A.; Medina, G.; Carter, C.; Leis, J. Role of Nedd4 and Ubiquitination of Rous Sarcoma Virus Gag in Budding of Virus-Like Particles from Cells. *J. Virol.* **2004**, *78*, 13943–13953, doi:10.1128/jvi.78.24.13943-13953.2004.
150. Stanke, N.; Stange, A.; Luftnegger, D.; Zentgraf, H.; Lindemann, D. Ubiquitination of the Prototype Foamy Virus Envelope Glycoprotein Leader Peptide Regulates Subviral Particle Release. *J. Virol.* **2005**, *79*, 15074–15083, doi:10.1128/jvi.79.24.15074-15083.2005.
151. Yang, Y.; He, Y.; Wang, X.; Liang, Z.; He, G.; Zhang, P.; Zhu, H.; Xu, N.; Liang, S. Protein SUMOylation modification and its associations with disease. *Open Biol.* **2017**, *7*, 170167, doi:10.1098/rsob.170167.
152. Han, Z.-J.; Feng, Y.-H.; Gu, B.-H.; Li, Y.-M.; Chen, H. The post-translational modification, SUMOylation, and cancer (Review). *Int. J. Oncol.* **2018**, *52*, 1081–1094, doi:10.3892/ijo.2018.4280.
153. Matunis, M.J.; Coutavas, E.; Blobel, G. A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell Biol.* **1996**, *135*, 1457–1470, doi:10.1083/jcb.135.6.1457.
154. Owerbach, D.; McKay, E.M.; Yeh, E.T.; Gabbay, K.H.; Bohren, K.M. A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. *Biochem. Biophys. Res. Commun.* **2005**, *337*, 517–520, doi:10.1016/j.bbrc.2005.09.090.
155. Mahajan, R.; Gerace, L.; Melchior, F. Molecular Characterization of the SUMO-1 Modification of RanGAP1 and Its Role in Nuclear Envelope Association. *J. Cell Biol.* **1998**, *140*, 259–270, doi:10.1083/jcb.140.2.259.
156. Saitoh, H.; Hinchey, J. Functional Heterogeneity of Small Ubiquitin-related Protein Modifiers SUMO-1 versus SUMO-2/3. *J. Biol. Chem.* **2000**, *275*, 6252–6258, doi:10.1074/jbc.275.9.6252.

157. Desterro, J.M.; Rodriguez, M.S.; Hay, R.T. SUMO-1 Modification of I $\kappa$ B $\alpha$  Inhibits NF- $\kappa$ B Activation. *Mol. Cell* **1998**, *2*, 233–239, doi:10.1016/s1097-2765(00)80133-1.
158. Nayak, A.; Müller, S. SUMO-specific proteases/isopeptidases: SENPs and beyond. *Genome Biol.* **2014**, *15*, 422, doi:10.1186/s13059-014-0422-2.
159. Tatham, M.H.; Kim, S.; Jaffray, E.; Song, J.; Chen, Y.; Hay, R.T. Unique binding interactions among Ubc9, SUMO and RanBP2 reveal a mechanism for SUMO paralog selection. *Nat. Struct. Mol. Biol.* **2004**, *12*, 67–74, doi:10.1038/nsmb878.
160. Eisenhardt, N.; Ilic, D.; Nagamalleswari, E.; Pichler, A. Biochemical characterization of SUMO-conjugating enzymes by in vitro sumoylation assays. *Methods Enzym.* **2019**, *618*, 167–185, doi:10.1016/bs.mie.2018.12.025.
161. Jeon, Y.J.; Yoo, H.M.; Chung, C.H. ISG15 and immune diseases. *Biochim. Biophys. Acta (BBA) Mol. Basis Dis.* **2010**, *1802*, 485–496, doi:10.1016/j.bbadis.2010.02.006.
162. Gurer, C.; Berthou, L.; Luban, J. Covalent Modification of Human Immunodeficiency Virus Type 1 p6 by SUMO-1. *J. Virol.* **2005**, *79*, 910–917, doi:10.1128/jvi.79.2.910-917.2005.
163. Jaber, T.; Bohl, C.R.; Lewis, G.L.; Wood, C.; West, J.T.; Weldon, R.A. Human Ubc9 Contributes to Production of Fully Infectious Human Immunodeficiency Virus Type 1 Virions. *J. Virol.* **2009**, *83*, 10448–10459, doi:10.1128/jvi.00237-09.
164. Martinez, N.W.; Xue, X.; Berro, R.G.; Kreitzer, G.; Resh, M.D. Kinesin KIF4 Regulates Intracellular Trafficking and Stability of the Human Immunodeficiency Virus Type 1 Gag Polyprotein. *J. Virol.* **2008**, *82*, 9937–9950, doi:10.1128/jvi.00819-08.
165. Lamoliatte, F.; McManus, F.P.; Maarifi, G.; Chelbi-Alix, M.K.; Thibault, P. Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using sequential peptide immunopurification. *Nat. Commun.* **2017**, *8*, 14109, doi:10.1038/ncomms14109.
166. Kim, W.; Tang, Y.; Okada, Y.; Torrey, T.A.; Chattopadhyay, S.K.; Pfeleiderer, M.; Falkner, F.G.; Dorner, F.; Choi, W.; Hirokawa, N.; et al. Binding of Murine Leukemia Virus Gag Polyproteins to KIF4, a Microtubule-Based Motor Protein. *J. Virol.* **1998**, *72*, 6898–6901, doi:10.1128/jvi.72.8.6898-6901.1998.
167. Weldon, A.R.; Sarkar, P.; Brown, S.M.; Weldon, S.K. Mason–Pfizer monkey virus Gag proteins interact with the human sumo conjugating enzyme, hUbc9. *Virology* **2003**, *314*, 62–73, doi:10.1016/s0042-6822(03)00348-9.
168. Yueh, A.; Leung, J.; Bhattacharyya, S.; Perrone, L.A.; Santos, K.D.L.; Pu, S.-Y.; Goff, S.P. Interaction of Moloney Murine Leukemia Virus Capsid with Ubc9 and PIASy Mediates SUMO-1 Addition Required Early in Infection. *J. Virol.* **2006**, *80*, 342–352, doi:10.1128/jvi.80.1.342-352.2006.
169. Wang, J.; Wen, S.; Zhao, R.; Qi, J.; Liu, Z.; Li, W.; An, J.; Wood, C.; Wang, Y. Covalent conjugation of the equine infectious anemia virus Gag with SUMO. *Biochem. Biophys. Res. Commun.* **2017**, *486*, 712–719, doi:10.1016/j.bbrc.2017.03.103.
170. Perng, Y.-C.; Lenschow, D.J. ISG15 in antiviral immunity and beyond. *Nat. Rev. Genet.* **2018**, *16*, 423–439, doi:10.1038/s41579-018-0020-5.
171. Potter, J.L.; Narasimhan, J.; Mende-Mueller, L.; Haas, A.L. Precursor Processing of Pro-ISG15/UCRP, an Interferon- $\beta$ -induced Ubiquitin-like Protein. *J. Biol. Chem.* **1999**, *274*, 25061–25068, doi:10.1074/jbc.274.35.25061.
172. Villarroja-Beltri, C.; Guerra, S.; Sánchez-Madrid, F. ISGylation—A key to lock the cell gates for preventing the spread of threats. *J. Cell Sci.* **2017**, *130*, 2961–2969, doi:10.1242/jcs.205468.
173. Zhao, C.; Collins, M.N.; Hsiang, T.-Y.; Krug, R.M. Interferon-induced ISG15 pathway: An ongoing virus–host battle. *Trends Microbiol.* **2013**, *21*, 181–186, doi:10.1016/j.tim.2013.01.005.
174. Hou, S.X.; Zheng, Z.; Chen, X.; Perrimon, N. The JAK/STAT Pathway in Model Organisms. *Dev. Cell* **2002**, *3*, 765–778, doi:10.1016/s1534-5807(02)00376-3.
175. Malakhov, M.P.; Malakhova, O.A.; Kim, K.I.; Ritchie, K.J.; Zhang, D.-E. UBP43 (USP18) Specifically Removes ISG15 from Conjugated Proteins. *J. Biol. Chem.* **2002**, *277*, 9976–9981, doi:10.1074/jbc.m109078200.
176. Levy, A.J.; Scott, I.; Mackewicz, C. Protection from HIV/AIDS: The importance of innate immunity. *Clin. Immunol.* **2003**, *108*, 167–174, doi:10.1016/s1521-6616(03)00178-5.
177. Shirazi, Y.; Pitha, P.M. Alpha interferon inhibits early stages of the human immunodeficiency virus type 1 replication cycle. *J. Virol.* **1992**, *66*, 1321–1328, doi:10.1128/jvi.66.3.1321-1328.1992.
178. Cheney, K.M.; Áine, M. Interferon-Alpha Mediates Restriction of Human Immunodeficiency Virus Type-1 Replication in Primary Human Macrophages at an Early Stage of Replication. *PLoS ONE* **2010**, *5*, e13521, doi:10.1371/journal.pone.0013521.
179. Pincetic, A.; Kuang, Z.; Seo, E.J.; Leis, J. The Interferon-Induced Gene ISG15 Blocks Retrovirus Release from Cells Late in the Budding Process. *J. Virol.* **2010**, *84*, 4725–4736, doi:10.1128/jvi.02478-09.
180. Woods, M.W.; Kelly, J.N.; Hattmann, C.J.; Tong, J.G.K.; Xu, L.S.; Coleman, M.D.; Quest, G.R.; Smiley, J.R.; Barr, S.D. Human HERC5 restricts an early stage of HIV-1 assembly by a mechanism correlating with the ISGylation of Gag. *Retrovirology* **2011**, *8*, 95, doi:10.1186/1742-4690-8-95.
181. Okumura, A.; Lu, G.; Pitha-Rowe, I.; Pitha, P.M. Innate antiviral response targets HIV-1 release by the induction of ubiquitin-like protein ISG15. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1440–1445, doi:10.1073/pnas.0510518103.
182. Morales, D.J.; Lenschow, D.J. The antiviral activities of ISG15. *J. Mol. Biol.* **2013**, *425*, 4995–5008, doi:10.1016/j.jmb.2013.09.041.
183. Zou, W.; Papov, V.; Malakhova, O.; Kim, K.I.; Dao, C.; Li, J.; Zhang, D.-E. ISG15 modification of ubiquitin E2 Ubc13 disrupts its ability to form thioester bond with ubiquitin. *Biochem. Biophys. Res. Commun.* **2005**, *336*, 61–68, doi:10.1016/j.bbrc.2005.08.038.
184. Ambler, R.P.; Rees, M.W.  $\epsilon$ -N-Methyl-lysine in Bacterial Flagellar Protein. *Nat. Cell Biol.* **1959**, *184*, 56–57, doi:10.1038/184056b0.
185. Levy, D. Lysine methylation signaling of non-histone proteins in the nucleus. *Cell. Mol. Life Sci.* **2019**, *76*, 2873–2883, doi:10.1007/s00018-019-03142-0.

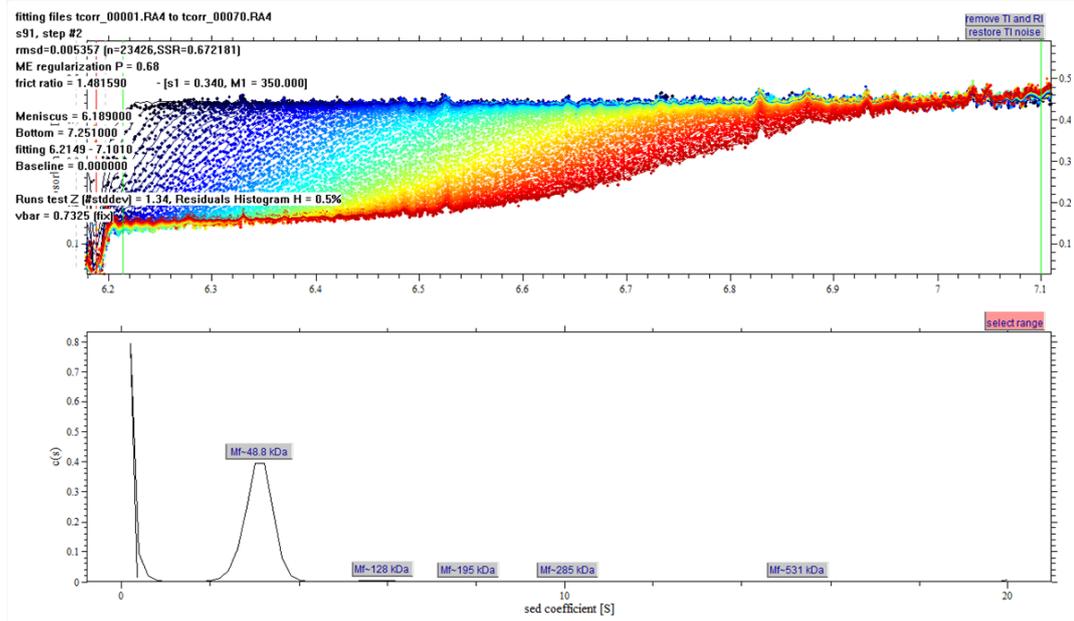
186. Lanouette, S.; Mongeon, V.; Figeys, D.; Couture, J. The functional diversity of protein lysine methylation. *Mol. Syst. Biol.* **2014**, *10*, 724, doi:10.1002/msb.134974.
187. Blanchet, F.; Cardona, A.; Letimier, F.A.; Hershfield, M.S.; Acuto, O. CD28 costimulatory signal induces protein arginine methylation in T cells. *J. Exp. Med.* **2005**, *202*, 371–377, doi:10.1084/jem.20050176.
188. Smith, W.A.; Schurter, B.T.; Wong-Staal, F.; David, M. Arginine Methylation of RNA Helicase A Determines Its Subcellular Localization. *J. Biol. Chem.* **2004**, *279*, 22795–22798, doi:10.1074/jbc.c300512200.
189. Willemsen, N.M.; Hitchen, E.M.; Bodetti, T.J.; Apolloni, A.; Warrilow, D.; Piller, S.C.; Harrich, D. Protein methylation is required to maintain optimal HIV-1 infectivity. *Retrovirology* **2006**, *3*, 92, doi:10.1186/1742-4690-3-92.
190. Lochmann, T.L.; Bann, D.V.; Ryan, E.P.; Beyer, A.R.; Mao, A.; Cochrane, A.; Parent, L.J. NC-mediated nucleolar localization of retroviral gag proteins. *Virus Res.* **2013**, *171*, 304–318, doi:10.1016/j.virusres.2012.09.011.
191. Invernizzi, C.D.F.; Xie, B.; Frankel, F.A.; Feldhammer, M.; Roy, B.B.; Richard, S.; Wainberg, M.A. Arginine methylation of the HIV-1 nucleocapsid protein results in its diminished function. *AIDS* **2007**, *21*, 795–805, doi:10.1097/qad.0b013e32803277ae.
192. Paris, J.; Tobaly-Tapiero, J.; Giron, M.-L.; Burlaud-Gaillard, J.; Buseyne, F.; Roingeard, P.; Lesage, P.; Zamborlini, A.; Saïb, A. The invariant arginine within the chromatin-binding motif regulates both nucleolar localization and chromatin binding of Foamy virus Gag. *Retrovirology* **2018**, *15*, 48, doi:10.1186/s12977-018-0428-z.

**Annexe 1 : Bussienne, C., Marquet, R., Paillart, J., C., and Bernacchi, S. (2021). Post-translational modifications of retroviral HIV-1 Gag precursors: an overview of their biological role. *Int. J. Mol. Sci.* 2021, 22, 2871**





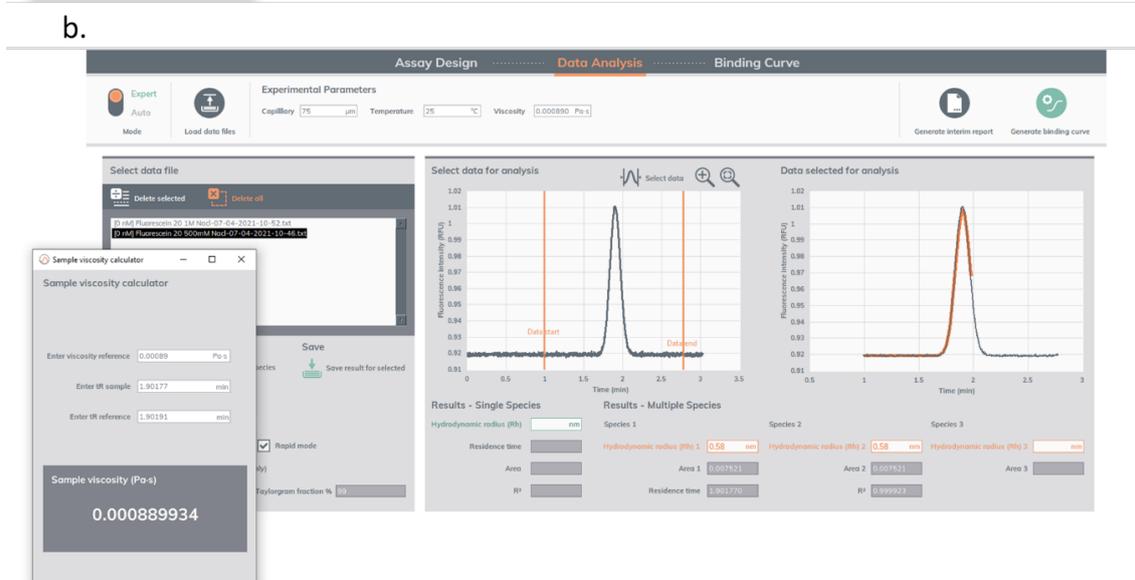
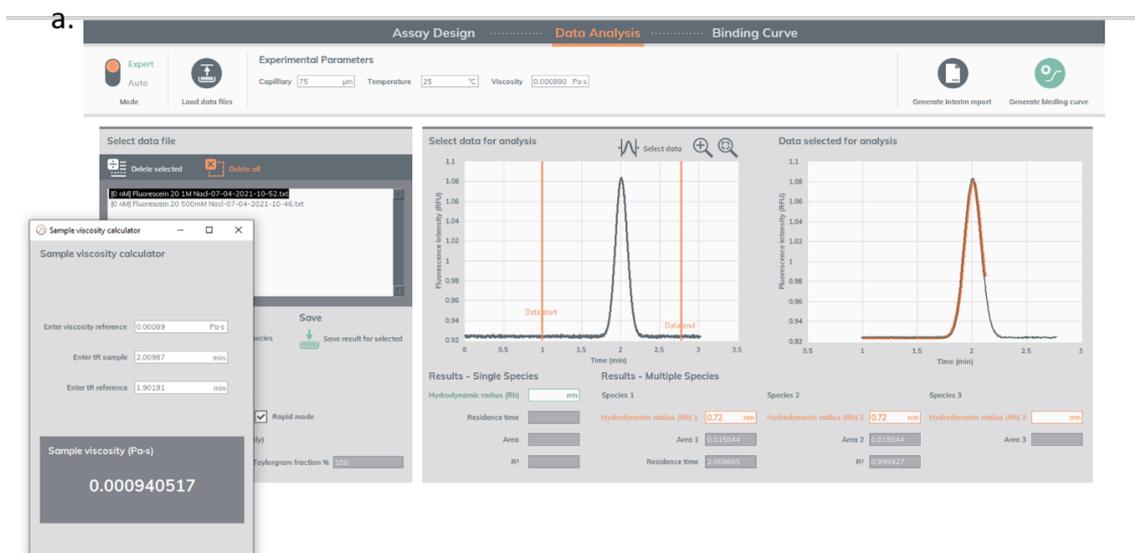
**C.**



Pics	Coef. sedimentation (S)	MW (Da)	Stock Radius (nm)
1	3.215	48774	3.58
2	6.106	127618	4.94
3	8.108	195306	5.69

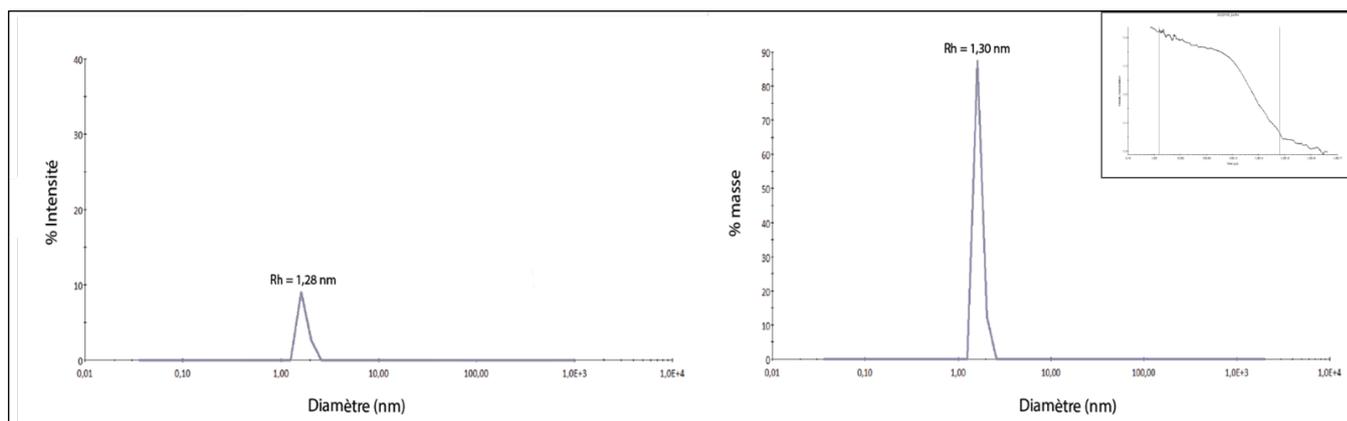
**Annexe 2 : Profils AUC obtenus après application du modèle distribution continue avec une variable discrète**

Les paramètres mesurés se trouvent en haut à gauche. Le premier carré correspond à l'alignement entre les courbes expérimentales et le modèle. Le second aux coefficients de sédimentations représentés pour chaque espèce. Un tableau récapitulant les valeurs de  $c(s)$ ,  $PM$  et  $R_h$  est ajouté en dessous de l'analyse. **a.** Correspond à la protéine 6His-Gag purifiée dans le tampon A1. **b.** Correspond à la protéine 6His-Gag purifiée dans le tampon A3. **c.** Correspond à la protéine 6His-Gag purifiée dans le tampon A1 et dialysé dans le tampon A3.



### Annexe 3 : Profils FIDAbio obtenu pour déterminer la viscosité de la solution tamponnée A1 et A3

Dans l'encadré à gauche le calcul détaillé dans la section Matériels et Méthodes (SIX.3) pour déterminer la viscosité de la solution tamponnée. La Courbe de gauche est la représentation de l'intensité du signal émis par la fluorescéine au cours du temps dans le tampon A1 ou A3. A droite : représentation de l'alignement entre le modèle et les valeurs expérimentales. **a.** Tampon A1. **b.** Tampon A3.



#### Annexe 4. Analyse DLS de la solution tamponnée A1

Représentation des distributions en intensité (à gauche) et en masse (à droite) en fonction des diamètres des protéines en solution. Le Rh est indiqué au-dessus du pic correspondant. Chaque graphique correspond à la moyenne de dix mesures. Les corrélogrammes sont représentés en bleu dans l'encadré en haut à droite.







## Etude structurale et fonctionnelle de la protéine Pr55Gag du VIH-1 en association avec le signal d'encapsidation de l'ARN génomique

### Résumé

Le virus de l'immunodéficience humaine de type 1 (VIH-1) sélectionne son ARN génomique (ARNg) parmi de nombreux ARN cellulaires et viraux épissés impliquant des interactions spécifiques entre la protéine Gag et le signal d'encapsidation de l'ARNg. Pour que l'ARN soit encapsidé dans les particules virales, son interaction avec Gag doit devenir non spécifique. Actuellement, les mécanismes moléculaires impliqués dans la **sélection et l'encapsidation** de l'ARNg par la protéine Gag ne sont pas entièrement compris.

Mon projet de thèse combine des analyses structurales et fonctionnelles pour caractériser et visualiser l'interaction de Gag avec l'ARNg et pour déterminer le rôle des partenaires d'un domaine de Gag (p6) dans les changements de spécificités d'interactions entre Gag et l'ARNg. Pour cela, j'ai optimisé la production des protéines en cellules de mammifères et j'ai résolu la structure de Gag à 7 Å par CryoEM. J'ai initié la caractérisation des complexes entre Gag et ses partenaires en utilisant différentes méthodes biophysiques.

Ces informations permettront d'identifier précisément les zones de contact protéine/ARN qui assurent la reconnaissance spécifique de l'ARN par Gag et de déterminer le rôle des partenaires de p6 pour l'encapsidation de l'ARNg. À plus long terme, ce travail apportera des informations pour identifier des inhibiteurs de cette interaction.

**Mots clé : VIH-1, assemblage viral, encapsidation de l'ARNg, Relation structure/fonction, précurseur Gag.**

## Structural and functional study of the HIV-1 Pr55Gag protein in association with the genomic RNA packaging signal

### Abstract

Human immunodeficiency virus type 1 (HIV-1) selects its genomic RNA (gRNA) from many cellular and spliced viral RNA involving specific interactions between Gag protein and the gRNA packaging signal. To allow RNA encapsidation in viral particles, its interaction with Gag must become nonspecific. Currently, the molecular mechanisms involved in the selection and packaging of gRNA by the Gag protein are not fully understood.

My PhD project combines structural and functional analyses to characterize and visualize the interaction of Gag with gRNA and to determine the role of p6 domain's partners of Gag in the changes of specificity of interactions between Gag and gRNA. For this purpose, I optimized proteins production in mammalian cells and solved the structure of Gag at 7 Å resolution by CryoEM. I initiated the characterization of the complexes between Gag and its partners using different biophysical methods.

This information will allow to identify precisely the protein/RNA contact areas that ensure the specific recognition of RNA by Gag and to determine the role of p6 partners for the encapsidation of gRNA. In the longer term, this work will provide information to identify inhibitors of this interaction.

**Key words : HIV-1, viral assembly, gRNA encapsidation, structure/function relationship, Gag precursor.**