

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

**Institut de Biologie Moléculaire et Cellulaire**

**UPR 9002 - Architecture et Réactivité de l'ARN**

**THÈSE** présentée par :

**Orian GILMER**

soutenue le : **30 Novembre 2022**

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

Discipline/ Spécialité : **Aspects Moléculaires et Cellulaires de la Biologie**

**Étude de la structure de la région 5' terminale de  
l'ARN génomique du virus de  
l'immunodéficience humaine de type 1 et des  
réarrangements induits par Gag et ses produits  
de maturation**

**THÈSE dirigée par :**

**Dr VIVET-BOUDOU Valérie**

IBMC, UPR 9002, Université de Strasbourg

**RAPPORTEURS :**

**Dr DE ROCQUIGNY Hugues**

**Dr SARGUEIL Bruno**

Inserm – U1259 MAVIVH – UFR Médecine Tours

CNRS UMR8015 – Université Paris Descartes

**AUTRES MEMBRES DU JURY :**

**Dr FRIANT Sylvie**

IPCB – GMGM, Université de Strasbourg

**INVITÉS :**

**Dr ALLMANG-CURA Christine**

IBMC, UPR 9002, Université de Strasbourg



# Remerciements

Tout d'abord, je tiens à remercier les membres de mon jury Dr Hugues De Rocquigny, Dr Bruno Sargueil, Dr Sylvie Friant et Dr Christine Allmang-Cura pour avoir accepté d'évaluer mon travail de doctorat. Je souhaite également remercier le ministère de l'enseignement supérieur et de la recherche ainsi que Sidaction pour le financement de mon doctorat.

J'aimerais aussi exprimer toute ma reconnaissance à Valérie, ma directrice de thèse, pour son soutien, sa disponibilité et sa patience. Du début de mon stage de M2 à la dernière ligne couchée sur ce manuscrit, merci de m'avoir guidé, corrigé et aidé à m'épanouir au laboratoire.

Merci JC pour ce premier stage de M1 qui m'a permis de découvrir le labo, pour avoir souvent pris le temps de répondre à mes questions, pour toute l'aide et pour l'énergie dont tu as fait preuve tout au long de ma thèse.

Merci à tous les membres de laboratoire que j'ai pu croiser lors de mon passage, anciens ou présents. Elodie pour m'avoir ouvert la voie du SHAPE, Noé, Damien pour vos conseils et votre bonne humeur, Tanja pour le calme et l'enthousiasme dont tu as fait preuve, mais aussi pour l'expertise en miniprep que tu m'as permis d'acquérir 😊, Benjamin pour m'avoir supporté pendant 3 ans sans (trop) bronché, Maud et Redmond pour votre soutien.

Merci également à Serena pour les discussions qui m'ont aidé à débloquer ma thèse quand les complexes refusaient de se former. Roland pour ton aide scientifique. Anne pour l'aide technique apporté et la bonne humeur en toute situation. Anne-Caroline pour tous tes conseils sur le séquençage et ta disponibilité. Antoine pour ton aide statistique précieuse (qui ne te servira malheureusement pas dans ton projet actuel 😊). Charlotte, Lisa, vos passages à l'IBM ont toujours été des bouffées d'air frais. Florine, je te souhaite le meilleur début de thèse possible. Rithu, may you keep your smile and stay away from the darkness 😊.

Cédric, merci pour tout, à toi, et à tous ceux du BTS (les tortues ninjas en particulier). Votre soutien, bonne humeur et amitié ont été un moteur pour moi pendant cette thèse.

PS : Cédric, arrête les blagues, s'il te plait. Merci.

Après presque cinq ans passés à l'institut, j'ai eu l'occasion de rencontrer de nombreuses personnes formidables. Merci à Guillaume et Phillipe pour l'animation régulière. Merci à David, Lucas, Matthieu et plus largement à la Romby Team et au box 436 pour tous les moments partagés. Merci à Noémie et Mathilde, pour avoir accepté mes caprices des pauses de midi (mais toujours sans violence, quoiqu'on en dise !). Merci à l'équipe Ryckelynck, présente et passée, pour l'hébergement, les nombreuses pauses café (en particulier avec Emilie), les discussions scientifiques et les délires partagés. Dans la même lignée, merci aux anciens de l'équipe Pfeffer, j'ai nommé Antoine, Olivier, Mister T, Monika et Morgane (bientôt), pour les discussions, scientifiques ou non, votre bonne humeur et votre présence.

Avant de conclure, je tiens à remercier mes amis, Etienne, Quentin, Clara, Ludo et David M. pour leur soutien depuis « presque » toujours.

Merci à tous mes amis du Master, et plus particulièrement, à David G., Marion, Monika, Kévin qui ont partagé mes galères et qui m'ont aidé et supporté ces cinq dernières années.

Un énorme merci à la princesse du Sundgau, j'ai nommé Watacha, et à son prince, Baptiste, qui m'ont fait l'honneur d'être témoin de leur union. Merci pour votre présence tout au long de ma thèse, pour le soutien moral et physique (souvent prodigué par Fluffy, certes, mais tout de même) et pour les nombreuses journées jeux de société et escape, passées et à venir.

Je tiens à remercier ma famille pour son soutien constant et inconditionnel. Un grand merci à ma sœur, pour son soutien moral passant par sa magie culinaire. Et à mes deux parents qui ont toujours été des modèles dans ce que j'ai entrepris. Merci aussi à la famille Rousseau pour m'avoir accueilli à bras ouvert et m'avoir épaulé tout au long de ma thèse.

Pour finir, je tiens à remercier Claire (depuis peu Dr Rousseau), pour son support, sa gentillesse et son amour, qui m'ont porté à travers ces quatre années de thèse. Rien n'aurait été pareil sans toi.





# Table des matières

<b>Table des matières</b> .....	
<b>Abréviations</b> .....	
<b>Liste des figures</b> .....	
<b>Liste des tableaux</b> .....	
<b>INTRODUCTION</b> .....	<b>1</b>
<b>I/ VIH-1 : généralités</b> .....	<b>1</b>
I.1/ Épidémiologie .....	1
I.2/ Taxonomie .....	2
I.3/ Structure de la particule virale.....	2
I.4/ Organisation du génome.....	3
I.5/ Cycle viral .....	7
I.5.1/ Phase précoce .....	9
I.5.1.1/ Reconnaissance, fusion et entrée.....	9
I.5.1.2/ Import nucléaire, décapsidation et transcription inverse .....	9
I.5.1.3/ Intégration.....	12
I.5.2/ Phase tardive.....	12
I.5.2.1/ Expression du génome viral.....	12
I.5.2.2/ Épissage, export nucléaire et traduction.....	13
I.5.2.3/ Assemblage, bourgeonnement et maturation .....	13
<b>II/ ARNg du VIH-1 : relation structure-fonction</b> .....	<b>16</b>
II.1/ Transcription .....	17
II.2/ Épissage.....	17
II.3/ Export nucléaire .....	18
II.4/ Traduction.....	18
II.5/ Dimérisation, sélection et encapsidation.....	19
II.6/ Transcription inverse .....	20
<b>III/ Méthodes d'études des structures d'ARN</b> .....	<b>21</b>
III.1/ Approches biophysiques .....	21
III.2/ Approches biochimiques.....	23
<b>IV/ Evolution des modèles structuraux rapportés dans la littérature</b> .....	<b>45</b>
IV.1/ Les motifs de l'ARNg du VIH-1 .....	46
IV.1.1/ La tige-boucle TAR .....	46
IV.1.2/ Le domaine RRE .....	47
IV.1.3/ Le domaine de « Frameshift ».....	48
IV.1.4/ Le domaine PBS.....	49
IV.1.5/ Le site d'initiation de la dimérisation.....	51
IV.1.6/ Le signal d'encapsidation .....	52
IV.2/ La région 5' terminale de l'ARNg du VIH-1 dans sa globalité.....	54
IV.2.1/ Le polymorphisme de la région 5' terminale et l'activité biologique.....	57
IV.2.2/ Les modèles structuraux <i>in virio</i> .....	61
IV.3/ Implication des protéines structurales virales sur la structure et le rôle de l'ARNg.....	65
IV.3.1/ La protéine Pr55 <sup>Gag</sup> .....	65
IV.3.2/ Les protéines NCp15 et NCp9 .....	67
IV.3.3/ La protéine NCp7 .....	67
<b>Problématique du sujet de recherche</b> .....	<b>69</b>

## TABLE DES MATIERES

<b>Axe 1.....</b>	<b>73</b>
<b>Matériel &amp; Méthodes :.....</b>	<b>77</b>
<b>I/ Production de l'ARN <i>in vitro</i>.....</b>	<b>77</b>
I.1/ Linéarisation du plasmide pNL4-3 clone 159 .....	77
I.2/ Transcription <i>in vitro</i> de l'ARNg 1-604 .....	78
I.3/ Purification de l'ARN sur colonne TSK G4000 .....	78
<b>II/ Formation des complexes ARN-protéines .....</b>	<b>79</b>
<b>III/ Modifications SHAPE à l'aide de NMIA.....</b>	<b>80</b>
III.1/ Modification de l'ARN seul ou en complexe .....	80
III.2/ Modification des complexes traités à l'AT-2 .....	80
III.3/ Modification des complexes traités à la protéinase K .....	80
<b>IV/ Transcription inverse des ADNc et séparation par électrophorèse capillaire .....</b>	<b>81</b>
<b>Résultats.....</b>	<b>83</b>
<b>I/ Production et purification de l'ARN 1-604 .....</b>	<b>83</b>
<b>II/ Optimisation du protocole hSHAPE.....</b>	<b>84</b>
II.1/ Protocole initial.....	84
II.1.1/ Formation des complexes ARN-Pr55 <sup>Gag</sup> .....	84
II.1.2/ Cartographie chimique <i>in vitro</i> : hSHAPE .....	86
II.2/ Mise au point .....	89
II.2.1/ Formation des complexes .....	89
II.2.2/ Validation des conditions expérimentales sur l'ARN nu .....	90
II.2.3/ Validation des nouvelles conditions expérimentales sur les complexes ribonucléoprotéiques ..	92
<b>III/ Structural maturation of the HIV-1 RNA 5' untranslated region by Pr55<sup>Gag</sup> and its maturation products.....</b>	<b>95</b>
<b>IV/ Analyses complémentaires .....</b>	<b>118</b>
IV.1/Analyse globale des réactivités associées à l'ARNg 1-604 .....	118
IV.2/Analyse globale des réactivités associées à l'ARNg 1-604 en présence de protéines .....	119
IV.3/Analyse spécifique des réactivités associées à l'ARNg 1-604 en présence de protéines et en fonction des domaines .....	120
IV.4/Comparaison avec la littérature .....	122
<b>Discussion : .....</b>	<b>126</b>
<b>Axe 2.....</b>	<b>131</b>
<b>Matériel &amp; Méthodes :.....</b>	<b>139</b>
<b>I/ Préparation des ARN modifiés pour le MOHCA-seq.....</b>	<b>139</b>
I.1/ Transcription <i>in vitro</i> avec 2'NH <sub>2</sub> ATP .....	139
I.2/ Purification de l'ARN transcrit <i>in vitro</i> .....	139
I.3/ Modification des ARN 2'NH <sub>2</sub> par l'ITCB-EDTA .....	140
I.4/ Chélation du fer sur l'ITBC-EDTA.....	140
<b>II/ MOHCA-seq : fragmentation des ARN.....</b>	<b>140</b>
II.1/ Clivages aux radicaux hydroxyles.....	141
II.2/ Déphosphorylation et rephosphorylation des fragments d'ARN.....	141
II.3/ Séparation des fragments en fonction de leur taille.....	141
<b>III/ Préparation de la librairie de séquençage à haut-débit .....</b>	<b>142</b>
III.1/ Ligation d'un adaptateur en 3' des fragments d'ARN : .....	142
III.2/ Transcription inverse des fragments ARN ligués : .....	142
III.3/ Ligation d'un adaptateur en 3' des ADNc : .....	143
III.4/ Amplification de la librairie d'ADNc par PCR.....	144



## TABLE DES MATIERES

III.5/ Purification de la librairie d'ADNc.....	144
<b>Résultats.....</b>	<b>147</b>
<b>I/ Approche 1: Cheng <i>et al.</i>, (2015) .....</b>	<b>150</b>
<b>II/ Approche 2: Kit NEBNext® Small RNA Library Prep Set for Illumina® .....</b>	<b>150</b>
<b>III/ Adaptation du MOHCA-seq à l'étude de la région 5' terminale de l'ARNg .....</b>	<b>155</b>
III.1/ Taux d'incorporation des nucléotides modifiés et temps de clivage.....	155
III.2/ Purification des ARN modifiés.....	159
<b>IV/ Approche 3: Kit NEBNext® Small RNA Library Prep Set for Illumina® .....</b>	<b>160</b>
<b>V/ Mise au point de la préparation de la banque de séquences pour le MOHCA-seq .....</b>	<b>165</b>
V.1/ Mise au point de la ligation en 3'.....	165
V.2/ Mise au point de la transcription inverse, de la seconde ligation et de l'amplification PCR .....	166
<b>VI/ Approche 4 : Préparation de banque de séquence adaptée à la région 5' de l'ARNg du VIH-1 pour un séquençage Illumina® .....</b>	<b>172</b>
<b>Discussion : .....</b>	<b>181</b>
<b>Discussion générale, perspectives et conclusion .....</b>	<b>187</b>
<b>Bibliographie .....</b>	<b>193</b>
<b>Annexes .....</b>	<b>213</b>



## ABREVIATIONS

# Abréviations

**(v/v)** : volume/volume

**%** : pourcentage

**°C** : degrés Celsius

**Å** : Angströms

**ADN** : Acide desoxyribonucléique

**ADNc** : ADN complémentaire

**ADNdb** : ADN double brin

**AMV** : *Avian Myeloblastosis Virus*

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

**ARN** : Acide Ribonucléique

**ARNdb** : ARN double brin

**ARNg** : ARN génomique

**ARNm** : ARN messenger

**ARNsb** : ARN simple brin

**ARNsn U1** : *small nuclear RNA U1*

**ARNt** : ARN de transfert

**ARNt<sup>Lys,3</sup>** : ARNt de la lysine, isoforme 3

**ASP** : AntiSense Protein

**AT-2** : aldrithiol-2

**ATP** : Adenosine triphosphate

**BMH** : *Branched Multiple Hairpins*

**BSA** : *Bovin Serum Albumin*

**BzCN** : Benzoyl cyanide

**CA** : Capside

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

**CD4** : *Cluster of Differentiation 4*

**CDC** : *Center for Disease Control*

**CLIP** : *Cross-linking immunoprecipitation*

**CRM1** : *Chromosomal maintenance 1*

**Cryo-EM** : Cryomicroscopie électronique

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

**CypA** : Cyclophiline A

**Da** : Dalton

**ddC** : didésoxyCytidine

**ddGTP** : didésoxyGuanine

**ddTTP** : didésoxyThymidine Tri-Phosphate

**DIS** : *Dimerisation initiation site*

**DLS** : *Dynamic Light Scattering*

**DMSO** : Dimethyl sulfoxide

**DNase** : Désoxyribonucléase

**dNTP** : désoxyNTP

**DSIF** : *DRB sensitivity-inducing factor*

**DTT** : dithiothréitol

**Env** : Enveloppe

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

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

**h** : heure

**H<sub>2</sub>O MQ** : eau Milli-Q

**hSHAPE** : *high-throughput Selective 2'Hydroxyl Acylation analysed by Primer Extension*

**IN** : Intégrase

**IRES** : Internal Ribosome Entry Site

**Kb** : Kilobase

**KDa** : Kilo Dalton

**kpb** : kilo paire de base

**LDI** : *Long distance interaction*

## ABREVIATIONS

**LEDGF** : *Lens epithelium-derived growth factor*

**LTR** : *Long terminal repeat*

**Lys** : Lysine

**M** : molaire

**MA** : Matrice

**MIME** : *Mutational interference mapping experiment*

**min** : minute

**mL** : millilitre

**mM** : millimolaire

**MOHCA-seq** : *Multiplexed •OH Cleavage Analysis with paired-end sequencing*

**MQ** : milli-Q

**NC** : Nucléocapside

**Nef** : *Negative factor*

**NELF** : *Negative elongation factor*

**NFAT** : *Nuclear Factor of activated T-cells*

**NFkB** : *Nuclear factor kappa-light-chain-enhancer of activated B cells*

**ng** : nanogramme

**NLS** : Nuclear Localization Signal

**nm** : nanomètre

**nM** : nanomolaire

**NMD** : *Nonsense-mediated decay*

**NMIA** : N-methyl-nitroisatoic anhydride

**NPC** : Complexe de pore nucléaire

**NTP** : Nucléoside Tri-Phosphate

**nts** : nucléotides

**NUP** : Nucléopore

**OMS** : Organisation Mondiale de la Santé

**ORF** : *Open Reading Frame* ou Cadre de lecture ouvert

**p-TEFb** : *positive transcription elongation factor*

**pb** : paire de base

**PBS** : *Primer Binding Site*

**PCR** : *Polymerase Chain Reaction*

**PDB** : *Protein Data Bank*

**pH** : potentiel hydrogène

**PIC** : *Pre-integration complex*

**PIP2** : Phosphatidylinositol 4,5-bisphosphate

**pmoles** : picomoles

**Pol** : Polymérase

**Pol II** : ARN polymérase II

**PPT** : *PolyPurine Tract*

**PR** : Protéase

**PR160<sup>Gag-Pol</sup>** : Précurseur de la polyprotéine de 160 KDa Gag-Pol

**Pr55<sup>Gag</sup>** : Précurseur de la polyprotéine de 55 KDa Gag

**Psi** : *Packaging signal*

**R** : Séquence répétée

**Ran-GTP** : *ras related nuclear protein*

**Rev** : *Regulator of expression of virion proteins*

**RMN** : Résonance magnétique nucléaire

**RNase** : Ribonucléase

**rpm** : Revolutions per minute

**RRE** : *Rev Response Element*

**RT** : *Reverse Transcriptase*

## ABREVIATIONS

**SAXS** : Diffraction des rayons X aux petits angles

**SD** : Site donneur majeur de l'épissage

**SHAPE** : *Selective 2'Hydroxyl Acylation analysed by Primer Extension*

**SIDA** : Syndrome d'immunodéficience acquise

**SIV** : Virus de l'Immunodéficience Simienne

**SL** : *Stem-Loop* ou Tige-boucle

**SMD** : *Staufen1-mediated decay*

**Sp1** : *Specificity protein 1*

**SP1/2** : *Spacer peptide -1/2*

**SSIV** : SuperScript IV

**SU-gp120** : glycoprotéine de surface de 120 KDa

**TAR** : *Trans-activation response element*

**Tat** : *Trans-activator of transcription*

**TBM** : Tris-Borate-MgCl<sub>2</sub>

**TM-gp41** : glycoprotéine transmembranaire de 41 KDa

**TSG101** : *Tumor Susceptibility Gene 101*

**U** : unité

**U3** : Région unique en 3'

**U5** : Région unique en 5'

**UNG2** : U N-Glycolase 2

**UPF1** : *Up Frameshift protein 1*

**UTR** : *Untranslated Region* ou Région non traduite

**Vif** : *Viral infectivity factor*

**VIH-1** : Virus de l'Immunodéficience humaine de type 1

**Vpr** : *Viral protein R*

**Vpu** : *Viral protein U*

**WC** : Watson-Crick

**XRC** : Cristallographie aux rayons X

**µg** : microgramme

**µL** : microlitre

**µm** : micromètre

**µM** : micromolaire



# Liste des figures

- Figure 1** : Prévalence mondiale de l'infection par le VIH-1 et pourcentages de décès causés par l'infection
- Figure 2** : Structures de la particule virale immature et mature, ainsi que ses composants
- Figure 3** : Organisation du génome viral
- Figure 4** : Représentation schématique des sites donneurs et accepteurs d'épissage, ainsi que des ARN viraux du VIH-1 issus des différents événements d'épissage
- Figure 5** : Représentation schématique du cycle viral du VIH-1
- Figure 6** : Représentation schématique du mécanisme de rétrotranscription de l'ARNg du VIH-1
- Figure 7** : Représentation schématique de la cascade protéolytique de Pr55<sup>Gag</sup> et évolution morphologique de la particule virale
- Figure 8** : Représentation schématique des régions de l'ARNg impliquées dans le cycle viral
- Figure 9** : Premiers modèles structuraux des tiges-boucles TAR et PolyA de l'ARNg du VIH-1
- Figure 10** : Premiers modèles structuraux de la région RRE présente dans le gène *env*
- Figure 11** : Premiers modèles structuraux pour la région du frameshift de Gag-Pol
- Figure 12** : Représentation du domaine PBS seul ou en complexe avec l'ARNt<sup>Lys,3</sup>
- Figure 13** : Premiers modèles de la tige-boucle SL1 du domaine Psi
- Figure 14** : Premiers modèles de la région contenant le domaine d'encapsidation Psi
- Figure 15** : Premières structures de la région 5' terminale de l'ARNg du VIH-1
- Figure 16** : Évolution des modèles structuraux de la région 5' terminale de l'ARNg du VIH-1
- Figure 17** : Représentation des premiers modèles de conformères et des interactions à longue distance
- Figure 18** : Représentation des modèles de conformères proposés par l'équipe de Summers
- Figure 19** : Représentation du modèle d'hétérotranscription proposé par Kharytonchyk *et al.*, (2016)
- Figure 20** : Représentation des modèles structuraux issus de données de cartographie *in vivo*
- Figure 21** : Approches méthodologiques employées pour l'étude de la région 5' de l'ARNg du VIH-1
- Figure 22** : Représentation de la carte plasmidique de pNL4-3 clone 159 et des profils de digestion avec *PvuII*
- Figure 23** : Purification des ARN transcrits *in vitro* sur colonne d'exclusion de taille
- Figure 24** : Vérification de la formation des complexes ribonucléoprotéiques entre l'ARN et Pr55<sup>Gag</sup> par retard sur gel d'agarose, TBM 0,5x à 4°C
- Figure 25** : Vérification de la qualité des complexes ribonucléoprotéiques formés avec le nouveau lot de Pr55<sup>Gag</sup>
- Figure 26** : Vérification de l'agrégation des protéines Pr55<sup>Gag</sup> par une approche de DLS
- Figure 27** : Expériences de formation de complexes entre l'ARNg 1-604 et les protéines Pr55<sup>Gag</sup>
- Figure 28** : Distributions des réactivités SHAPE de l'ARNg 1-604 en absence de protéine
- Figure 29** : Comparaison des données obtenues en hSHAPE aux sites de fixation proposés par Nikolaitchik *et al.*, (2020)
- Figure 30** : Comparaison des données obtenues en hSHAPE à des données issues de la littérature

## LISTES DES FIGURES

- Figure 31** : Représentation schématique de l'approche du MOHCA-seq
- Figure 32** : Représentation de réactifs principaux de l'approche de MOHCA-seq
- Figure 33** : Représentation de la sphère d'effet des radicaux hydroxyles produits en réaction au fer
- Figure 34** : Représentation schématique des biais de l'approche du MOHCA-seq
- Figure 35** : Représentation schématique des approches de préparation de banque de séquences réalisées
- Figure 36** : Analyse sur gel d'agarose des résultats préliminaires issus du kit NEBNext®
- Figure 37** : Analyse des profils des banques de séquençage obtenus sur Bioanalyser
- Figure 38** : Modélisation des données de séquençage à haut-débit
- Figure 39** : Validation de l'incorporation des nucléotides ATP 2'NH<sub>2</sub> et de l'ITCB-EDTA
- Figure 40** : Validation de l'étape de clivage aux radicaux hydroxyles
- Figure 41** : Vérification de l'impact du temps de réaction sur le clivage des ARN modifiés ou non
- Figure 42** : Gel théorique des résultats du contrôle au Bioanalyser pour les banques de séquences réalisées à l'aide du kit NEBNext®
- Figure 43** : Modélisation des données de séquençage à haut-débit sous forme de cartes bidimensionnelles
- Figure 44** : Contrôle de la qualité des ARN produits *in vitro* et validation de l'étape de ligation en 3' sur une puce RNA 6000 Nano au Bioanalyser
- Figure 45** : Validation sur une puce RNA 6000 Nano au Bioanalyser de l'utilisation de la soude (NaOH) pour la dégradation de l'ARN matriciel
- Figure 46** : Vérification sur gel d'agarose de l'impact du traitement à l'ExoI sur la seconde ligation lors de la préparation de la banque
- Figure 47** : Validation de la séparation des ARN en fonction de la taille et de la seconde ligation sur les ADNC
- Figure 48** : Schéma représentatif des structures attendues dans les conditions expérimentales testées en MOHCA-seq
- Figure 49** : Impact de l'oligonucléotide antisens U5 (@U5) sur la dimérisation
- Figure 50** : Impact de l'oligonucléotide antisens AUG (@AUG) sur la dimérisation
- Figure 51** : Vérification de la pureté de l'ARN modifié 2'NH<sub>2</sub> et des clivages aux radicaux hydroxyles lors du protocole de MOHCA-seq
- Figure 52** : Contrôles externes de chaque étape de la préparation des banques de séquences pour le MOHCA-seq
- Figure 53** : Contrôle des amplifications PCR associées à chaque condition de MOHCA-seq
- Figure 54** : Vérification de la qualité des banques de séquences après purification sur gel



# Liste des tableaux

**Tableau 1** : Les protéines du VIH-1, fonctions et partenaires

**Tableau 2** : Séquences des amorces utilisées en hSHAPE

**Tableau 3** : Coefficients de corrélation entre les moyennes de réactivités SHAPE des conditions en absence (NoProt) et présence (10 Gag, 40 Gag et 60 Gag) de la protéine Pr55<sup>Gag</sup>

**Tableau 4** : Comparaison 2 à 2 des données de réactivités pour les conditions avec ou sans glycérol, repliées dans un tampon HEPES (pH 7,5 ou 8,0) et modifiées au NMIA ou au BzCN

**Tableau 5** : Comparaison 2 à 2 des données de réactivités pour les conditions avec ou sans Pr55<sup>Gag</sup>, repliées dans un tampon HEPES (pH 7,5 ou 8,0) et modifiées au NMIA ou au BzCN.

**Tableau 6** : Tests de Wilcoxon entre la condition « NoProt » et les conditions en présence de protéines pour l'ARNg 1-604 en fonction de la nature des nucléotides

**Tableau 7** : Tests de Wilcoxon entre la condition « NoProt » et les conditions en présence de protéines pour l'ARNg 1-604 en fonction de la nature des nucléotides et des domaines NoPsi et Psi

**Tableau 8** : Séquences des oligonucléotides utilisés lors du MOHCA-seq

**Tableau 9** : Récapitulatif des conditions étudiées en MOHCA-seq et des index utilisés pour chaque échantillon

**Tableau 10** : Table de la Loi de Poisson de paramètre  $\lambda$  et d'espérance  $k$ , représentant la probabilité d'incorporation de nucléotides 2'NH<sub>2</sub>

**Tableau 11** : Résultats préliminaires de l'analyse des banques de séquences réalisées par la plateforme EpiRNA-Seq

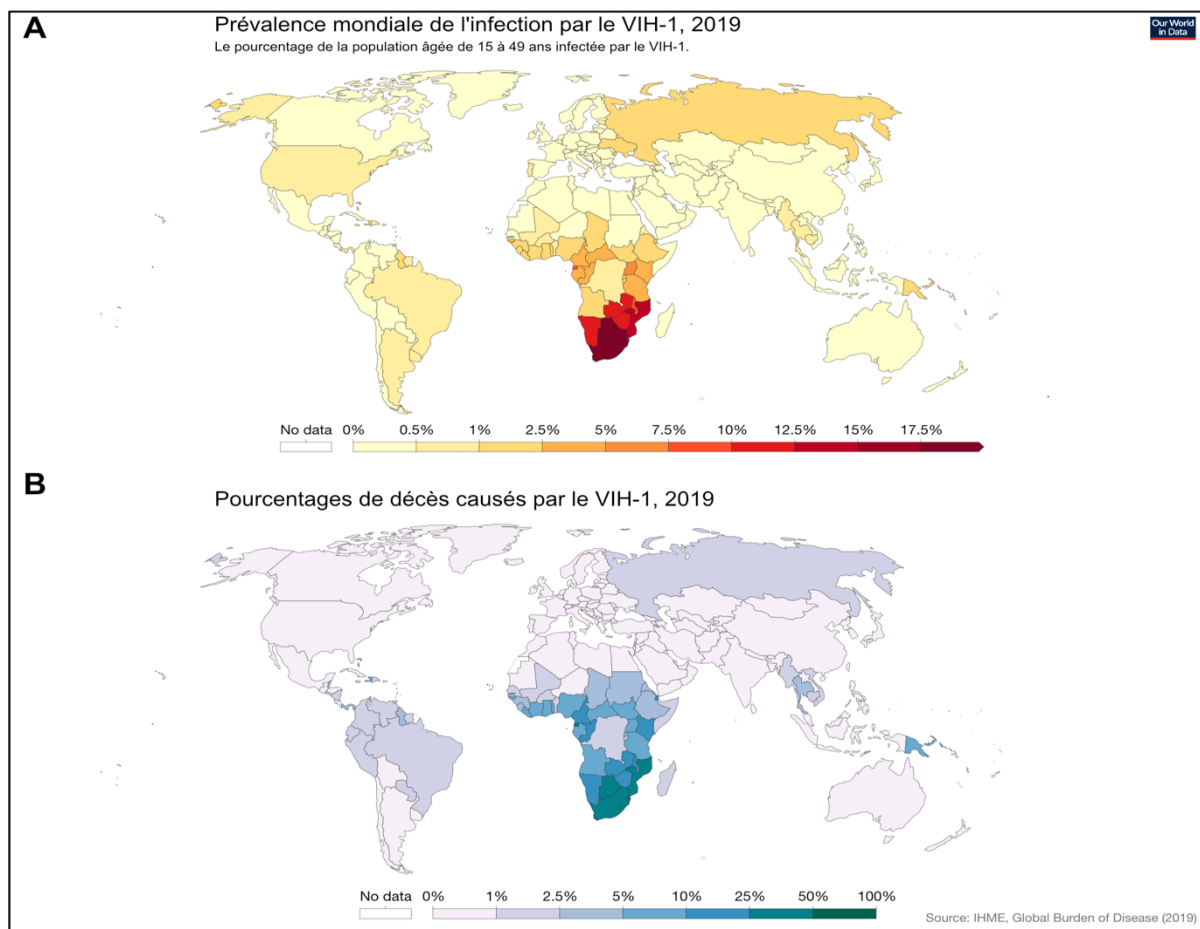


# INTRODUCTION

## I/ VIH-1 : généralités

### I.1/ Épidémiologie

En 1981, la pandémie de syndrome d'immunodéficience acquise (SIDA) éclate avec les premiers cas rapportés par le *Center for Disease Control* (CDC). Deux ans plus tard, l'agent étiologique est isolé à partir d'un ganglion de patient infecté et identifié comme appartenant à la famille des rétrovirus (1). Depuis le début de la pandémie, le nombre de décès causés par le virus est estimé à 40,1 millions par l'Organisation Mondiale de la Santé (OMS). De plus, 38,4 millions de personnes vivaient toujours avec le virus de l'immunodéficience humaine de type 1 (VIH-1) en 2021, faisant de cette infection un problème de santé majeur dans le monde (Figure 1).



**Figure 1 :** (A) Prévalence mondiale de l'infection par le VIH-1 et (B) pourcentages de décès causés par l'infection.

## INTRODUCTION

### 1.2/ Taxonomie

Les *Retroviridae* sont une grande famille de virus enveloppés et possédant un génome à acide ribonucléique (ARN) qui est rétrotranscrit en acide désoxyribonucléique (ADN) lors de l'infection et intégré dans le génome de la cellule infectée. Cette famille peut être séparée en deux catégories, les rétrovirus simples et les rétrovirus complexes, que l'on distingue principalement par leur génome (2). En effet, l'ensemble des rétrovirus contiennent les séquences codant les protéines *gag* (*group-specific antigen*), *pol* (*polymerase*) et *env* (*envelope*), responsables de la synthèse des protéines structurales, enzymatiques et d'enveloppe, respectivement. En revanche, les rétrovirus complexes codent également pour des protéines régulatrices supplémentaires, généralement issues d'ARN messagers (ARNm) épissés. De plus, la famille des rétrovirus peut être subdivisée en 7 genres, définis par leur parenté évolutionnaire (ICTV, 2022).

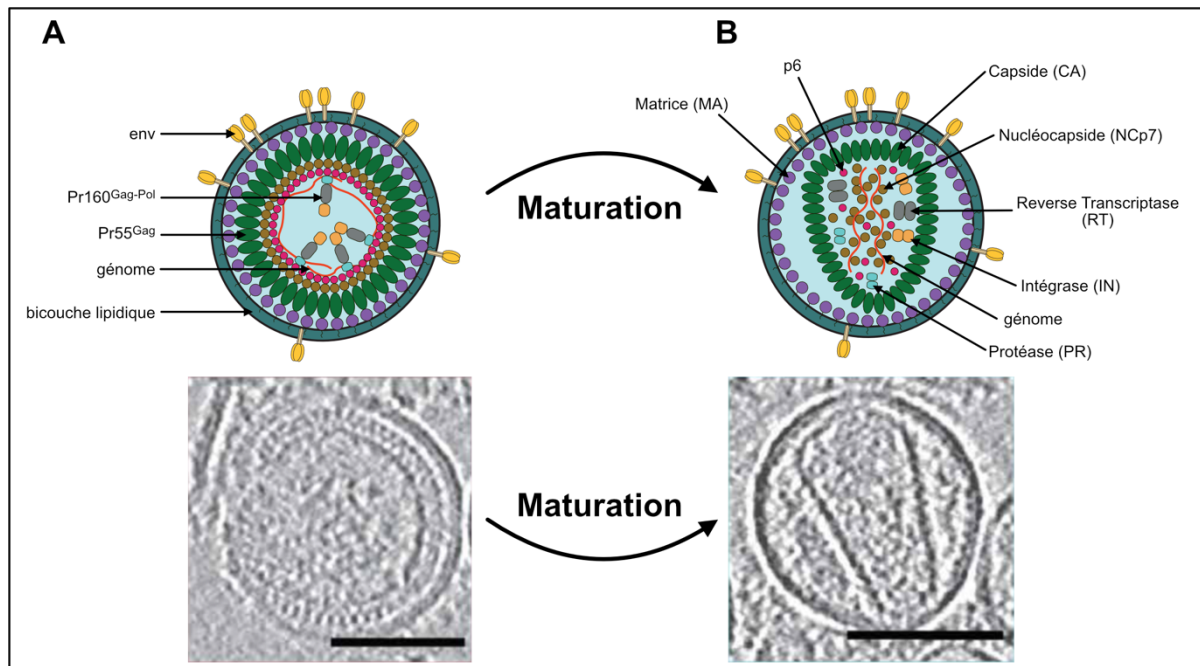
Le VIH-1, membre du genre *Lentivirus*, est un rétrovirus complexe divisé en 4 groupes. Chaque groupe provient d'un événement de franchissement d'espèces. Ainsi, les groupes M (« main ») et N (« non-M, non-O ») trouvent leur origine chez un virus de l'immunodéficience simienne (SIV) issu du chimpanzé, et les groupes O (« outlier ») et P chez un SIV infectant le gorille (3,4). Les différents isolats responsables de la pandémie actuelle sont regroupés dans le groupe M.

### 1.3/ Structure de la particule virale

Les virions produits par le VIH-1 sont des particules sphériques d'approximativement 120 nm, exposant des trimères de glycoprotéines virales enchâssés dans une bicouche lipidique. Initialement, ces particules bourgeonnent de la membrane des cellules infectées sous forme de virions immature (5). A l'intérieur de l'enveloppe, entre 2500 et 5000 protéines virales structurales Pr55<sup>Gag</sup> et Pr160<sup>Gag-Pol</sup> forment un maillage à la membrane plasmique (5,6). Cependant, ces particules immatures ne sont pas infectieuses (7) (**Figure 2A**). En effet, l'infectivité virale n'est acquise qu'après la maturation des particules. Cette étape est médiée par l'activation de la protéase virale et consiste en une cascade protéolytique des différents précurseurs Pr55<sup>Gag</sup> et Pr160<sup>Gag-Pol</sup>, ainsi qu'au réarrangement du dimère d'ARN génomique (ARNg). A la fin de l'étape de maturation, les particules virales infectieuses sont compétentes

## INTRODUCTION

pour initier un nouveau cycle réplcatif. Ces particules sont morphologiquement différentes des particules immatures avec la présence d'un noyau coniforme composé de protéines de capsidie et encapsulant un génome dimérique compacté (**Figure 2B**), appelé capsidie virale (8). La capsidie renferme également les protéines de nucléocapsidie en interaction avec le génome, les protéines enzymatiques avec la *reverse transcriptase* (RT), l'intégrase (IN) et la protéase (PR), et d'autres protéines virales comme la *Viral protein R* (Vpr), le *Viral infectivity factor* (Vif) ou encore le *negative factor* (Nef) (9).



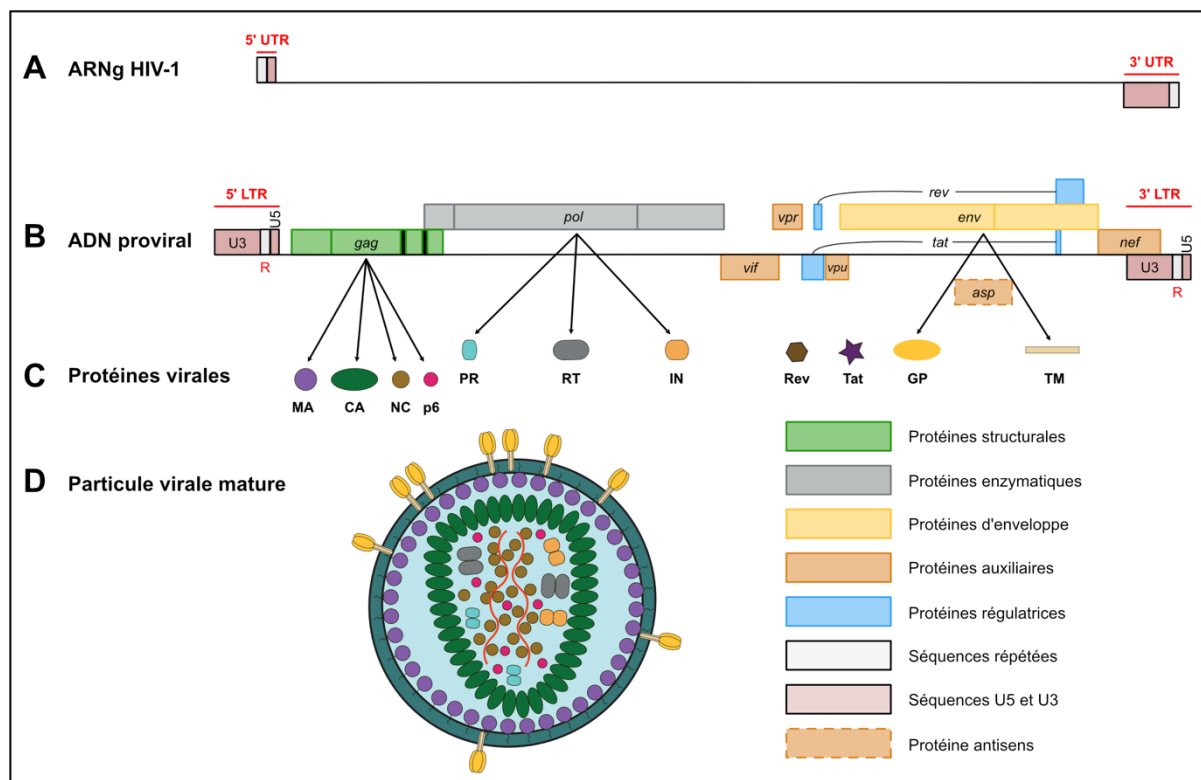
**Figure 2** : Structures de la particule virale (A) immature et (B) mature, ainsi que ses composants. (B) La particule mature contient la structure caractéristique en forme de cône de la nucléocapsidie. Les clichés de cryotomographie électronique sont issus de Mallery *et al.*, (2021) (10). L'échelle correspond à 100 nm. Les protéines Vif, Vpr et Nef ne sont pas représentées.

### 1.4/ Organisation du génome

Le génome du VIH-1 est constitué de deux copies monocaténares d'ARN de polarité positive, coiffées et polyadénylées. Les deux extrémités du génome viral sont flanquées de séquences répétées (R), ainsi que de séquences uniques en 5' (U5) et en 3' (U3), constituant les régions non traduites (5' et 3' UTR) (**Figure 3A**). Lors de la transcription inverse, les régions U5 et U3 sont dupliquées pour être présentes à chaque extrémité dans les *Long Terminal Repeat* (LTR) (**Figure 3B**). L'ADN viral, dont la séquence a été obtenue en 1985 (11), est ensuite intégré dans le génome de la cellule hôte (12). Les séquences U3 rétrotranscrites en 5' de

## INTRODUCTION

l'ADN viral contiennent les séquences promotrices régulant la transcription des ARN viraux (13). Les régions R et U5 contiennent divers éléments impliqués dans la régulation de différentes étapes du cycle réplcatif (14). Au sein de la séquence codante, 10 *Open Reading Frame* (ORF) codent pour les 16 protéines virales (**Figure 3B**). Les gènes *gag*, *pol* et *env* codent pour trois précurseurs polyprotéiques. Les protéines régulatrices sont encodées par les gènes *tat* et *rev* et sont nécessaires au cycle viral. Les gènes *vif*, *vpr*, *vpu*, *nef* et *asp* permettent l'expression des protéines auxiliaires. Le gène codant la protéine ASP (AntiSense Protein) est localisé sur le brin antisens du génome proviral, chevauchant le gène *env* (15). La fonction de cette protéine, bien qu'inconnue à ce jour, présente un avantage sélectif pour le virus (16). Le **Tableau 1** récapitule les fonctions principales portées par chacune de ces protéines.



**Figure 3** : Organisation du génome viral. (A) L'ARNg viral, long de 9,2 kb, est flanqué des régions 5' et 3' UTR et sert de matrice pour la transcription inverse en ADN. (B) Une fois intégré, le provirus de 9,8 kpb est transcrit en ARN qui sera épissé ou non pour permettre l'expression de 10 ORF. La région codante est flanquée par des séquences identiques (5' et 3' LTR) dupliquées lors de la transcription inverse. (C) Les protéines virales impliquées dans le cycle réplcatif et/ou retrouvées dans (D) la particule virale sont représentées, à l'exception de Vif, Vpr et Nef.

## INTRODUCTION

Protéine virale	Fonctions	Partenaires
<b>Matrice, MA</b>	Cible la membrane plasmique pour l'assemblage, incorpore les protéines d'enveloppe, impliquée dans des processus post-entrée	Protéine d'enveloppe TM, Phosphatidylinositol 4,5-bisphosphate (PIP2), ARNt
<b>Capside, CA</b>	Multimérise pendant l'assemblage, forme la capside coniforme dans les particules matures, participe à l'adressage du génome au noyau	Capside CA, cyclophiline A, CSFP6, nucléoporines et transportines
<b>Nucléocapside, NC</b>	Sélection spécifique du génome dimérique, encapsidation, chaperon de l'ARN	ARN génomique, ARNt, APOBEC3G/F
<b>p6</b>	Bourgeoisement, spécificité de sélection de l'ARNg	ESCRT complex, Vpr
<b>Protéase, PR</b>	Maturation protéolytique de Gag et Gag-Pol	Protéase PR, Pr55 <sup>Gag</sup> , Pr160 <sup>Gag-Pol</sup>
<b>Rétro transcriptase, RT</b>	Synthèse de l'ADN viral, dégradation de l'ARNg par le domaine RNase H	ARNt <sup>Lys,3</sup> , ARNg, Intégrase IN
<b>Intégrase, IN</b>	Insertion de l'ADN viral dans le génome de la cellule hôte	Intégrase IN, Rétro Transcriptase RT, LEDGF/p75, UNG2, ADN viral
<b>Glycoprotéine de surface, SU, gp120</b>	Reconnaissance des récepteurs membranaires CD4, permet l'entrée du virus dans la cellule	Protéine d'enveloppe TM, CD4, CCR5, CXCR4,
<b>Glycoprotéine transmembranaire, TM, gp41</b>	Contient le peptide de fusion, permet la fusion des membranes virales et cellulaires	Protéine d'enveloppe SU, Matrice MA
<b>Viral infectivity factor, Vif</b>	Inhibe l'expression des facteurs de restriction APOBEC3G/3F	ARNg, APOBEC3G/3F, complexe ubiquitine ligase,
<b>Viral protein R, Vpr</b>	Améliore l'infectivité viral, arrêt du cycle G2/M	p6, UNG2, nucléoporines
<b>Trans-Activator of Transcription, Tat</b>	Activation de la transcription virale par phosphorylation du CTP de la Pol II	ARNg (TAR), p-TEFb
<b>Regulator of expression of virion proteins, Rev</b>	Régule l'export des ARN viraux partiellement ou non épissés	ARNg (RRE), CMR1, Exportine 1, Importine β
<b>Viral protein U, Vpu</b>	Diminution du nombre de récepteurs CD4, facilite le bourgeoisement des particules virales	CD4
<b>Negative Factor, Nef</b>	Diminution du nombre de récepteurs CD4, améliore l'infectivité virale, bloque l'apoptose	CD4, kinases, dynamine 2
<b>Antisense Protein, Asp</b>	Rôle putatif dans l'induction de l'autophagie Inconnu	Inconnu

**Tableau 1 :** Les protéines du VIH-1, fonctions et partenaires. Adapté de Swanson et Malim, (2008) (17). La liste de ces fonctions et partenaires n'est pas exhaustive.

Les précurseurs polyprotéiques Pr55<sup>Gag</sup> et Pr160<sup>Gag-Pol</sup> contiennent les différentes protéines structurales et enzymatiques, respectivement (**Figure 3C**). La maturation protéolytique débute par l'activation de la protéase virale PR lors de sa dimérisation. Après son auto-excision de Pr160<sup>Gag-Pol</sup>, PR va induire la maturation protéolytique des deux précurseurs polyprotéiques. Les clivages successifs de Pr55<sup>Gag</sup> permettent l'individualisation

## INTRODUCTION

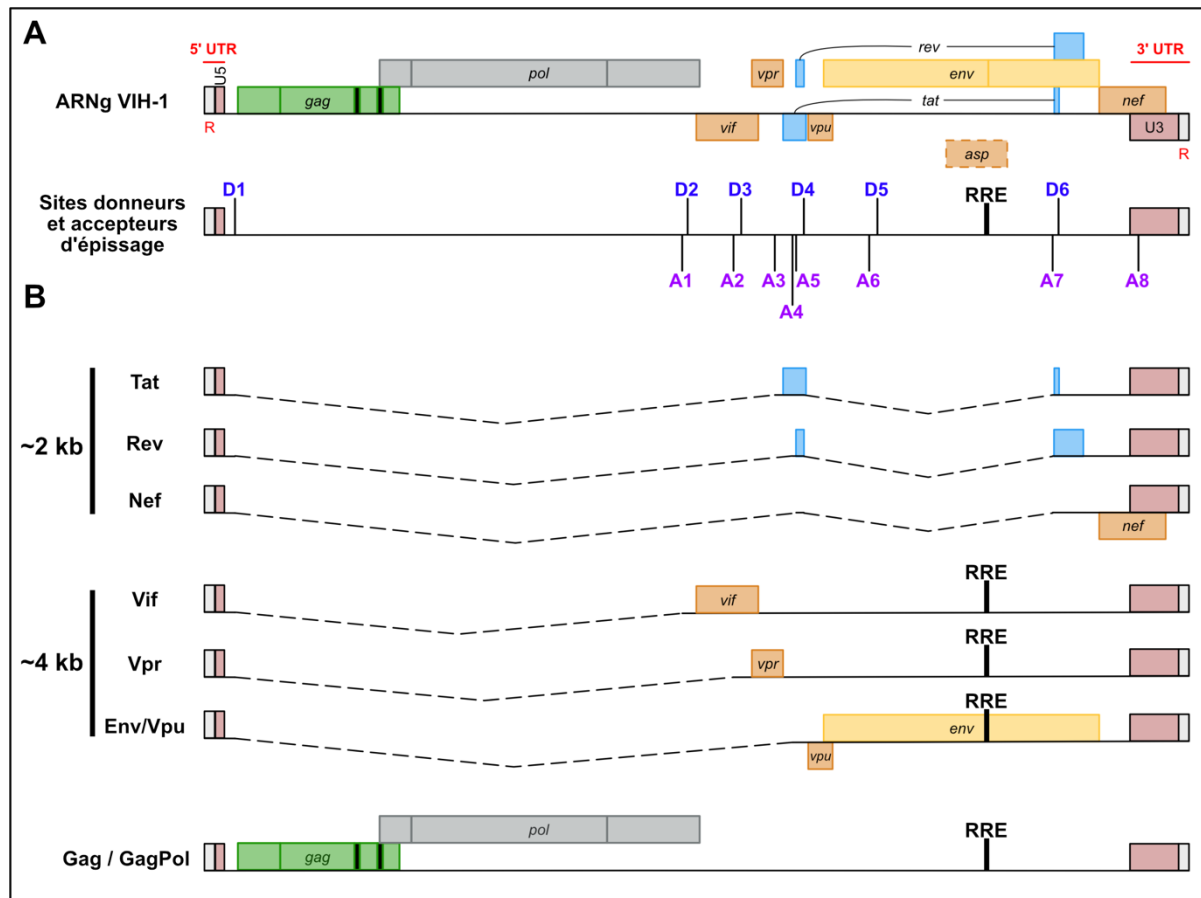
des domaines de matrice (MA), capsid (CA), nucléocapsid (NC) et p6. De la même façon, Pr160<sup>Gag-Pol</sup> est maturée pour singulariser la protéase (PR), la rétrotranscriptase (RT) et l'intégrase (IN). Ces clivages protéolytiques sont réalisés au sein de la particule virale lors de l'étape de maturation. Comme Pr55<sup>Gag</sup> et Pr160<sup>Gag-Pol</sup>, les protéines d'enveloppe sont synthétisées sous forme de précurseur polyprotéique avec gp160. En revanche, cette polyprotéine est maturée dans le Golgi par la furine, une protéase cellulaire, pour générer les glycoprotéines d'enveloppe de surface (SU – gp120) et transmembranaires (TM – gp41) (18). Les protéines maturées sont toutes retrouvées dans la particule virale après le bourgeonnement. Cependant, parmi les 7 autres protéines synthétisées par le VIH-1, seules les protéines Vif, Vpr et Nef sont également encapsidées (8) (**Figure 3D**).

Lors de la transcription, un unique ARN de 9,2 kb est produit par le provirus. Cet ARN polycistronique contient différents sites donneurs et accepteurs d'épissage (**Figure 4A**) qui permettent l'obtention d'une centaine d'ARN épissés alternatifs (19,20). Parmi l'ensemble de ces ARN codant pour les différentes protéines virales, trois catégories d'ARN messagers se dessinent (**Figure 4B**). La première catégorie comprend les ARN complètement épissés, d'environ 2 kb, qui codent pour les protéines *Transactivator of Transcription* (Tat), *Regulator of expression of virion proteins* (Rev) et Nef. Ces protéines sont synthétisées en priorité. En effet, Tat joue un rôle dans la transcription des ARN viraux et Rev est nécessaire à l'export nucléaire des ARN contenant des introns via le *Rev Response Element* (RRE). La protéine Nef est impliquée, entre autres, dans la régulation négative des protéines réceptrice *Cluster of differentiation 4* (CD4), permettant la reconnaissance des particules virales de leur cellules cibles. La seconde catégorie inclue les ARN partiellement épissés de taille approximative de 4 kb. Ces ARN permettent la traduction des protéines Vif, Vpr, Vpu et Env. Vif est requise pour l'inhibition des facteurs de restriction cellulaires *Apolipoproteins B mRNA editing enzyme, catalytic polypeptide-like 3F/3G* (APOBEC3F/3G) (21). Vpu accompagne Nef dans sa fonction de diminution du nombre de protéine CD4 et permet le bourgeonnement des particules virales en contrecarrant l'action du facteur de restriction BST2 (22). Vpr est importante lors des étapes précoces du cycle répliatif. Les protéines d'enveloppe (Env) sont requises pour les étapes de reconnaissance et d'entrée. Finalement, la dernière catégorie compte les ARN non épissés, ou ARN génomiques (ARNg) de 9,2 kb, qui servent de matrice à la synthèse des



## INTRODUCTION

protéines structurales (Pr55<sup>Gag</sup>) et enzymatiques (Pr160<sup>Gag-Pol</sup>), mais qui peuvent également être encapsidés sous forme de dimère dans les particules virales.



**Figure 4 :** Représentation schématique des sites donneurs et accepteurs d'épissage, ainsi que des ARN viraux du VIH-1 issus des différents événements d'épissage. (A) L'ARNg du VIH-1, non épissé, contient les séquences de l'ensemble des gènes viraux. Les principaux sites donneurs (D) et accepteurs (A) d'épissages sont localisés sur le génome. (B) Les différentes catégories d'ARN viraux codant les protéines sont représentées avec les ARN épissés (~2 kb), les ARN partiellement épissés (~4 kb) et l'ARNg non épissé (~9,2 kb). Le *Rev Responsive Element* (RRE) est indiqué.

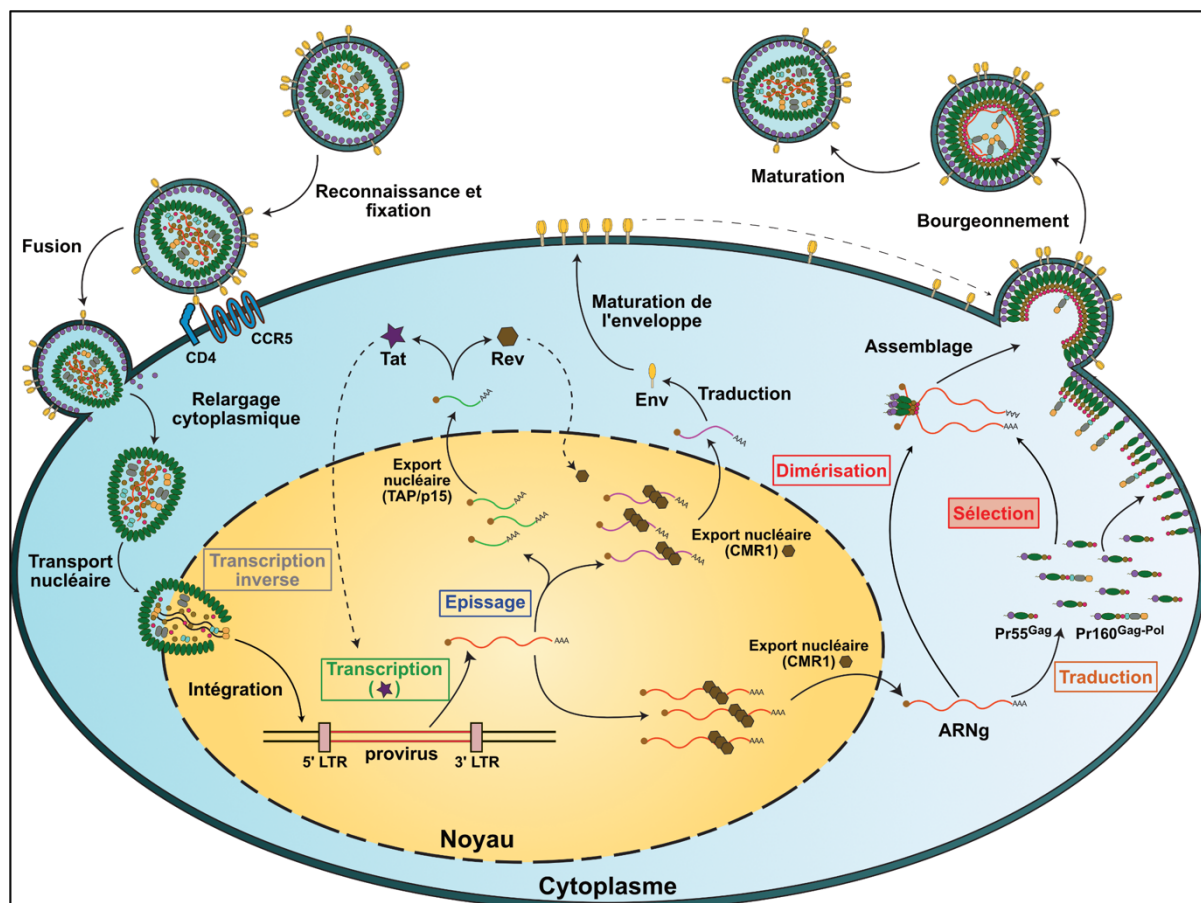
### 1.5/ Cycle viral

Comme l'ensemble des virus, le VIH-1 est un parasite intracellulaire obligatoire dont la réplication passe par le détournement des machineries cellulaires des hôtes infectés. Les cibles principales d'une infection par le VIH-1 sont les cellules présentant des récepteurs CD4 à la surface de leur membrane, tels que les lymphocytes T CD4+, mais également les macrophages et les cellules dendritiques (23). L'infection débute avec la fusion de la particule

## INTRODUCTION

virale à la membrane plasmique de la cellule hôte et termine avec le bourgeonnement de particules virales qui initieront d'autres cycles réplicatifs (**Figure 5**).

Ces cycles peuvent être divisés en deux phases appelées précoce et tardive. Lors de la phase précoce, une particule virale mature va reconnaître et fixer les récepteurs CD4 à l'aide des protéines d'enveloppe gp120, avant d'initier une fusion des membranes (24). La capsid virale est alors relarguée dans le cytoplasme puis dirigée vers le noyau (25). La transcription inverse de l'ARNg en ADN viral est complétée et l'ADN viral est intégré dans le génome de la cellule hôte sous forme de provirus (26,27). La phase tardive débute avec la transcription d'un ARNg de 9,2 kb. Cet ARNg est épissé ou non avant d'être exporté dans le cytoplasme où il sert de matrice à la synthèse des protéines virales. L'ARNg peut également former un homodimère dans le cytoplasme qui est sélectionné par des protéines structurales virales. Le cycle réplicatif se termine avec l'encapsidation du dimère d'ARNg dans une particule virale, son bourgeonnement et sa maturation dans le milieu extracellulaire (28).



**Figure 5** : Représentation schématique du cycle viral du VIH-1. Le cycle débute par la reconnaissance et la fixation d'une particule mature à la membrane d'une cellule cible. Après fusion membranaire, la nucléocapside virale est relarguée dans le cytoplasme et transportée vers le noyau. Le génome viral est rétrotranscrit et intégré dans le

## INTRODUCTION

génomique de la cellule hôte sous forme de provirus. La phase tardive débute avec la transcription d'un ARNg viral qui sera, ou non, épissé. Les ARN viraux épissés permettent la synthèse des protéines d'enveloppe, régulatrices et auxiliaires. Les ARN non épissés sont exportés dans le cytoplasme pour servir de matrice à la traduction des protéines structurales et enzymatique, avant d'être sélectionnés sous forme de dimères pour être encapsidés dans une particule naissante. Après le bourgeonnement, la particule virale subit une étape de maturation pour acquérir son infectivité et initier un nouveau cycle. Les étapes colorées sont celles régulées par des motifs structuraux de l'ARNg (**Figure 8**).

### I.5.1/ Phase précoce

#### *I.5.1.1/ Reconnaissance, fusion et entrée*

La première étape de l'infection d'une nouvelle cellule cible par le VIH-1 débute par la reconnaissance du récepteur CD4 et d'un co-récepteur *C-C chemokine receptor type 5* (CCR5) ou *C-X-C chemokine receptor type 4* (CXCR4) selon le tropisme cellulaire (29). Cette reconnaissance est médiée par le trimère de glycoprotéine de surface, gp120, et va induire un changement conformationnel menant à l'exposition d'un peptide de fusion (24). Le peptide de fusion est inséré dans la membrane plasmique et forme un domaine de 6 hélices qui rapproche les deux membranes, provoquant leur fusion. L'étape d'entrée s'achève lorsque la nucléocapside virale est libérée dans le cytoplasme.

#### *I.5.1.2/ Import nucléaire, décapsidation et transcription inverse*

La libération de la nucléocapside dans le cytoplasme se déroule dans la périphérie de la cellule, au niveau de la membrane. La nucléocapside va s'attacher au réseau de microtubules et être activement transportée vers le noyau. A ce jour, trois modèles permettant de décrire les étapes de décapsidation et d'import ont été proposés (25). Le premier postule une décapsidation cytoplasmique débutant lors de l'entrée et une transcription inverse complète dans le cytoplasme (30). Le complexe de rétrotranscription est alors remodelé en complexe de pré-intégration (PIC) contenant l'ADN double brin (ADNdb) viral (31). Un second modèle propose une décapsidation concomitante à l'entrée nucléaire, lorsque la nucléocapside se fixe au complexe de pore nucléaire (NPC) (32). Dans ce modèle, la transcription inverse est complétée en amont de la décapsidation. Récemment, une dernière hypothèse a été présentée où la nucléocapside pénètre le noyau sous une forme intacte

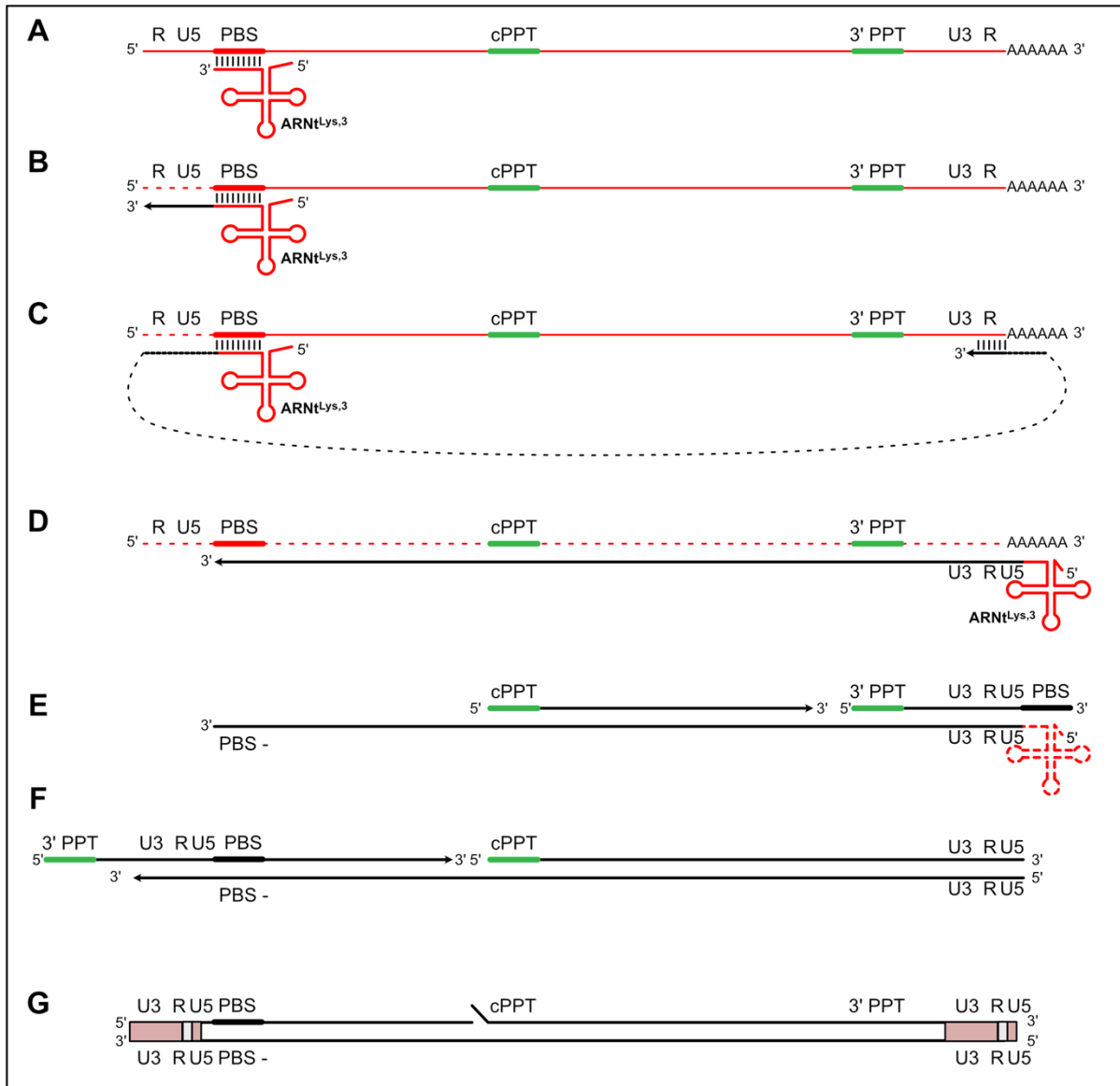
## INTRODUCTION

(26,27). La transcription inverse est finalisée dans le noyau et la décapsidation se déroule à proximité des sites d'intégration du génome viral.

Bien que la chronologie des étapes de décapsidation et de rétrotranscription ne fasse pas encore consensus, il est clair que ces étapes sont nécessaires au cycle viral. Leur réalisation dépend de différents facteurs cellulaires et viraux. En effet, la capsid virale interagit avec la cyclophiline A (CypA) qui stabilise la structure conique (33), mais aussi avec les protéines de nucléoporines 62 (NUP62), 88 (NUP88), 153 (NUP153), 214 (NUP214) et 358 (NUP358), le *Cleavage and Polyadenylation Specificity Factor* (CPSF6) (34) et les transportines 1 (TRN-1) et 3 (TNPO3) (35,36) pour l'import nucléaire.

Le processus de transcription inverse a lieu entre l'entrée de la nucléocapside dans le cytoplasme et la fin de l'import nucléaire. Cette étape consiste en la synthèse d'un ADN viral double brin à partir des deux molécules d'ARNg encapsidées (37,38). Elle repose sur les trois activités enzymatiques portées par la RT virale qui sont celles de rétrotranscriptase, de RNase H et d'ADN polymérase ADN dépendante. La transcription inverse débute par l'hybridation des 18 nucléotides en 3' de l'ARN de transfert (ARNt) pour l'isoforme 3 de la lysine (ARNt<sup>Lys,3</sup>) sur une région en aval de U5, appelée *Primer Binding Site* (PBS). La RT virale ajoute alors séquentiellement des nucléotides en 3' de l'ARNt jusqu'à atteindre l'extrémité 5' terminale de l'ARNg (**Figure 6A**). L'activité RNase H de la RT agit en parallèle de la synthèse de l'ADN viral, permettant la dégradation de la région 5' (**Figure 6B**). L'ADN viral, contenant les séquences complémentaires aux régions R et U5, subit son premier transfert de brin et s'hybride à l'extrémité 3' de l'ARNg, sur la région R (**Figure 6C**). La synthèse de l'ADNc viral reprend jusqu'à ce que la totalité de l'ARNg soit rétro-transcrit et que la majorité de la matrice soit dégradée (**Figure 6D**). Seules certaines régions riches en purines (cPPT et 3'PPT) sont protégées de la dégradation. Ces courtes séquences servent d'amorces à la synthèse du second brin d'ADN par l'activité ADN polymérase ADN dépendante de la RT virale (**Figure 6E**). Lorsque la partie complémentaire au PBS de l'ARNt<sup>Lys,3</sup> est synthétisée, l'ARNt est dégradé et un second saut de brin est réalisé. Les séquences complémentaires PBS et PBS (-) s'hybrident et la synthèse de l'ADN viral reprend (**Figure 6F**). La synthèse de l'ADN viral se conclut au site de terminaison central, légèrement en amont du cPPT. Le provirus résultant contient les séquences U3 – R – U5 (LTR) aux deux extrémités et une ouverture centrale qui sera réparée par les protéines cellulaires.

## INTRODUCTION



**Figure 6 :** Représentation schématique du mécanisme de rétrotranscription de l'ARNg du VIH-1. (A et B) La synthèse de l'ADN viral commence par l'hybridation de l'ARNt<sup>Lys,3</sup> et qui sert d'amorce pour rétrotranscrire l'extrémité 5' de l'ARNg. (C) La courte séquence d'ADN viral synthétisée réalise un premier saut de brin pour que les séquences complémentaires des régions R s'hybrident. (D) La transcription inverse reprend et l'ADN viral (-) est synthétisé. Des séquences riches en purines, *central PolyPurine Tract* et *3' PolyPurine Tract* (cPPT et 3' PPT), résistent à la dégradation RNase H et (E) servent de matrice à la synthèse du brin d'ADN viral (+). La séquence PBS est copiée avant d'être dégradée avec l'ARNt. (F) Un second saut de brin a lieu pour que les deux séquences PBS interagissent et que la synthèse du double brin d'ADN viral se termine. (G) L'ADN pré-proviral obtenu contient les séquences des 5' et 3' LTR ainsi qu'une zone non liguée en amont du cPPT qui sera réparée par des protéines cellulaires.

Lors du processus de transcription inverse, deux sauts de brins sont réalisés. Bien qu'il soit possible que l'ADN viral soit formé à partir d'une unique molécule d'ARNg, les événements

## INTRODUCTION

de transferts sont le plus souvent réalisés entre les deux molécules d'ARNg, menant à des recombinaisons génétiques (39). Pour finir, le complexe de rétrotranscription implique différentes protéines virales qui peuvent influencer le processus (37). En effet, les protéines NCp7 facilitent la transcription inverse par leur activité chaperon de l'ARN tandis que la protéine Vpr interagit avec les protéines cellulaires *U N-Glycolase 2* (UNG2) impliquée dans les mécanismes de réparation de l'ADN.

### *1.5.1.3/ Intégration*

Après l'import nucléaire et la transcription inverse, l'ADN pré-proviral est intégré dans le génome de la cellule hôte. Cette étape d'intégration est nécessaire afin d'induire une expression importante d'ARN viral (40) et est réalisée par des dimères d'IN fixés sur chaque LTR de l'ADN viral. L'intégration débute par le clivage d'un dinucléotide en 3' de chaque extrémité de l'ADN. Les extrémités 3'OH obtenues induisent une cassure sur l'ADN chromosomique, causée par une attaque nucléophile. L'ADN viral est intégré dans la cassure réalisée et la machinerie cellulaire répare les dégâts, complétant l'intégration du génome pré-proviral. Différentes protéines cellulaires sont également impliquées dans l'intégration génomique de l'ADN viral, comme les protéines *lens epithelium-derived growth factor* (LEDGF/p75) qui dirigent le complexe de pré-intégration vers les régions transcriptionnellement active du génome (41).

### *1.5.2/ Phase tardive*

#### *1.5.2.1/ Expression du génome viral*

Le génome viral intégré, appelé provirus, est long de 9,8 kpb et flanqué par les régions LTR. Le domaine U3 présent dans le 5' LTR contient différents éléments requis pour la transcription de l'ARNg viral qui débute à la jonction entre U3 et R (42). La transcription est réalisée par l'ARN polymérase II (Pol II) cellulaire en association avec d'autres facteurs de transcription cellulaires tels que *Nuclear factor kappa-light-chain-enhancer of activated B cells* (NF- $\kappa$ B), *Nuclear Factor of activated T-cells* (NFAT) et *Specificity protein 1* (Sp1). La transcription basale de l'ARNg du VIH-1 est relativement faible du fait de l'interaction de

## INTRODUCTION

*Negative elongation factor* (NELF) et *DRB sensitivity-inducing factor* (DSIF) qui inhibent l'élongation de la transcription par la Pol II après la transcription des premiers éléments de l'ARN (43,44). Parmi les régions transcrites, la tige-boucle *Transactivation response element* (TAR) est retrouvée. Cette dernière permet le recrutement de la protéine virale Tat qui recrute elle-même le complexe protéique *positive transcription elongation factor* (p-TEFb). La phosphorylation de NELF et de DSIF par p-TEFb force la dissociation des protéines. De même p-TEFb phosphoryle le domaine C terminal de la Pol II, ce qui résulte en une activation de la transcription avec une amélioration de la processivité et une augmentation de la quantité d'ARN viraux produits.

### *1.5.2.2/ Épissage, export nucléaire et traduction*

Comme décrit dans la partie I.4/ Organisation du génome, les ARN viraux peuvent être épissés ou non (**Figure 4B**). Les premiers ARN produits sont complètement épissés par le spliceosome et sont exportés dans le cytoplasme par la voie canonique d'export des ARNm cellulaires, TAP/p15 (45). Ils sont alors pris en charge par la machinerie traductionnelle cellulaire et permettent la synthèse des protéines Tat, Rev et Nef. Tat et Rev sont redirigées vers le noyau par les signaux d'imports nucléaires (NLS) qu'elles portent. Comme décrit précédemment, Tat permet d'augmenter l'expression des ARN viraux et Rev permet l'export nucléaire des ARN viraux partiellement épissés et non épissés par la voie *Chromosomal maintenance 1* ou *Exportine 1* (CRM1) (46,47). Cet export est réalisé par la reconnaissance spécifique du motif RRE et la multimérisation de Rev. Une protéine CRM1 en association à *ras related nuclear protein* (Ran-GTP) permet la translocation des ARN viraux dans le cytoplasme où ils servent de matrice à la synthèse des protéines virales.

### *1.5.2.3/ Assemblage, bourgeonnement et maturation*

Comme les autres rétrovirus, le VIH-1 sélectionne spécifiquement un homodimère d'ARN génomique, non épissé, lors de l'étape d'encapsidation (48,49). La dimérisation de cet ARNg est initiée dans le cytoplasme par une interaction non covalente. Le dimère ainsi formé est spécifiquement reconnu par un petit nombre de protéines Pr55<sup>Gag</sup> et Pr160<sup>Gag-Pol</sup> via leur domaine de NC, parmi l'ensemble des ARN cellulaires et viraux présents dans le cytoplasme

## INTRODUCTION

(49,50). Cette reconnaissance est médiée par le signal d'encapsidation, appelé *Packaging Signal* (Psi ou «  $\Psi$  »), qui est localisé dans la région 5' terminale de l'ARNg du VIH-1. Historiquement, ce signal a été identifié sur une structure en tige-boucle (appelée SL3 ou «  $\Psi$  »), en amont du site donneur d'épissage D1, expliquant l'encapsidation des ARN viraux non épissés (51–53). Cependant, il a été montré que la délétion de cette tige-boucle n'empêchait pas l'encapsidation de l'ARN viral et la région Psi a été élargie entre les nucléotides 227 et 337 (54–58). L'interaction spécifique entre le dimère d'ARNg et les protéines est médiée par les deux motifs protéiques en doigt de zinc de type « CCHC » (59) et semble également influencée par le domaine C-terminal p6 (60).

Après la formation des complexes ARN-Gag, ces derniers sont dirigés vers la membrane plasmique grâce aux groupements myristiques présents dans la région N terminale du domaine MA. L'acide myristique permet l'ancrage des protéines dans la membrane au niveau de radeaux lipidiques contenant des niveaux élevés de PIP2 (Phosphatidylinositol 4,5-bisphosphate), de cholestérol et de sphingolipides (28,61). Le domaine MA interagit également avec la queue cytoplasmique de la glycoprotéine d'enveloppe transmembranaire gp41 pour la recruter sur le lieu de bourgeonnement de la particule virale. A la suite de l'ancrage membranaire des complexes ARNg-Gag, les protéines Pr55<sup>Gag</sup> et Pr160<sup>Gag-Pol</sup> vont multimériser via leur domaine CA, permettant de nucléer l'assemblage de la particule immature. Le bourgeonnement des particules virales est régulé par le domaine p6 et la protéine Vpu. Le domaine p6 recrute le complexe *Endosomal Sorting Complex Required for Transport* (ESCRT) via une interaction avec *Tumor Susceptibility Gene 101* (TSG101), une sous-unité du complexe (62). Ce complexe permet la scission entre la membrane cellulaire et les particules, tandis que Vpu améliore le taux de libération des virions (63).

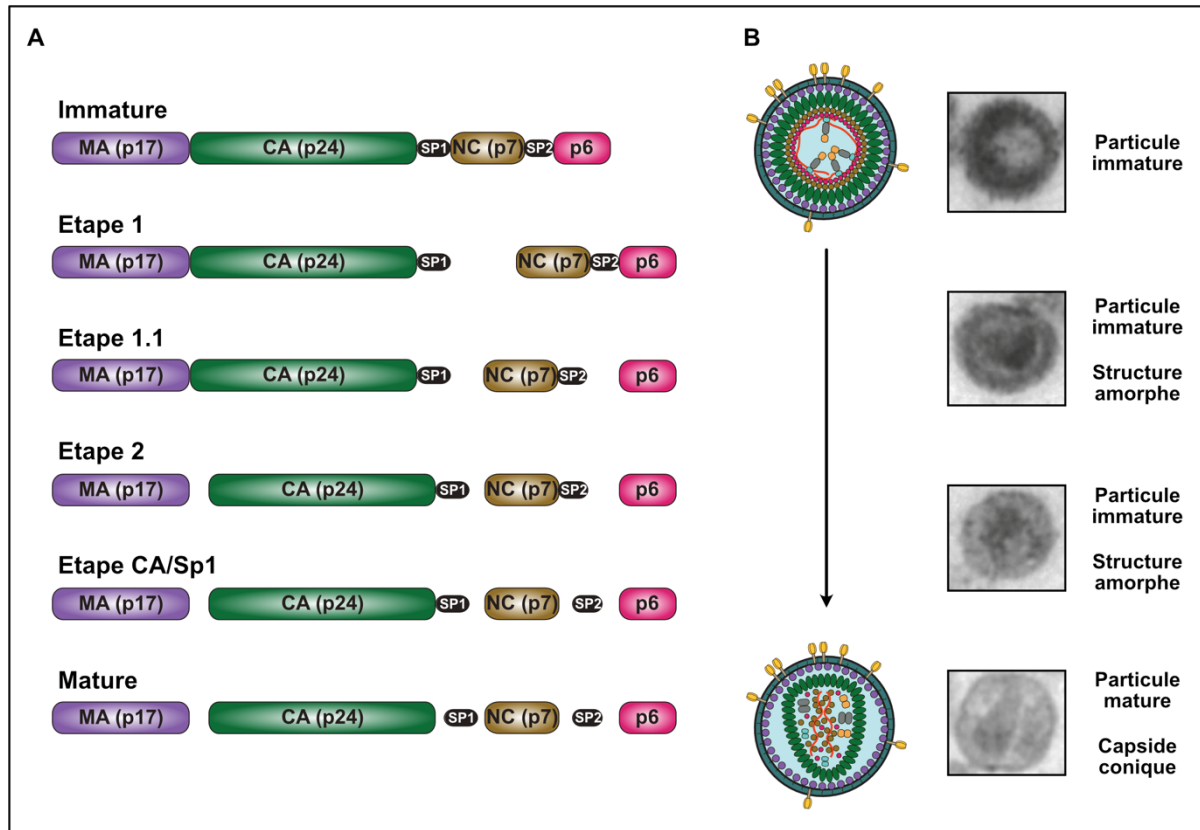
Les particules virales bourgeonnant des cellules infectées sont appelées immatures et sont non infectieuses. L'acquisition du potentiel infectieux est réalisée par une maturation des particules virales dans le milieu extracellulaire et médiée par l'activité protéolytique portée par un dimère de protéases virales (PR) (64,65). La maturation va induire un changement drastique de la particule virale par les clivages séquentiels des polyprotéines Pr55<sup>Gag</sup> et Pr160<sup>Gag-Pol</sup> (**Figures 2 et 7**). Ces clivages vont permettre la libération des protéines virales (**Figure 7A**) et permettre la formation de la capsid virale conique contenant le dimère d'ARNg mature (**Figure 7B**). En effet, une maturation génomique du dimère d'ARNg est concomitante à la maturation protéolytique. Cette maturation permet l'augmentation de la stabilité du



## INTRODUCTION

dimère d'ARN (66,67), et est requise pour mener à bien l'étape de transcription inverse (68,69).

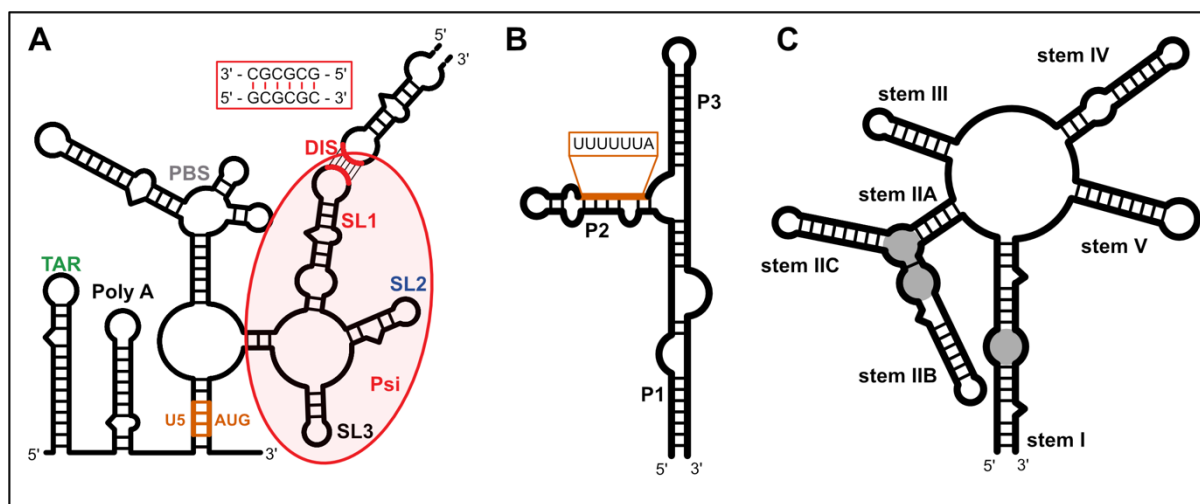
A l'issue du processus de maturation, les particules virales ont acquis leur potentiel infectieux et peuvent initier de nouveaux cycles répliatifs. Cependant, les mécanismes moléculaires impliqués dans la maturation du dimère d'ARNg sont peu explorés.



**Figure 7** : Représentation schématique de la cascade protéolytique de Pr55<sup>Gag</sup> et évolution morphologique de la particule virale. (A) Le dimère de protéases virales induit la maturation protéolytique de Pr55<sup>Gag</sup> (immature) en libérant l'ensemble des domaines protéiques composant le précurseur par 5 clivages successifs. (B) Lors de cette maturation protéolytique, la particule virale va également subir des changements structuraux qui ont été capturés en microscopie électronique à transmission (64).

## II/ ARNg du VIH-1 : relation structure-fonction

La majorité des génomes rétroviraux sont des dimères de molécules d'ARN dont la taille est comprise entre 8 et 10 kb (70). L'ARNg du VIH-1 est une molécule de 9,2 kb coiffée d'une guanine triméthylée en 5' et polyadénylée en 3' (71). Ces ARNg sont impliqués dans deux processus essentiels au cycle réplcatif. En effet, ils servent à la fois de matrice pour la synthèse des protéines structurales et enzymatiques, et de génome à encapsider dans les particules virales pour procéder aux cycles réplcatifs suivants. Cette dualité fonctionnelle suppose la présence de signaux permettant de discriminer les ARN destinés à la traduction de ceux encapsidés. Ainsi, la structure de l'ARN a été étudiée afin d'associer les différentes fonctions identifiées à des motifs structuraux. Lors de ces études, trois régions ont été mises en évidence pour leurs implications dans des processus viraux distincts. Ces trois régions sont l'extrémité 5' terminale, la jonction entre les gènes *gag* et *pol* et un motif d'environ 350 nts localisé dans la séquence codante du gène *env* (**Figure 8**). Ces trois régions ont été montrées comme étant requises pour les processus : de transcription, d'épissage, d'export nucléaire, de traduction, de dimérisation, de sélection, d'encapsidation et de transcription inverse. Chaque élément structural a été montré comme régulant au moins l'une de ces étapes.



**Figure 8** : Représentation schématique des régions de l'ARNg impliquées dans le cycle viral. (A) La région 5' terminale comprend les 350 premiers nucléotides de l'ARNg. Elle est impliquée dans les étapes de transcription (TAR et polyA), d'épissage (SL2), de traduction (U5:AUG), de dimérisation (SL1 - DIS), de sélection et d'encapsidation (Psi) et de rétrotranscription (PBS) (adapté de 72). (B) La région comprenant les nucléotides 1570 à 1700 englobe la structure responsable du *frameshift* -1. Cette structure permet la synthèse des protéines enzymatiques Pr160<sup>Gag-Pol</sup> (adapté de 65). (C) Les nucléotides 7700 à 8050 localisés dans le gène *env* se replient en une structure capable de former entre 4 et 5 tiges-boucles, appelée *Rev-Responsive Element* (RRE). Cette

## INTRODUCTION

structure est impliquée dans l'export nucléaire des ARN viraux non épissés et partiellement épissés via l'interaction avec la protéine Rev dont les sites de fixation sont grisés (adapté de 66). Les séquences encadrées (A) en rouge et (B) en orange correspondent aux séquences de dimérisation et glissante du *frameshift*, respectivement.

### II.1/ Transcription

La transcription de l'ARNg du VIH-1 est contrôlée par un promoteur localisé dans la région U3 du 5' LTR proviral. Cependant, la transcription basale est relative faible. En effet, l'élongation de la majorité des transcrits est terminée prématurément après que la Pol II s'arrête à proximité du site d'initiation de la transcription (43). Ces arrêts produisent des transcrits de 59 nucléotides (75) qui sont capables de former la tige-boucle TAR (**Figure 8A** – vert). Cet élément structural a été démontré comme essentiel pour la transactivation de la transcription (76,77). En effet, la protéine virale Tat reconnaît et fixe la boucle riche en U proche de la boucle apicale de TAR (78) et recrute le complexe p-TEFb sur la boucle apicale (79,80). L'étude de différents sous-types viraux ont montré que la présence de mutation dans la tige basale de TAR n'affectait pas significativement la structure secondaire ou la sensibilité de la TAR à la fixation de Tat (81,82). Cependant, TAR n'est pas l'unique motif impliqué dans la régulation de la transcription. En effet, après que l'élongation de la transcription reprenne, une seconde tige-boucle est synthétisée par la Pol II. Cette tige-boucle, appelée PolyA, présente un signal de polyadénylation dans sa boucle apicale (<sup>73</sup>AAUAAA<sup>78</sup>) qui a été montré comme étant nécessaire à l'encapsidation, mais qui inhibe la transcription par une polyadénylation précoce (83,84). Ce signal est réprimé par la fixation du petit ARN nucléaire U1 (ARN snU1) et en particulier de la protéine 70K du complexe U1 snRNP (85).

### II.2/ Épissage

La transcription du provirus permet la synthèse d'un unique ARNg de 9,2 kb permettant la traduction des protéines Pr55<sup>Gag</sup> et Pr160<sup>Gag-Pol</sup>. Cependant, l'ensemble des autres protéines virales sont issues d'événements d'épissages, dont 100% d'entre eux commencent au site majoritaire d'épissage D1 (**Figure 4A**). Ce site donneur est localisé dans la région 5' non traduite de l'ARN, au niveau de la boucle apicale de la tige-boucle SL2 (**Figure 8A** – bleu). La séquence de D1 (<sup>289</sup>GGUGAGU<sup>295</sup>) est complémentaire à celle de

## INTRODUCTION

l'ARNsnU1 (5'- ACUUACCUG -3'), ce qui permet l'interaction avec le petit ARN, le recrutement du spliceosome et l'épissage subséquent de l'ARN viral (86). Cependant, il a été montré que la structure en tige-boucle de SL2 régule négativement le mécanisme d'épissage (87,88) et qu'il est probable que la structure de l'ARN permet de contrebalancer le processus d'épissage (84). En effet, l'épissage des ARNg n'est plus requis lors des phases tardives du cycle viral et il est nécessaire de conserver l'intégrité du génome à encapsider.

### II.3/ Export nucléaire

Parmi l'ensemble des ARN viraux, la majorité présente un profil où l'épissage n'a pas été complété. Dans une cellule eucaryote, ces ARN sont considérés comme aberrants et dégradés par les systèmes de régulations post-transcriptionnel (89). En effet, la présence de séquences introniques et de codons de terminaison prématurés peuvent être reconnus comme des signes d'épissage défectueux par les protéines cellulaires. Pour contourner ces mécanismes, l'export des ARN viraux repose sur une interaction entre la protéine virale Rev et le RRE (**Figure 8C**). Le RRE est une structure présentant 4 à 5 tiges-boucles, selon les études (74), ainsi que différentes boucles internes riches en purines. Ces dernières semblent responsables de la reconnaissance et de la fixation spécifique de Rev sur le RRE (90). Pour assurer l'export, la multimérisation de Rev est nécessaire (91) et permet ensuite de former un complexe avec différentes protéines, comprenant CMR1 et Ran-GTP. Cependant, Rev interagit également avec les protéines *Up Frameshift protein 1* (UPF1) et Staufén, des protéines centrales aux mécanismes de dégradation de l'ARN du NMD (*Nonsense-mediated decay*) et du SMD (*Staufen1-mediated decay*) (92–94). Ainsi, le VIH-1 assure l'export nucléaire de ses ARN via une interaction médiée par la structure de ses ARN et une séquestration/exclusion des facteurs permettant la dégradation de ces derniers.

### II.4/ Traduction

Dans le cytoplasme, l'ensemble des ARN viraux peuvent être pris en charge par la machinerie traductionnelle cellulaire et permettre la synthèse des protéines virales (**Figure 8A** et **8B** - orange). Ces protéines sont majoritairement issues d'un mécanisme de traduction coiffe-dépendant. Cependant, le VIH-1 met également en place différentes voies de

## INTRODUCTION

traductions alternatives (95). En effet, différentes études ont mis en évidence un mécanisme non canonique pour la synthèse de la protéine Env à partir de l'ARNm bicistronique Vpu-Env (**Figure 4B**). Env serait donc produit via un mécanisme de *ribosome shunting* (96,97) bien que l'hypothèse initiale ait été celle d'une synthèse reposant sur le mécanisme de *leaky scanning* (98). De même, la synthèse des protéines Pr55<sup>Gag</sup> et Pr160<sup>Gag-Pol</sup> a été montrée comme pouvant être initiée par un mécanisme coiffe-dépendant suivi d'un *frameshift* -1 (99–101), ou à partir de deux *Internal Ribosome Entry Site* (IRES) (102–104). Parmi ces différents mécanismes traductionnels, ceux de *frameshift*, d'IRES et *ribosome shunting* reposent sur des éléments structuraux de l'ARN (105).

Dans le cas de la traduction des protéines Pr55<sup>Gag</sup> et Pr160<sup>Gag-Pol</sup>, la structure adoptée par la région 5' terminale de l'ARNg est supposée être également impliquée (106). En effet, un interrupteur structural basé sur les interactions U5:DIS et U5:AUG régulerait (**Figure 8A** – orange) le devenir des ARN viraux, les dirigeant vers la traduction ou l'encapsidation, respectivement. Cet aspect structural de l'ARNg du VIH-1 sera traité plus en détail dans la partie IV.2.1. de l'introduction.

### II.5/ Dimérisation, sélection et encapsidation

En général, la moitié des ARNg du VIH-1 sont activement traduits dans le cytoplasme sans séparation distincte des ARNg non traduits (107). La dimérisation ARNg non traduits se produit vraisemblablement dans le cytosol des cellules infectées (108) bien que certaines études placent ce processus à la membrane plasmique (109). Cette dimérisation repose sur l'interaction non covalente de deux séquences palindromiques de 6 nucléotides exposés par la boucle apicale de SL1 de chaque molécule d'ARNg (110,111) (**Figure 8A** – rouge).

La sélection spécifique des ARNg du VIH-1 dans le cytoplasme parmi l'ensemble des ARN cellulaires et viraux suivi de leur encapsidation à la membrane plasmique est régie par l'interaction entre le domaine NC des protéines Pr55<sup>Gag</sup> et la région Psi de l'ARNg (70,112–114). Historiquement, la tige-boucle SL3 a été considérée comme le site de haute affinité de fixation de Pr55<sup>Gag</sup> pour la sélection et l'encapsidation (115–117). Cependant, de nouvelles études localisent le site principal de fixation au niveau de la boucle interne de SL1 (49,118). De façon générale, plusieurs protéines Pr55<sup>Gag</sup> sont capables de reconnaître l'ARNg pour le diriger vers le lieu d'encapsidation (50,119,120). De la même façon que la partie précédente,

## INTRODUCTION

les étapes de dimérisation, sélection et encapsidation sont traitées plus en détail dans la partie IV.2.1. de l'introduction.

### II.6/ Transcription inverse

Après l'encapsidation, le dimère d'ARNg n'est pas le seul acide nucléique présent dans la particule virale. En effet, les particules néoformées encapsident également des ARN cellulaires, tels que des ARN 7SL, des ARN snU1 ou différents ARN de transfert (121). Parmi les ARNt retrouvés dans les particules virales, les ARNt<sup>Lys,1, 2 et 3</sup> sont significativement enrichis. La présence de l'ARNt<sup>Lys,3</sup> reflète une importance particulière en tant qu'amorce requise à l'étape de transcription inverse. Cet ARNt interagit avec le domaine PBS présent dans la région 5' terminale de l'ARNg (**Figure 8A** – gris) et initie la transcription inverse (**Figure 6B**). Cette interaction longue de 18 nucléotides est rendue possible par la dynamique intrinsèque du domaine PBS où plusieurs nucléotides restent accessibles (122).

Dans le cas de la transcription inverse avec le domaine PBS, ou les autres étapes du cycle présentées précédemment, les structures adoptées par les motifs présents dans l'ARNg du VIH-1 sont étroitement liées aux fonctions et aux interactions mises en place. Ainsi, il est critique de résoudre des modèles précis des différentes régions car les meilleures représentations structurales permettront les modèles mécanistiques les plus complets.

### III/ Méthodes d'études des structures d'ARN

De la simple traduction à la régulation de l'expression génique en passant par la catalyse de processus biologiques essentiels, les deux dernières décennies ont permis de mettre en évidence la diversité des fonctions portées par les molécules d'ARN (123,124). Ces fonctions des ARN sont intrinsèquement connectées aux structures complexes qu'ils sont capables de former. Ainsi, la pleine compréhension des fonctions biologique d'un ARN passe par l'étude des structures qu'il adopte pour remplir ces fonctions d'intérêt.

Différentes méthodes existent pour déterminer les structures 2D ou 3D d'un ARN (125–127). Ces approches peuvent être divisées en deux catégories. La première catégorie comprend les approches biophysiques qui interrogent directement la structure atomique. La seconde catégorie contient l'ensemble des approches biochimiques, indirectes, qui interrogent certaines caractéristiques des nucléotides impliqués dans une structure (flexibilité, exposition au solvant, interaction WC-WC, capacité à être modifié).

#### III.1/ Approches biophysiques

L'étude de la structure des ARN par des approches biophysiques est généralement réalisée par 4 méthodes qui sont la cristallographie aux rayons X (XRC), la diffraction des rayons X aux petits angles (SAXS), la résonance magnétique nucléaire (RMN) et la cryo-microscopie électronique (cryo-EM). Chacune de ces méthodes s'intéresse directement à la résolution atomique des molécules étudiées. Cependant, leur mise en place se révèle bien plus compliquée que celles des approches biochimiques (128).

En effet, les approches de cristallographie et de cryo-EM permettent de caractériser les structures des molécules d'ARN à haute résolution par l'obtention de cartes de densité (129). Cependant, dans le cas de la cristallographie, l'obtention de cristaux n'est pas aisée pour des molécules flexibles et dynamiques telles que les ARN qui peuvent adopter de nombreuses conformations distinctes (130). Cette limitation est bien représentée par le faible nombre et la faible diversité des structures d'ARN issues de cette approche. En effet, la majorité de ces structures obtenues concernent des ARNt et des ribosomes, molécules relativement stables et souvent complexées à des protéines. En revanche, la limitation provenant de la nécessité de cristallisation est levée dans le cas d'une approche basée sur la cryo-EM. En effet, la cryo

## INTRODUCTION

microscopie électronique requiert de figer des molécules pures à des températures cryogéniques dans un environnement d'eau vitrifiée (131). Jusqu'à présent, une limite de taille restreignait l'application de cette approche aux macromolécules et la majorité des structures de qualité obtenues étaient celles de protéines et de complexes protéiques. Cette limite de taille a récemment été franchie avec la détermination du riboswitch SAM-IV, d'une taille inférieure à 40 kDa, soit moins de 120 nucléotides (132). Cette limite avait également été dépassé l'année précédente, mais avait nécessité de coupler la cryo-EM à de la RMN (133). Cependant, dans le cas de larges molécules d'ARN, qui permettraient théoriquement les meilleures résolutions, le problème des nombreuses conformations adoptées reste une limitation majeure (134).

Parmi la totalité des structures d'ARN seul rapportées dans la PDB en octobre 2022, environ 33% ont été obtenues par RMN. De la même façon que pour la cryo-EM, la cristallisation de l'échantillon étudié n'est pas nécessaire en RMN. Ainsi, les informations structurales sont issues de l'études des fréquences de résonance propres à chaque noyau atomique (135). Ces études permettent de déterminer la proximité des atomes dans l'espace et donc de modéliser des structures d'ARN. Un des avantages de cette approche est sa capacité à obtenir des informations sur les interactions non canoniques impliquées dans les repliements de structures d'ARN, tels que les paires de bases non Watson-Crick, les empilements de bases, les interactions triples ou les pseudo-nœuds (136). Cependant, tout comme les autres approches, la RMN est rapidement limitée par la taille de l'ARN qui est généralement inférieure à 100 nucléotides (137).

Pour finir, la diffraction des rayons X aux petits angles a montré quelques résultats prometteurs dans la résolution des structures d'ARN (138,139). Cette méthode permet l'acquisition rapide d'informations globales sur les structures d'ARN par l'étude des profils de diffraction des rayons X traversant la molécule d'intérêt. En fonction des conditions expérimentales, des informations précises jusqu'à 1 nm peuvent être obtenues, bien que la résolution soit souvent inférieure à celle des autres méthodes si l'approche n'a pas été couplée à une autre technique.







## INTRODUCTION

### III.2/ Approches biochimiques

A l'inverse des approches biophysiques, les méthodes biochimiques réalisent des mesures quantitatives basées sur un sondage de la structure adoptée par l'ARN d'intérêt. L'évolution de ces méthodes et leurs applications à l'étude de génomes viraux sont présentées dans la revue « Chemical and Enzymatic Probing of Viral RNAs: From Infancy to Maturity and Beyond » (140).

Review

# Chemical and Enzymatic Probing of Viral RNAs: From Infancy to Maturity and Beyond

Orian Gilmer <sup>†</sup>, Erwan Quignon <sup>†</sup> , Anne-Caroline Jousset <sup>†</sup> , Jean-Christophe Paillart <sup>†</sup> , Roland Marquet <sup>\*</sup>   
and Valérie Vivet-Boudou <sup>\*</sup>

Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR9002, F-67000 Strasbourg, France; o.gilmer@ibmc-cnrs.unistra.fr (O.G.); e.quignon@ibmc-cnrs.unistra.fr (E.Q.); ac.jousset@ibmc-cnrs.unistra.fr (A.-C.J.); jc.paillart@ibmc-cnrs.unistra.fr (J.-C.P.)

\* Correspondence: r.marquet@ibmc-cnrs.unistra.fr (R.M.); v.vivet@ibmc-cnrs.unistra.fr (V.V.-B.);  
Tel.: +33-(0)3-88-41-70-54 (R.M.); +33-(0)3-88-41-70-35 (V.V.-B.)

<sup>†</sup> These authors contributed equally to this review.

**Abstract:** RNA molecules are key players in a variety of biological events, and this is particularly true for viral RNAs. To better understand the replication of those pathogens and try to block them, special attention has been paid to the structure of their RNAs. Methods to probe RNA structures have been developed since the 1960s; even if they have evolved over the years, they are still in use today and provide useful information on the folding of RNA molecules, including viral RNAs. The aim of this review is to offer a historical perspective on the structural probing methods used to decipher RNA structures before the development of the selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) methodology and to show how they have influenced the current probing techniques. Actually, these technological breakthroughs, which involved advanced detection methods, were made possible thanks to the development of next-generation sequencing (NGS) but also to the previous works accumulated in the field of structural RNA biology. Finally, we will also discuss how high-throughput SHAPE (hSHAPE) paved the way for the development of sophisticated RNA structural techniques.

**Keywords:** RNA; structure; enzymatic probe; chemical probe; SHAPE; capillary electrophoresis; high-throughput sequencing; mutational profiling



**Citation:** Gilmer, O.; Quignon, E.; Jousset, A.-C.; Paillart, J.-C.; Marquet, R.; Vivet-Boudou, V. Chemical and Enzymatic Probing of Viral RNAs: From Infancy to Maturity and Beyond. *Viruses* **2021**, *13*, 1894. <https://doi.org/10.3390/v13101894>

Academic Editors: Anne Simon and Stuart F.J. Le Grice

Received: 27 August 2021  
Accepted: 16 September 2021  
Published: 22 September 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

RNA molecules are key players in a variety of biological events, and this is particularly true for viral RNAs. It has been known since 1949 [1] that the infectivity of some viruses is driven by viral RNA (vRNA), and that vRNA alone can infect plants [2]. Therefore, to better understand the replication of those pathogens and try to block this process, special attention has been paid to the structure of their RNAs. Indeed, many functions that are essential to viruses are encoded by specific RNA domains that are engaged in long-range interactions or promote specific interactions with metal ions, proteins, or other RNAs. Additionally, vRNAs can be replicated, spliced, translated, or packaged as genomic RNAs (gRNAs) into new virions. Because these processes are highly dependent on interactions enabled or prevented by RNA structure, full understanding of the molecular mechanisms involved requires characterization of these RNA structures. However, RNAs are dynamic molecules that can fold into various conformations, depending on environmental conditions. Their flexibility and large size often make them unsuitable for study by the most powerful structural approaches, such as X-ray crystallography [3,4], nuclear magnetic resonance, [5] or cryo-electron microscopy [6], although this situation is changing in recent years thanks to the coupling of cryo-EM with high-throughput biochemistry and computer modeling of the 3D structure [7]. Nevertheless, by the second half of the 20th century, methods to probe the structure of RNA in solution were developed; even if these techniques have evolved

over the years, they are still in use today and provide useful information on the folding of RNA molecules, including vRNAs.

The aim of this review is to offer a historical perspective on the structural probing methods used to decipher RNA structures before the development of the high-throughput selective 2'-hydroxyl acylation analyzed via the primer extension (hSHAPE) methodology, discussing examples taken from the world of viruses and how they have influenced the probing techniques since 2005. Indeed, nowadays, we have numerous sophisticated tools to assess RNA structure in vitro, in cellula, and even in viro, either alone or in complex with various macromolecules. These technological breakthroughs using enzymatic and chemical probes, which are detected through advanced methods [8], were made possible thanks to next-generation sequencing (NGS) approaches but also to the previous works accumulated in the field of structural RNA biology [9]. Therefore, to understand where most of the currently used methods originated from, it is necessary to go back to the "Pre-SHAPE era" of RNA structural biology. Finally, we will also discuss how hSHAPE paved the way for the development of sophisticated RNA structural techniques that provide more detailed structural information. However, the crosslinking-based structural methods, which have also been developed recently and are complementary to the RNA probing methods, are not reviewed here (for recent reviews of these techniques, see [10–12]).

## 2. History of the RNA Structure Epic: From Scratch to SHAPE

Since the first mention of RNA, referred to as "yeast nucleic acid" in the early 1900s, numerous key discoveries have been made concerning its structure. In 1954, the hypothesis that RNA is a "2'-3'-5' branched" molecule was abandoned in favor of the linear 5'-3' chain [13]. In 1956, Rich and Davies [14] showed that RNA is able to form double-stranded structures, in a "DNA-like" fashion. The next year, Felsenfeld et al. [15] uncovered the complexity of the RNA structure, with a triple-stranded RNA structure observed by combining X-ray diffraction and sedimentation studies. From that point onwards, special attention was brought to RNA structure and the methods to determine it.

### 2.1. Birth of RNA Structure Probing

Long before the first observation of a structured RNA molecule (by Rich and Davis in 1956), ribonucleases were known to degrade ribonucleic acids [16]. However, little was known about their specificity towards nucleotide identity or RNA structures. An early report demonstrating the nucleotide specificity of ribonucleases was published in 1957, albeit with some errors, when RNase T1 and RNase T2 were purified from *Aspergillus oryzae* by Sato and Egami [17]. Then, in 1962, it was hypothesized that double-stranded RNA is resistant to nucleases [18–20]. Finally, in 1965, Holley et al. [21] took advantage of the properties of two RNases, RNase T1 (specific cleavage of RNA at the 3' phosphate of unpaired Gs) [22] and pancreatic RNase (specific cleavage at the 3' of unpaired Us and Cs [23]), to study the structure of yeast tRNA<sup>Ala</sup>. Comparing the short and long fragments obtained under complete and partial digestion, respectively, they were able to identify and arrange the different fragments, in relation to each other. They proposed three structural models for tRNA<sup>Ala</sup>, including the now classical cloverleaf structure.

In parallel to the structural studies with enzymes, chemical reagents that are able to interact or modify nucleic acids were also investigated, in order to understand the properties of RNA. Historically, one of the first chemicals known for its ability to interact with nucleic acids was 1-chloro-2-[(2-chloroethyl)sulfonyl]ethane, also known as mustard gas [24]. This compound, as well as other alkylating agents, was shown to efficiently inactivate viruses, such as the tobacco mosaic virus [25]. Some years later, methylation of adenosine, cytosine, and guanosine, by the nowadays widely used dimethylsulfate (DMS), was demonstrated by Lawley and Brookes [26]. In the same period, Gilham described, for the first-time, a carbodiimide chemical reagent able to modify uridine and guanosine, thus rendering these nucleotides less sensitive to RNase cleavage [27]. In 1965, the link with the secondary structure of nucleic acids was made indirectly by using the

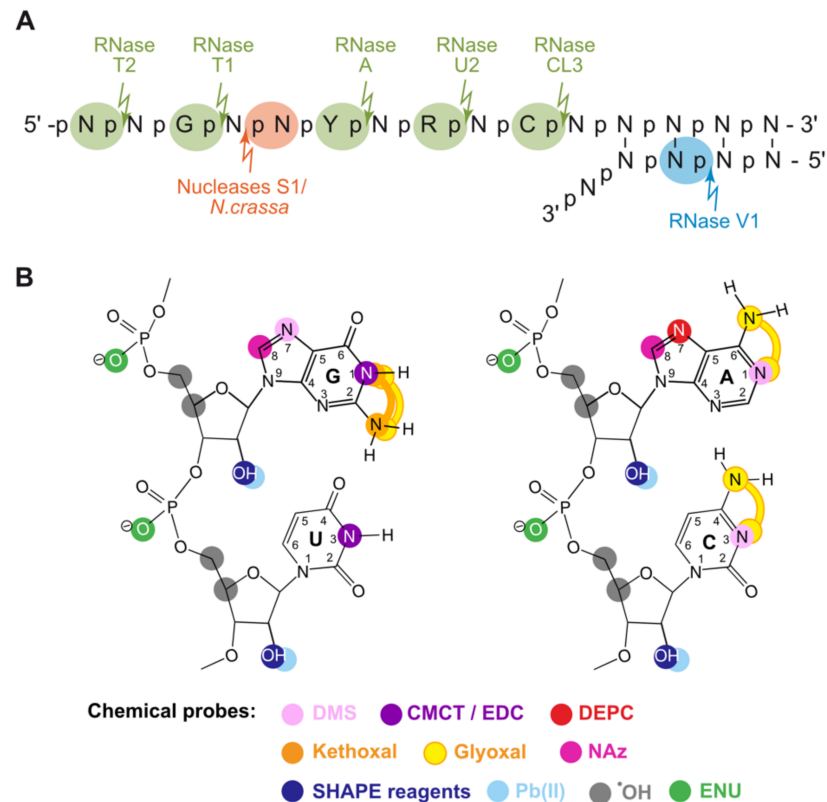
property of carbodiimide derivatives to protect the U and G nucleotides from specific RNase digestion [28,29] and not yet as a direct chemical probe by itself, as it is now used.

## 2.2. Childhood of RNA Probing and Investigation of Viral RNA Structures

### 2.2.1. Characteristics and Specificities of the Pre-SHAPE Probes

From 1965 to 2005, 40 years of RNA probing research has led to the development of a wide range of tools available to scientists (Figure 1 and Table 1).

The most commonly used RNases specifically cleave single-stranded nucleotides. Some of them cleave RNA molecules without sequence specificity, while others have a more or less pronounced preference for one or several nucleotides. For example, *Neurospora crassa* endonuclease [30], RNase T2 [17], and nuclease S1 from *Aspergillus oryzae* [31] cleave single-stranded RNA (ssRNA), with limited dependency on the nucleobase identity. RNase T2, together with RNase T1 (see below), presents the additional advantage of being active at neutral and basic pH, even in the absence of divalent cations and is, thus, ideal for studying the effects of pH and  $Mg^{2+}$  ions on the RNA secondary structure [32]. RNase T1 from *Aspergillus oryzae* specifically cleaves RNA at the 3' phosphate of unpaired G's [22], and chicken liver RNase 3 (CL3) cleaves ssRNA at Cs and, to a lesser extent, at As and Us [33]. RNase U2, from *Ustilago sphaerogena*, cleaves the 3'-5' phosphodiester bonds of unpaired As and Gs [34], and RNase A cuts after unpaired pyrimidine residues, with a preference for Cs and especially CpA motifs [35]. In contrast to these ssRNA-specific enzymes, RNase V1, also known as cobra venom RNase (since it was isolated from the venom of the Naja snake *Naja oxiana* [36,37]), selectively cleaves double-stranded RNA (dsRNA), or staked nucleotides, by generating 5'-phosphate ends [38]. It has to be noted that nowadays, RNase V1 is still one of the very few probes that gives a positive signal for the ds regions, together with hydroxyl radicals generated from the intercalating agent MPE-Fe(II). The activity and optimal conditions of use for many nucleases are detailed in Desai et al. [39]. The main limitations of RNases as structural probes come from their size, which makes them sensitive to steric hindrance and prevents their use in cells or viral particles. Furthermore, if the reaction conditions are too strong, secondary cleavages, that do not reflect the native RNA secondary structure, may occur.



**Figure 1.** Enzymatic and chemical probes available nowadays. (A) Common enzymatic probes and their targets on ss- and dsRNA. The arrows and the highlighted nucleotides indicate whether the fragment formed after strand scission is 3' or 5' phosphate. (B) Main chemical probes and their target positions on base, sugar, and phosphate. DMS: dimethylsulfate; CMCT: 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluenesulfonate; EDC: 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide; DEPC: diethylpyrocarbonate; kethoxal: 3-ethoxy-1,1-dihydroxy-2-butanone; SHAPE reagents are NMIA: N-methylisatoic anhydride, BzCN: benzoylcyanide, and NAI: 2-methyl nicotinic acid imidazolid; FAI: 2-methyl-3-furoic acid imidazolid; 1M6: 1-methyl-6-nitroisatoic anhydride; 1M7: 1-methyl-7-nitroisatoic anhydride and 2A3: 2-aminopyridine-3-carboxylic acid imidazolid; NAz: nicotinoyl-azide; Glyoxal: ethanedial [40]; ENU: ethyl-nitrosourea; Pb(II): lead ion; \*OH: hydroxyl radical.

Beside enzymatic probes, chemical probes were also developed and have several advantages. Indeed, they are smaller in size, potentially effective in vivo, compatible with most buffer components, and active over a wide range of pH, salt, and temperature. Most of them target ssRNAs and show some base specificity. Kethoxal, which was originally known as an antiviral agent (thanks to its ability to react with Gs [41]), was used to probe N1- and NH<sub>2</sub>-2-G of the ssRNA regions [42]. In 1970, Öberg et al. showed that diethylpyrocarbonate (DEPC) modifies ssRNA, but not dsRNA, by conducting studies on polioviruses [43]. It was not until a year later that DEPC was found to react with N7-A [44]; in 1980, Peattie and Gilbert [45] described a procedure to probe stacked adenosines by treating RNA with DEPC and then inducing a cleavage with aniline. Additionally, 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metha-p-toluene sulfonate (CMCT) targets N3-Us and N1-Gs [46], whereas DMS modifies N3 and N1 of unpaired Cs and As, respectively, and the N7 of Gs was not involved in tertiary interactions [45]. Interestingly, methylation of the N1-A position was only discovered in 1985, since it can be detected solely with the primer extension readout method (see below) [47]. The modification of the Hoogsteen face of Gs is interesting, since it allows for the use of DMS to probe G-quadruplexes [48]. Other chemicals react with the sugar-phosphate backbone and, therefore, do not depend

on the nature of the base. Such compounds include ethyl-nitrosourea (EtNu or ENU), which alkylates the oxygens of the phosphate groups of ss- and ds nucleic acids that are not engaged in tertiary interactions or in cation coordination. After an additional alkaline treatment, the phosphotriester is hydrolyzed, resulting in RNA strand scission [49,50]. Lead (II) cleaves the phosphate backbone of ssRNAs, with pronounced cleavage in metal ion binding pockets; the highly reactive hydroxyl radicals probe the solvent accessibility of the nucleic acid backbone when generated from Fe(II) coordinated to soluble EDTA [51] or double-stranded structures when generated from Fe(II) in complex with EDTA linked to the intercalator methidiumpropyl (MPE-Fe) [52].

In the first experiments in the 1960s, detection of the modified nucleotides was achieved by paper or gel electrophoresis of the small fragments obtained after the successive digestion of RNAs and the laborious work of puzzle reconstruction [21,53]. From the 1970s, the detection of modification or cleavage sites was achieved by one of two methods, both of which involve  $^{32}\text{P}$  (or sometimes  $^{35}\text{S}$ ) labeling. For probing experiments of small RNAs (or of the extremities of long RNAs), direct detection of strand scissions of 3' or 5' end-labeled RNAs is the method of choice. In that case, the cleavage of the ribo-phosphate backbone can, of course, be achieved by enzymatic probes or chemical probes, such as lead (II) or hydroxyl radicals. However, other chemical reagents that, under normal conditions, only modify RNA can be combined with a subsequent treatment that induces strand scission at the modification site. For instance, DMS modified guanosines, when reduced by  $\text{NaBH}_4$ , then treated with aniline, results in strand cleavage [54]; whereas N3 alkylated cytosines need to be treated with hydrazine, and then with aniline, for strand scission [55]. It has to be noted that treatment with hydrazine is not specific, since it may induce cleavage at the U positions, although these nucleotides are not probed by DMS. DEPC is also compatible with this detection method when coupled with aniline treatment [56], as well as ENU, since the phosphotriester formed after ENU treatment are easily hydrolyzed under mild alkaline conditions [57]. The RNA fragments are separated by polyacrylamide gel electrophoresis and by comparison with a sequencing ladder; thus, the positions of the cleavages in the RNA molecule are identified.

**Table 1.** Enzymatic and chemical probes from 1965 to 2005. Abbreviations are as follows: A: adenosine; G: guanosine; C: cytidine; U: uracil; N: any nucleotide; ssN: any unpaired nucleotide, regardless of the base; ssA (G, C, or U): unpaired adenosine (guanosine, cytidine, or uracil); dsN: any paired nucleotide; HDV: hepatitis D virus; TYMV: turnip yellow mosaic virus; HIV-1/2: human immunodeficiency virus type 1/2; EBV: Epstein–Barr virus; •OH: hydroxyl radical. \* indicate the publications that concern nucleic acids.

Probe	Target	In Cell-In Viro	Original Publication (* Related to Nucleic Acid)	Early Use as Structural Probe	Application to Viral RNA
Enzymatic					
RNase A	ssC and U	No	Markham and Smith (1952) * [58]	tRNA <sup>Ala</sup> (1965) [21]	HDV (1991) [59]
RNase T1	ssG	No	Sato and Egami (1957) * [22]	tRNA <sup>Ala</sup> (1965) [21]	TYMV (1982) [60,61]
Nuclease S1	ssN	No	Harada and Dahlberg (1975) * [31]	5S rRNA (1977) [62]	Bromoviruses (1981) [63]
RNase CL3	ssC>>A>U	No	Levy and Karpetsky (1980) * [33]	TYMV RNA (1982) [60]	TYMV (1982) [60]
RNase V1	4–6 nts in helices or stacked nts	No	Favorova et al. (1981) * [38]	5S rRNA (1982) [64]	TYMV (1982) [60,61]
RNase T2	ssA>ssN	No	Sato and Egami (1957) * [17]	yeast tRNA (1984) [32]	Poliovirus (1989) [65]
N. crassa endonuclease	ssN	No	Linn and Lehman (1965) [30]	tRNA <sup>Trp</sup> (1984) [66]	HIV-1 (1993) [67]
RNase U2	ssA>G>>C>U	No	Uchida et al. (1970) * [34]	16S rRNA (1987) [68]	HDV (1991) [59]
Chemical					
Kethoxal	N1 and N2- ssG	No	Stahelin (1959) * [41]	Yeast tRNA (1969) [42]	EBV (1988) [69]
DMS	N7-G/N1-ssA N3-ssC	Yes	Lawley and Brookes (1963) * [26]	Yeast tRNA <sup>Phe</sup> (1980) [45]	TYMV (1982) [61]

Table 1. Cont.

Probe	Target	In Cell-In Viro	Original Publication (* Related to Nucleic Acid)	Early Use as Structural Probe	Application to Viral RNA
DEPC	N7-A	No	Oberg (1970) * [43]	Yeast tRNA <sup>Phe</sup> (1980) [45]	TYMV (1982) [61]
ENU	Phosphates of ssN and dsN	No	Singer (1976) * [49]	tRNA <sup>Phe</sup> (1980) [50]	TYMV (1982) [60]
CMCT	N3-U and N1-G ss	No	Augusti and Brown (1965) * [28]	16S rRNA (1986) [46]	EBV (1988) [69]
MPE-Fe(II) (*OH source)	Phosphate/sugar backbone dsN	No	Hertzberg and Dervan (1982) * [70]	tRNA <sup>Phe</sup> (1984) [52]	HIV-2 (2013) [71]
Fe-EDTA (*OH source)	Solvent accessibility	Yes	Tullius and Dombroski (1986) * [72]	pre-mRNA (1989) [51]	HIV-1 (1999) [73]
Lead (II)	Phosphate/sugar backbone ssN	Yes	Werner et al. (1976) * [74]	16S rRNA (1989) [75]	HIV-1 (1997) [76]

An alternative method emerged when Qu et al. [77] and Moazed et al. [78] found that it was possible to map the location of enzyme cleavage sites or chemical modification at Watson–Crick positions (kethoxal, DMS, and CMCT) with reverse transcription stops. This technique involves the extension of a radiolabeled primer with reverse transcriptase (RT) and the subsequent resolution of the cDNA population, by denaturing polyacrylamide gel electrophoresis and detection, either by autoradiography or phosphorimaging.

### 2.2.2. Early Applications to Viral RNAs

The biochemical probing techniques mentioned above were often developed to solve eukaryotic and prokaryotic RNA structures, but they were also used alone or in combination and proved their efficiency in the structure determination of viral RNAs transcribed in vitro (extracted from virions or directly inside the particles) (Table 1). In the following sections, we describe some examples of the early applications of RNA probing for the elucidation of vRNA structures.

#### 3'-Terminal Region of Plant Viruses: tRNA Mimicry for Replication

Since the first RNA structures ever identified were tRNAs, it was not surprising that the first recognized viral structures were tRNA-like structures present in the 3' termini of ssRNA plant viruses (3'TLS). The specific features of some plant viral genomes were discovered after the observation of a linkage between the terminal adenosine of the turnip yellow mosaic virus (TYMV) RNA and a valine amino acid [79]. Structural and functional studies were performed to try to understand why vRNA was aminoacylated. In the early 1980s, Ahlquist et al. [63] studied the RNA 3'-terminal region of different bromoviruses and a cucumovirus. Using S1 nuclease to probe the RNA secondary structure, they showed that the 3'-terminal region was highly conserved, suggesting a strong functional constraint, even though the TLS secondary structures were quite different from that of canonical tRNAs. Nevertheless, the cellular aminoacyl-tRNA synthetases were able to recognize the 3'TLS, in order to aminoacylate the terminal CCA<sub>3'OH</sub> at levels similar to those of the cellular tRNA molecules.

Following these findings, many studies were performed on the 3'TLS from TYMV. Florentz et al. used ENU to probe the RNA structure, and they complemented that tool with a combination of enzymes, such as nuclease S1, RNase CL3, T1, and V1 [60]. In the meantime, Rietveld et al. [61] proposed a secondary structure for the TYMV 3' terminal region. They used RNase T1, V1, and nuclease S1, as well as the chemical probes DMS and DEPC. At that time, DMS had already been used to decipher the global ss or ds behavior of an RNA structure [80] and to study the structure of yeast tRNA [46]. Rietveld et al. demonstrated the existence of a pseudoknot, which was absolutely necessary for the aminoacylation of the 3'TLS. Furthermore, they showed that the 3'-end of TYMV gRNA does not present the conventional acceptor stem of tRNA, but they proposed a tertiary structure, which nevertheless resembled the acceptor arm of tRNA. This arrangement is conserved in other plant viruses of the tymovirus and tobamovirus groups. In 1983, the probing of cucumovirus and bromovirus gRNA 3'-end with RNase T1 was performed by

Joshi et al. [81]. These authors studied the minimal sequence necessary to form TLS, as well as the implication of  $Mg^{2+}$  ions into aminoacylation. Up to now, the exact function of the 3'TLS remains unclear [82]. Yet, numerous insights on 3'TLS were acquired. Three different classes of 3'TLS have been identified, characterized by the amino acid bound to their 3'CCA<sub>OH</sub>, with three different types of secondary structures giving rise to very similar 3D structures [83].

#### 3'-Terminal Region of Plant Viruses: Cap Independent Translation Enhancers

In ssRNA plant viruses, TLS are not the only structure found in the 3'terminal region. Some viruses devoid of 5' cap and 3' polyA tail use a non-canonical mechanism to recruit ribosomes to their 5'-end, by a specific RNA element [84]. These motifs, named 3' cap independent translation enhancers (3'CITE) [85], were discovered in 1993 in a satellite tobacco necrosis virus [86]. The 3'CITEs are able to interact with the initiation factor complex eIF4E, a component of the translation machinery, through their 3'-end structure (for a review of the structure of 3'CITE, see [87]). To assess the structure of the tomato bushy stunt virus (TBSV) 3'CITE, a combination of enzymatic (RNase S1 and V1) and chemical probing (DEPC and CMCT) was used and allowed to model the 5'-3' interaction responsible for the delivery of the initiation factor to the 5'-end of the genome, through long-distance base-pairing [88]. The initiation factor recruits the 40S subunit of the ribosome, facilitating the translation of the viral RNA. Finally, even if the molecular mechanisms involved in enhanced translation differ from one 3'CITE to another; they remain interchangeable and fulfill a similar function [87].

#### RNA Structures of Human Pathogenic Viruses

The experiments performed with plant vRNAs paved the way for the chemical and enzymatic probing of other viruses. In this regard, a particular interest was expressed for human pathogenic viruses. One of them was poliovirus, a well-known member of the *Picornaviridae*. Indeed, structural elements involved in the recruitment of host ribosomes to promote translation of the viral proteins were discovered in the 5'-end region of picornaviruses [89,90]. One of the first internal ribosome entry site (IRES) secondary structure models to be built was the one from the encephalomyocarditis virus (EMCV), which was obtained through a combination of chemical (DMS) and enzymatic (RNase V1 and S1) probing [91]. Later, IRESes were found to be present in many viral families, with more or less complex structures able to promote translation initiation, usually in a cap-independent manner, without requiring the complete host translation machinery (for review, see [92]). Today, the structures and molecular mechanisms involved in the translation induced by IRES are extensively studied.

Moreover, picornaviruses structures were also studied to gain a better understanding of their virulence. In 1989, probing with kethoxal (along with DMS and RNases T1, T2, and V1) was used to identify the structural differences between a virulent and an attenuated poliovirus [65]. These data confirmed and completed a previous computer-predicted minimal energy structure [93] and provided additional evidence for the existence of an interaction in a region important for neurovirulence.

RNA viruses are not the only human pathogen relying on RNA structure for their infectivity. Indeed, the Epstein-Barr virus (EBV) (a *Herpesviridae*) codes for two small ncRNAs, EBER-1 and EBER-2, involved in latency and tumorigenic phenotypes. The structure of these two RNAs was studied using a combination of DMS, kethoxal, and CMCT to obtain information on all four nucleotides [69]. Associated with RNase V1 and nuclease S1, these reagents allowed the determination of the structure of EBER1 and EBER2 ncRNAs, either naked or in cell from extracted ribonucleoprotein (RNP) complexes. The modifications and cleavage sites were mapped by primer extension, except for the extreme 20 nucleotides at the 3'-end of the RNAs, for which enzymatic digestion and PAGE of the digests was necessary. Moreover, the authors determined the binding sites of the La



protein, which constitutes the RNP complexes by immunoprecipitation with La-antigen followed by RNase T1 or RNase A digestion.

The appearance of the acquired immunodeficiency syndrome (AIDS) pandemic in the 1980s has led to a large number of probing studies of the human immunodeficiency virus type 1 (HIV-1) genomic RNA (gRNA) at different stages of the viral cycle and with a wide range of probes. For example, the structural domains, located in the 5' terminal region, and their potential role in dimerization were studied by chemical probing with enzymatic and chemical probes [94,95]. *Neurospora crassa* endonuclease was used to study the tRNA<sup>Lys,3</sup>-gRNA reverse transcription initiation complex, together with a panel of chemical probes [67,96]. In addition to the predictable interaction between the viral primer binding site (PBS) and the 3' extremity of the tRNA<sup>Lys,3</sup>, the authors were able to identify additional and specific contacts between an A-rich loop (located upstream of the PBS and the anticodon loop of tRNA<sup>Lys,3</sup>) by labeling the 3' extremity of the tRNA and performing a denaturing PAGE on the RNA digests. To gain insight into the three-dimensional structure of the gRNA-tRNA<sup>Lys,3</sup> complex, a complementary study with hydroxyl radicals was also performed [73]. In 1997, a variety of chemical probes, including lead ions, were used to study the influence of the purines flanking the dimerization initiation site (DIS) palindromic sequence on HIV-1 gRNA dimerization [76]. Pb<sup>2+</sup> ions were used to detect subtle structural alterations between monomeric and dimeric forms of gRNA. The data allowed for the conclusion that the three flanking purines played major roles in the kinetics of RNA dimerization and in the stability of the dimers formed [76]. Later on, lead ions were also used to monitor the specific binding of aminoglycosides at the DIS in vitro and in cells [97]. Structural studies of the 5'-UTR (untranslated region) of HIV-1 gRNA were important for understanding the fate of this RNA between translation and packaging. Based on enzymatic and chemical probing, different models with alternative long-range interactions have been published, which could drive gRNA to the translation or packaging pathways [98–100].

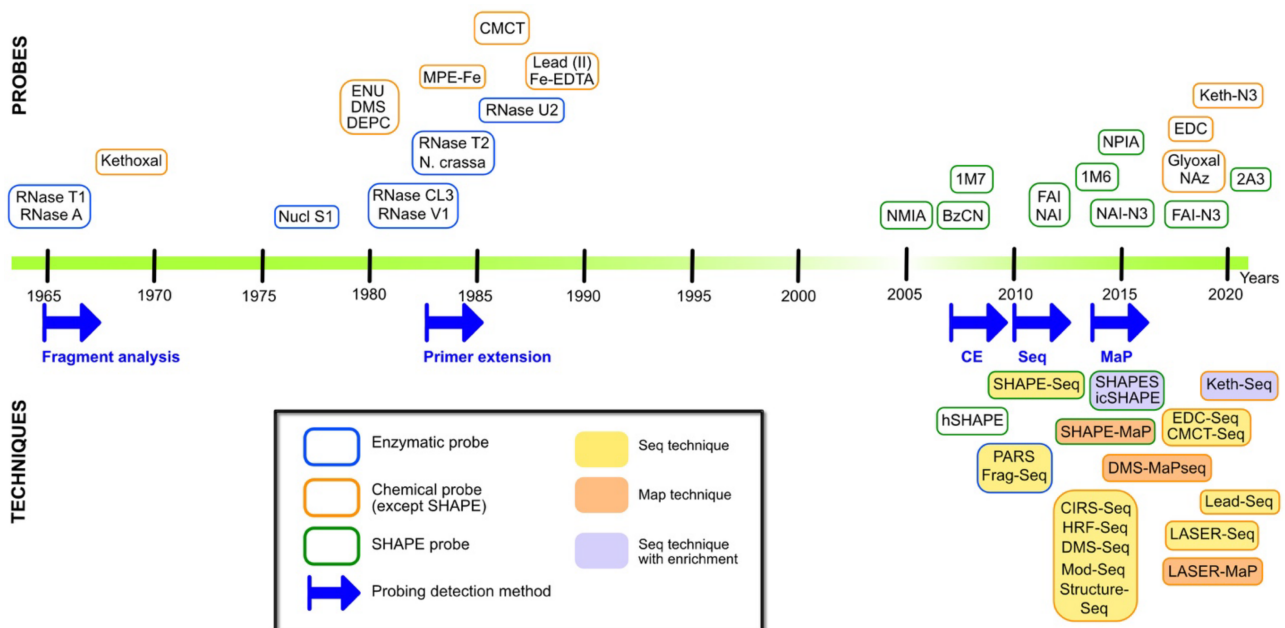
#### Enzymatic Viral RNAs: Self-Cleaving Ribozymes

Many viral RNA structures are known to carry out important functions during replication. These functions often require the viral or/and cellular partners recruited by structural RNA motifs. However, some of the structures present in viruses, viroids, and virusoids are functional on their own once they have adopted a specific 3D structure [101]. This is the case for viral ribozymes, such as the hammerhead ribozymes, which are found in many ss circular RNAs and are responsible for the self-cleaving activity of the RNA during rolling circle replication. The first identified RNA with autocatalytic cleavage ability was the *Tetrahymena thermophila* IVS RNA [102]. In this study, the authors established that the secondary/tertiary structure of the RNA was involved in the cleaving activity. In 1989, this activity was also found in the hepatitis D virus (HDV) [103], and the secondary structures associated with the self-cleavage function in gRNA and antigenomic RNA were confirmed through enzymatic probing with RNases A, V1, T1, U2 [60], and [<sup>32</sup>P] 5'-end radiolabeled RNA. Since the discovery of these self-cleaving RNAs, they have been widely used as tools for controlling gene expression [104].

### 3. The SHAPE (r)Evolution

Since the beginning of chemical probing, probes have been selected for their ability to highlight the exposure to the solvent, flexibility, or ss nature of the nucleotides. A combination of a few probes was usually required, in order to establish an experimentally validated secondary structure model of an RNA of interest. This changed in 2005, with the work of Merino et al. [105]. Indeed, in previous work on 2'-NH<sub>2</sub>-nucleotides, this laboratory had noticed that the 2' position of nucleotides with a flexible sugar ring, which generally correspond to unpaired nucleotides, was more reactive than that of base-paired nucleotides [106]. Therefore, they designed a new chemical probe, the N-methylisatoic anhydride (NMIA), capable of acylating the 2'-hydroxyl group of nucleosides and tested

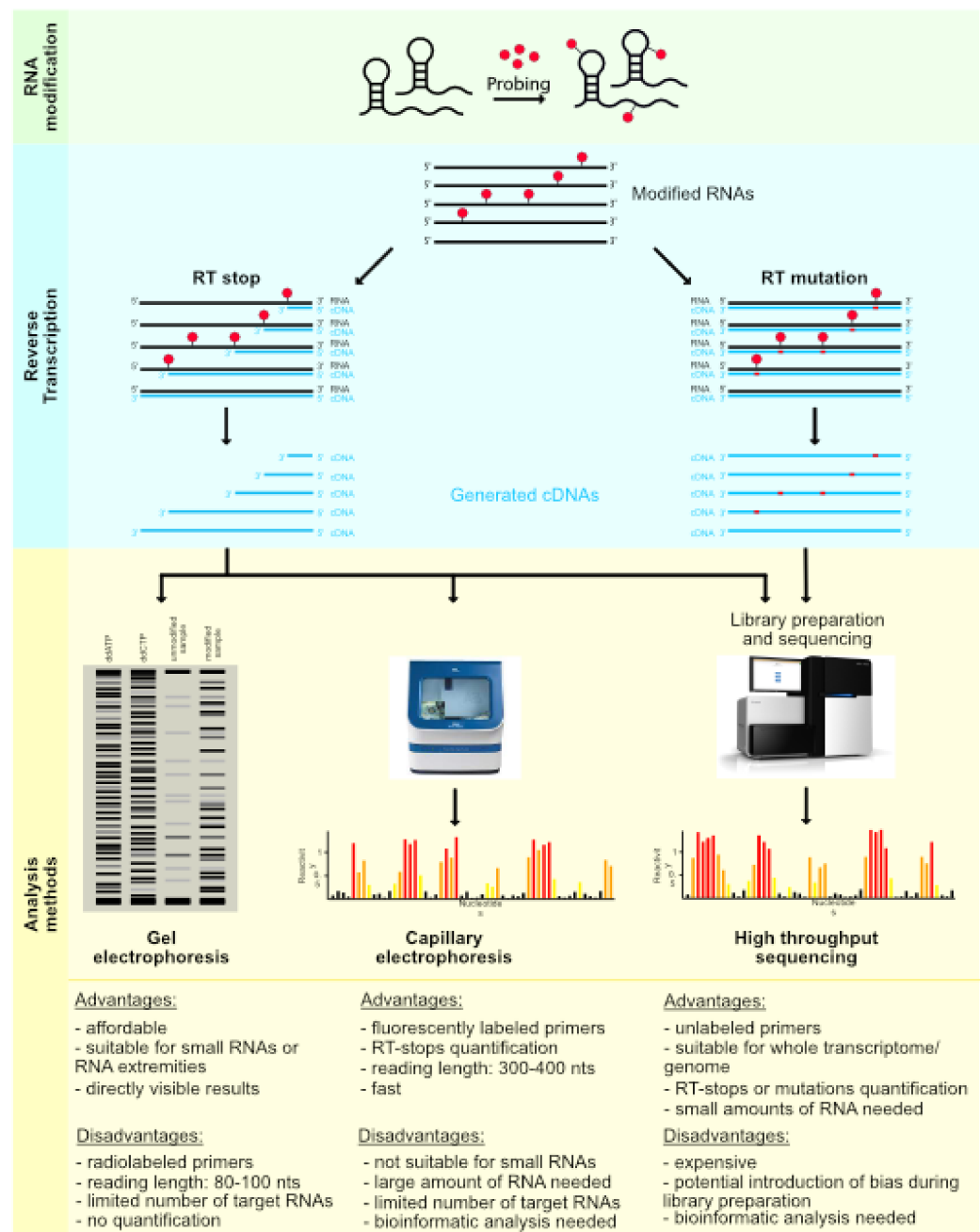
its potential as a structural probe by studying the structure of yeast tRNA<sup>Asp</sup> [105]. The resulting mapping method was named SHAPE for selective 2'-hydroxyl acylation analyzed by primer extension (Figures 1 and 2).



**Figure 2.** Timeline of RNA probes and probing techniques evolution. Upper part, enzymes are boxed in blue, chemical reagents (with the exception of SHAPE reagents) are boxed in orange, and SHAPE probes are boxed in green; the center of the box corresponds to the year of publication. Lower part, the blue arrows indicate the implementation of a new method to detect sites of cleavage or modification. CE: capillary electrophoresis; Seq: deep-sequencing; MaP: mutational profiling. Methods combining probing and deep-sequencing are highlighted in yellow. Methods based on mutational profiling are highlighted in orange, and techniques using enrichment by selection via the probe are labeled in purple. PARS: parallel analysis of RNA structure; FragSeq: fragmentation sequencing; CIRS-Seq: DMS and CMCT probing and sequencing; HRF-Seq: hydroxyl radical cleavage and sequencing; Mod-Seq: DMS probing and sequencing; Structure-Seq: DMS probing and sequencing; SHAPES: probing with NPIA and selection; icSHAPE: probing with NAI-N<sub>3</sub> and selection; Keth-seq: N<sub>3</sub>-kethoxal probing, selection, and sequencing; LASER/-seq/-MaP: light-activated structural examination of RNA/analyzed by sequencing/mutational profiling.

The first notable improvement brought by SHAPE was that it allowed for the interrogation of the 4 nucleotides of RNA in a single experiment and, thus, to map all the reactivities of an RNA molecule using a single probe [105,107,108]. NMIA was only the first of a series of reagents that emerged in the following years (Table 2). First, introducing a nitro group in para of the electrophilic carbonyl, 1-methyl-7-nitroisatoic anhydride (1M7) was found to be much more reactive than NMIA (half live of 14 s vs. 10 min) [109,110] and the reaction kinetic was found to be even faster for the commercial reagent BzCN, which reacts on 2'-OH in 0.25 s. [109]. While BzCN is well suited for the study of dynamic RNAs and transient RNA structures [111], other compounds were designed to be less reactive, in order to be compatible with the study of RNA structures. This is particularly the case for 2-methyl nicotinic acid imidazolide (NAI) and 2-methyl-3-furoic acid imidazolide (FAI) ( $t_{1/2} = 33$  min and 73 min, respectively) [112,113].

Another important improvement concerns the coupling of SHAPE mapping with adduct detection by capillary electrophoresis (CE), in order to allow for high-throughput and quantitative analysis (hSHAPE). Indeed, the initial approach consisted of extending radioactive primers and visualizing the cDNAs obtained by PAGE, but the methodology quickly evolved with the use of fluorescently labeled primers. In that case, the cDNAs obtained after the reverse transcription step can be separated by capillary electrophoresis and quantified on DNA sequencers [109] (Figure 3).



**Figure 3.** Schematic representation of RNA probing methods based on chemical modification. Folded RNAs were modified with the chemical probe of interest (green panel). For the RT stop strategy (left, blue panel), modified RNAs were reverse-transcribed with (1) radiolabeled primers for gel electrophoresis visualization (left, yellow panel), (2) fluorescently labeled primers for capillary electrophoresis detection (center, yellow panel), or (3) unlabeled primers for high-throughput sequencing after adapter ligation and library preparation (right, yellow panel). For the mutational profiling approach (right, blue panel), the modified RNAs were reverse-transcribed under mutation-inducing conditions with unlabeled primers. The cDNAs were ligated to adapters, and a library was prepared for high-throughput sequencing (right, yellow panel). The main advantages and disadvantages of the three analysis methods were featured at the bottom of the yellow panel.

The switch from PAGE to capillary electrophoresis has had multiple consequences. First of all, the reproducibility and resolution of the results have greatly improved, which has allowed for obtaining information for every nucleotide and to increase the length of the reading windows from approximately 80–100 nucleotides to 300–400 nucleotides

in a single experiment, making the study of large RNAs more accessible [114]. Second, the number of experiments to be analyzed in one working day (with a DNA sequencing analyzer instrument equipped with a capillary array) was also increased, allowing for the study of more RNAs and/or experimental conditions. Thirdly, the new format of the generated data is digital, similar to DNA sequencing traces, so they must be processed before being used. In addition, the analysis must be able to compare the different experiments; consequently, there was a need for bioinformatics tools to handle all these steps in efficient, automated, and accurate processes. Over the years, several software have been developed to solve this problem, such as CAFA [115], ShapeFinder [114], HiTRACE [116], FAST [117], QuShape [118], and RiboCAT [119]. Some tools have been developed in recent years to complement the analysis performed by the previously mentioned software, such as RNATHOR [120] and RNAPROBE [121].

**Table 2.** SHAPE and post-SHAPE chemical probes.

Probe	Target	In Cell/In Viro Probing	Early Use as Structural Probe
SHAPE	2'-OH ssN	Yes except BzCN	[105,107–109,112,122,123]
EDC	N3-ssU and N1-ssG	Yes	[124]
	N1-N2-ssG		
Glyoxal	N1-N6-ssA	Yes	[40]
	N3-N4-ssC		
NAz	C8-A and C8-G	Yes	[125]

SHAPE was successfully applied to the study of the RNA structures of several human pathogens. Using 1M7, Watts et al. [126] were able to interrogate 99.4% of the 9173 nucleotides of HIV-1 gRNA and to obtain a complete SHAPE-validated RNA secondary structure model. Poliovirus RNA was also studied with the NMIA probe [127] and hepatitis C virus with NAI [128].

Besides the hSHAPE technique, other “old probes” have also been coupled to CE detection, in order to gain more information on RNA structures. Actually, SHAPE probes interrogate the sugar face of the RNA, but information on the Watson–Crick face accessibility are also of interest; chemical probes, such as DMS [129–131] and CMCT [130,131], which interrogate the nucleobases, along with RNases [130], have successfully been applied to RNA structure determination in a way similar to hSHAPE.

## 4. A New Era for Enzymatic and Chemical Probing

### 4.1. The Deep-Sequencing Leaps

Even if hSHAPE and its further modifications constituted important advances in RNA structure determination by decreasing the workload of the experiments and facilitating quantification, the improved, but still limited, output and the large amount of RNA required for the detection of the modified nucleotides were the two biggest limitations of this technique. Indeed, for studies that are limited to small regions (and when the amount of RNA available is low), it is preferable to use the older method, consisting of the extension of radiolabeled primers, followed by analysis by PAGE, instead of the extension of fluorescent primers analyzed on a DNA sequencer. The coupling of RNA probing with next-generation sequencing (NGS) allowed the community to overcome these limitations and step into a new era of transcriptome- and genome-wide studies of RNA structure (structurome).

Structurome studies began in 2010 with enzymatic probes, and two methods were developed almost simultaneously: parallel analysis of RNA structure (PARS) [132] and fragmentation sequencing (Frag-Seq) [133]. PARS combined nuclease S1 and RNase V1 digestion, providing information on the ss- and dsRNA regions, respectively, followed by high-throughput sequencing of the dsDNA library, whereas Frag-Seq used only one enzyme (nuclease P1) but took two controls into account to gain information. Although the PARS method has been improved, notably in the form of PARTE [134,135] and nextPARS [136],

methods involving enzyme probes have given way to chemical probe-based techniques. Indeed, soon after the development of PARS and Frag-seq, SHAPE probes were coupled to NGS (SHAPE-Seq [137] and SHAPE-Seq 2.0 [138]), since, unlike enzymes, SHAPE probes can be used in cells, as well as in a wider range of temperature and buffer conditions (Figure 2). The millions of reads generated by deep-sequencing make the SHAPE-Seq a more sensitive technique than hSHAPE. Furthermore, this method can be adapted to get transcriptome-wide structural information or by using sequence-specific reverse transcription primers to focus on the structure of an RNA of interest [12,139]. However, particular attention must be paid to the construction of the library because biases can be introduced at this level, especially during the ligation of adapters and/or during the PCR amplification step (Figure 3).

#### 4.2. Old Probes Back in the Spotlight

Given the advances made possible by SHAPE-seq, “old” probes have also been revisited, in combination with deep-sequencing detection. A multitude of techniques have emerged that often vary only in one of the steps of the library preparation. HRF-Seq [140] takes advantage of hydroxyl radical cleavages; DMS was used in structure-Seq [141], DMS-Seq [142], Mod-Seq [143], and CIRS-seq [144]. This last technique actually combines DMS and CMCT to probe all four nucleotides. More recently, *N*-cyclohexyl *N'*-(2-morpholinoethyl) carbodiimide, the reactive group of CMCT, was also used in the Seq technique to map the pseudouridine positions in rRNA [145], as well as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), a cell permeable derivative of CMCT [124,146] (Figure 2 and Table 2). In a similar way, lead(II) [147] has been coupled with NGS and is also experiencing a renewal. Since each probe possesses its own bias or specificity, increasing the number of probes available for studying RNA structures and RNA-protein interactions allows us to provide more pieces to the RNA and RNP structure puzzles.

The increasing evidence that G-quadruplexes (G4s) structures are important in the virus life cycle [148,149], and the fact that they may constitute a therapeutic target, has led to the application of the high-throughput mapping approaches for their study (for a review, see [150]). The previously established techniques, such as RT stop footprinting [151] and SHALiPE [150], have been improved to rG4-Seq [152], SHALiPE-Seq [153], or G4RP-Seq [154]. Another high-throughput based method, Keth-Seq [155], can also be applied to G4s studies, even if it is a non-rG4-specific method and it cannot differentiate G4s from dsGs.

Moreover, other improvements have been made to the “PROBE-Seq” techniques by introducing a second and orthogonal reactive group inside pre-existing probes. Using bi-functional probes, such as *N*-propanone isatoic anhydride (NPIA) in the SHAPES technique [122], NAI-N3 and FAI-N3 in icSHAPE [123] (which has been recently used to probe SARS-CoV-2 gRNA structure [156]), and Khetoxal-N3 in Keth-Seq [155], it was possible to link biotin to the modified RNA molecules via a click chemistry reaction. Selection with streptavidin beads allowed for the enrichment of the RNA of interest, thereby increasing the signal-to-noise ratio and significantly improving the sensitivity.

#### 4.3. Mutational Profiling

Mutational profiling (MaP) was perhaps the most important improvement in determining the structure of RNAs by simplifying and reducing biases during library preparation. It was first adapted to SHAPE as SHAPE-MaP [157,158] and later, with the widely used DMS probe, as DMS-MaPseq [159]. These new techniques are reminiscent of whole-genome bisulfite sequencing (WGBS) [160], which has allowed the study of the human DNA methylomes by detecting mutations. These techniques use the capacity of reverse transcriptase to misread SHAPE or DMS-modified nucleotides and incorporate mutations in the newly synthesized cDNA. Whereas in the SHAPE-Seq and DMS-Seq strategies, modifications are identified as 3'-ends of the cDNA library, generated by stops during the reverse transcription; in the MaP strategies, mutations are detected all along the cDNA fragments. The

generated cDNA library presents more homogeneity in size and can give more information, as it can contain multiple probe-induced mutations inside the same fragment (Figure 3). These features make it possible to get rid of the adapter ligation bias and to better identify the positions and frequencies of modification, even with very small amounts of RNA.

Recently, new chemical probes were designed and adapted to the MaP strategy (Figures 1 and 2). Under light-activation, nicotinoyl-azide (NAz) generates nicotinoyl nitrenium ions that react with the C8 position of Gs and As, whether they are ss or ds. This probe was originally used in a “classical way” in the LASER (light activated structural examination of RNA) technique [125], to interrogate the RNA solvent accessibility and the ligand binding sites. However, NAz is compatible with deep-sequencing in LASER-Seq and LASER-MaP [161], as it induces RT-stops, as well as misreadings. A new SHAPE probe, compatible with mutational profiling, was also developed to improve in cellula studies: 2-aminopyridine-3-carboxylic acid imidazolide (2A3) [162].

#### 4.4. The Latest Developments

Nevertheless, the probing techniques described above are still limited because, for example, they do not directly identify base-pairing interactions, rather they infer them from the compatibility that may exist between the observed reactivity and the proposed structure. That is why even more sophisticated methods have been developed. Among them, the “RNA interaction groups analyzed by mutational profiling” (RING-MaP) technique [163] allows for the identification of nucleotides that interact with each other. Indeed, when RNA is highly modified with DMS, these nucleotides are seen, after reverse transcription, as correlated patterns of mutations in the cDNAs [164]. Depending on how the correlation data are interpreted, it is possible with this technique to (1) detect RNA duplexes with the “pairing ascertained from interacting RNA strands measured by mutational profiling” (PAIR-MaP) method [164,165], (2) identify multiple RNA subpopulation structures present in the solution [163], and (3) reveal tertiary interactions [163,166,167]. This technique was further improved by replacing DMS with the highly reactive trimethyl oxonium (TMO), which allowed for the time-resolved probing of RNA [168]. In order to identify alternative conformations of RNAs, data from DMS-MaPseq experiments can also be analyzed by “detection of RNA folding ensembles using expectation-maximization” (DREEM) [169]. This algorithm allows the detection of heterogeneous regions inside the HIV-1 gRNA (and more precisely, at splice sites) that may explain the regulation of the gRNA packaging. In another approach, merging mutate-and-map [170,171] with massively parallel DNA sequencing technology in M2-seq [172] generates helix signatures. Here, mutations are first introduced by error-prone PCR, and the effects of the mutations on nucleotide-pairing are detected by DMS probing.

Another interesting application of MaP is the localization of the protein interaction sites on RNAs by the RNP-MaP technique [173]. Indeed, with classical mapping methods, it is quite complicated to differentiate nucleotides that are protected by base pairing from those protected by the binding of a protein. In RNP-MaP, the cell permeable reagent NHS-diazirine reacts with the lysine of proteins, thanks to its succinimidyl ester function and with nucleic acids, after the activation of its diazirine moiety by long-wavelength UV. The RNA-protein crosslinks are then detected by mutational profiling reverse transcription, and the protein binding sites can be deduced in a more reliable way. This is a very interesting first implementation for footprint experiments, but the development of a set of probes, similar to what happened with the chemical mapping probes, would be a plus to extend this technique.

There is no doubt that these latest developments have greatly improved the study of RNA structures and will pave the way for more in-depth studies. However, performing such experiments, involving a high rate of RNA chemical modifications, may induce structural changes during the probing experiment, and careful controls must be carried out to avoid analysis bias.

#### 4.5. Applications of the MAP Strategies to Viral RNAs

These powerful techniques were applied to the study of several viruses. In recent years, the global structure of the influenza A virus (IAV) genome was undertaken by SHAPE-MaP experiments, with a 1M7 probe in vitro and inside virions [174]. The authors were able to detect both intra- and inter-interactions between the ss viral RNA fragments. DMS-MaPseq was also used to better understand IAV infection, at that time by studying mRNA in infected living cells; Simon et al. [175] identified the stable structural domains important for IAV replicative capacity. The SHAPE-MaP technique with 1M7 has also been used to study the chikungunya virus (CHIKV) gRNA [176]. The high-throughput flow allowed for the identification of regions where the RNA secondary structure is conserved between CHIKV isolates; some of them, already identified, are known for their importance in the functionality of the virus.

Since 2020, the highly pathogenic coronavirus, severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), has also been studied with mutational profiling approaches. SHAPE-MaP [177,178] and DMS-MaPseq [179,180] have been used to study the full RNA genome in infected cells, and these techniques identified conserved ss segments, as well as folded structures that could constitute RNA drug targets. It is reasonable to think that in the coming years, the expansion of the MaP strategy to the other chemical probes will be realized for the study of RNA structures.

### 5. Conclusions

Research on enzymatic and chemical probes in the second half of the last century has led, in barely 20 years, to the creation of a toolbox of probes that complement each other and that have allowed for the identification of many RNAs structures, as well as ligand-binding sites. Quickly after their discovery, these techniques have successfully been applied to the study of viral RNAs and have enabled a better comprehension of the biological roles of their structural features. However, the notable shortcomings of these methods were their inability to quickly provide robust structural information, as the generation of data was slow and quantification of the experiments was difficult, complicating data analysis.

In 2005, these shortcomings were addressed with the development of a new class of chemical probes targeting the sugar 2'-OH, regardless of the nature of the base and, later, the use of fluorescent primers and analysis of the cDNAs by capillary electrophoresis. These advances gave a new impetus to the RNA structural probing field by allowing for large-scale analyses in a shorter time, with quantifiable results that were easier to reproduce and compare.

Most importantly, these steps in the direction of high-throughput probing have paved the way for even more performant techniques: the next-generation sequencing-based RNA structure probing technologies. The methods initially developed with enzymatic probes were abandoned over time, in favor of the more versatile chemical probes. Indeed, one after another, almost all the former chemical probes have been coupled to deep-sequencing, which provides a more diverse and adaptable array of tools, available for the study of many RNA species and contexts.

### 6. Perspectives

As the novel combinations of probes with various methods based on NGS technology become more common, new challenges will arise in the RNA structural probing field. Thanks to the decades of development of the processing algorithms, the RNA secondary structures can be predicted in silico with or without probing information, even if experimental data are required to sharpen and confirm the predicted models. This points toward an increasing need for tools, in order to address the 3D structures of RNA molecules, as many functions and mechanisms rely on long-range interactions and complex folding structures. Some steps have already been taken toward this goal, with the recent development of SHAPE-JuMP [181], but also with other methods, such as LIGR-seq [182], PARIS [183], COMRADE [184], and MOHCA-seq [185].

Besides, a new window is presently opening in the field of RNA probing with nanopore sequencing, which allows for the direct detection of the RNA modification on individual (even very long) molecules, without the need for a reverse transcription step. The coupling of nanopore sequencing with RNA sequencing should prove especially useful in the analysis of co-existing RNA structures, which is difficult with ensemble techniques [186].

Finally, it is also crucial to remember that the structural information in the RNA field is required to decipher the functions carried out by RNA molecules. In order to establish a clear link between RNA motifs and their molecular functions, new ultrahigh-throughput methods were developed, which showed promising results in predicting the impact on single elements of RNA functions [187].

**Author Contributions:** Conceptualization, R.M.; writing—original draft preparation, O.G., E.Q., A.-C.J. and V.V.-B., writing—review and editing, O.G., A.-C.J., J.-C.P., R.M. and V.V.-B. 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 the CNRS. O.G. and E.Q. were supported by fellowships from the French ministry of Research and Higher Education, as well as from SIDACTION (O.G.) and University of Strasbourg (E.Q.).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors thank Pascale Romby (UPR 9002 CNRS) and Anjana Krishnan (College of Medicine and Health Sciences, UAE) for the critical reading of this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Markham, R.; Smith, K.M. Studies on the Virus of Turnip Yellow Mosaic. *Parasitology* **1949**, *39*, 330–342. [[CrossRef](#)]
2. Gierer, A.; Schramm, G. Infectivity of Ribonucleic Acid from Tobacco Mosaic Virus. *Nature* **1956**, *177*, 702–703. [[CrossRef](#)]
3. Ennifar, E.; Walter, P.; Ehresmann, B.; Ehresmann, C.; Dumas, P. Crystal Structures of Coaxially Stacked Kissing Complexes of the HIV-1 RNA Dimerization Initiation Site. *Nat. Struct. Biol.* **2001**, *8*, 1064–1068. [[CrossRef](#)] [[PubMed](#)]
4. Suddala, K.C.; Zhang, J. High-Affinity Recognition of Specific TRNAs by an mRNA Anticodon-Binding Groove. *Nat. Struct. Mol. Biol.* **2019**, *26*, 1114–1122. [[CrossRef](#)] [[PubMed](#)]
5. Keane, S.C.; Heng, X.; Lu, K.; Kharytonchyk, S.; Ramakrishnan, V.; Carter, G.; Barton, S.; Hosic, A.; Florwick, A.; Santos, J.; et al. Structure of the HIV-1 RNA Packaging Signal. *Science* **2015**, *348*, 917–921. [[CrossRef](#)] [[PubMed](#)]
6. Zhang, K.; Li, S.; Kappel, K.; Pintilie, G.; Su, Z.; Mou, T.-C.; Schmid, M.F.; Das, R.; Chiu, W. Cryo-EM Structure of a 40 KDa SAM-IV Riboswitch RNA at 3.7 Å Resolution. *Nat. Commun.* **2019**, *10*, 5511. [[CrossRef](#)] [[PubMed](#)]
7. Kappel, K.; Zhang, K.; Su, Z.; Watkins, A.M.; Kladwang, W.; Li, S.; Pintilie, G.; Topkar, V.V.; Rangan, R.; Zheludev, I.N.; et al. Accelerated Cryo-EM-Guided Determination of Three-Dimensional RNA-Only Structures. *Nat. Methods* **2020**, *17*, 699–707. [[CrossRef](#)]
8. Mailler, E.; Paillart, J.-C.; Marquet, R.; Smyth, R.P.; Vivet-Boudou, V. The Evolution of RNA Structural Probing Methods: From Gels to next-Generation Sequencing. *WIREs RNA* **2019**, *10*, e1518. [[CrossRef](#)] [[PubMed](#)]
9. Rich, A. The Era of RNA Awakening: Structural Biology of RNA in the Early Years. *Q. Rev. Biophys.* **2009**, *42*, 117–137. [[CrossRef](#)]
10. Nguyen, T.C.; Zaleta-Rivera, K.; Huang, X.; Dai, X.; Zhong, S. RNA, Action through Interactions. *Trends Genet.* **2018**, *34*, 867–882. [[CrossRef](#)]
11. Kudla, G.; Wan, Y.; Helwak, A. RNA Conformation Capture by Proximity Ligation. *Annu. Rev. Genom. Hum. Genet.* **2020**, *21*, 81–100. [[CrossRef](#)]
12. Wang, X.-W.; Liu, C.-X.; Chen, L.-L.; Zhang, Q.C. RNA Structure Probing Uncovers RNA Structure-Dependent Biological Functions. *Nat. Chem. Biol.* **2021**, *17*, 755–766. [[CrossRef](#)]
13. Brown, D.M.; Heppel, L.A.; Hilmoie, R.J. Nucleotides. Part XXIV. The Action of Some Nucleases on Simple Esters of Monoribonucleotides. *J. Chem. Soc.* **1954**, 40–46. [[CrossRef](#)]
14. Rich, A.; Davies, D.R. A New Two Stranded Helical Structure: Polyadenylic Acid and Polyuridylic Acid. *J. Am. Chem. Soc.* **1956**, *78*, 3548–3549. [[CrossRef](#)]
15. Felsenfeld, G.; Davies, D.R.; Rich, A. Formation of a Three-Stranded Polynucleotide Molecule. *J. Am. Chem. Soc.* **1957**, *79*, 2023–2024. [[CrossRef](#)]
16. Dubos, R.J. The Decomposition of Yeast Nucleic Acid by a Heat Resistant Enzyme. *Science* **1937**, *85*, 549. [[CrossRef](#)]



17. Sato, K.; Egami, F. Studies on Ribonucleases in Takadiastase. I. *J. Biochem.* **1957**, *44*, 753–767. [[CrossRef](#)]
18. Gomatos, P.J.; Tamm, I.; Dales, S.; Franklin, R.M. Reovirus Type 3: Physical Characteristics and Interaction with L Cells. *Virology* **1962**, *17*, 441–454. [[CrossRef](#)]
19. Cantoni, G.L.; Gelboin, H.V.; Luborsky, S.W.; Richards, H.H.; Singer, M.F. Studies on Soluble Ribonucleic Acid of Rabbit Liver III. Preparation and Properties of Rabbit-Liver Soluble RNA. *Biochim. Biophys. Acta* **1962**, *61*, 354–367. [[CrossRef](#)]
20. Geiduschek, E.P.; Moohr, J.W.; Weiss, S.B. The Secondary Structure of Complementary RNA. *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 1078–1086. [[CrossRef](#)]
21. Holley, R.W.; Everett, G.A.; Madison, J.T.; Zamir, A. Nucleotide Sequences in the Yeast Alanine Transfer Ribonucleic Acid. *J. Biol. Chem.* **1965**, *240*, 2122–2128. [[CrossRef](#)]
22. Sato, K.; Egami, F. The specificity of T1 ribonuclease. *Comptes Rendus Seances Soc. Biol. Fil.* **1957**, *151*, 1792–1796.
23. Doctor, B.P.; Connelly, C.M.; Rushizky, G.W.; Sober, H.A. Studies on the Chemical Structure of Yeast Amino Acid Acceptor Ribonucleic Acids. *J. Biol. Chem.* **1963**, *238*, 3985–3990. [[CrossRef](#)]
24. Elmore, D.T.; Gulland, J.M.; Jordan, D.O.; Taylor, H.F.W. The Reaction of Nucleic Acids with Mustard Gas. *Biochem. J.* **1948**, *42*, 308–316. [[CrossRef](#)] [[PubMed](#)]
25. Fraenkel-Conrat, H.; Singer, B.; Tsugita, A. Purification of Viral RNA by Means of Bentonite. *Virology* **1961**, *14*, 54–58. [[CrossRef](#)]
26. Lawley, P.D.; Brookes, P. Further Studies on the Alkylation of Nucleic Acids and Their Constituent Nucleotides. *Biochem. J.* **1963**, *89*, 127–138. [[CrossRef](#)] [[PubMed](#)]
27. Gilham, P.T. An Addition Reaction Specific for Uridine and Guanosine Nucleotides and Its Application to the Modification of Ribonuclease Action. *J. Am. Chem. Soc.* **1962**, *84*, 687–688. [[CrossRef](#)]
28. Augusti-Tocco, G.; Brown, G.L. Reaction of N-Cyclohexyl, N'-Beta (4-Methylmorpholinium) Ethyl Carbodiimide Iodide with Nucleic Acids and Polynucleotides. *Nature* **1965**, *206*, 683–685. [[CrossRef](#)]
29. Naylor, R.; Ho, N.W.Y.; Gilham, P.T. Selective Chemical Modifications of Uridine and Pseudouridine in Polynucleotides and Their Effect on the Specificities of Ribonuclease and Phosphodiesterases. *J. Am. Chem. Soc.* **1965**, *87*, 4209–4210. [[CrossRef](#)]
30. Linn, S.; Lehman, I.R. An Endonuclease from *Neurospora Crassa* Specific for Polynucleotides Lacking an Ordered Structure. *J. Biol. Chem.* **1965**, *240*, 1294–1304. [[CrossRef](#)]
31. Harada, F.; Dahlberg, J.E. Specific Cleavage of tRNA by Nuclease S1. *Nucleic Acids Res.* **1975**, *2*, 865–871. [[CrossRef](#)]
32. Vary, C.P.; Vournakis, J.N. RNA Structure Analysis Using T2 Ribonuclease: Detection of PH and Metal Ion Induced Conformational Changes in Yeast tRNA<sup>Phe</sup>. *Nucleic Acids Res.* **1984**, *12*, 6763–6778. [[CrossRef](#)]
33. Levy, C.C.; Karpetsky, T.P. The Purification and Properties of Chicken Liver RNase: An Enzyme Which Is Useful in Distinguishing between Cytidylic and Uridydic Acid Residues. *J. Biol. Chem.* **1980**, *255*, 2153–2159. [[CrossRef](#)]
34. Uchida, T.; Arima, T.; Egami, F. Specificity of RNase U2. *J. Biochem.* **1970**, *67*, 91–102. [[CrossRef](#)]
35. Holley, R.W.; Apgar, J.; Everett, G.A.; Madison, J.T.; Marquisee, M.; Merrill, S.H.; Penswick, J.R.; Zamir, A. Structure of a Ribonucleic Acid. *Science* **1965**, *147*, 1462–1465. [[CrossRef](#)]
36. Vasilenko, S.K.; Babkina, G.T. Isolation and properties of ribonuclease isolated from cobra venom. *Biokhimiia* **1965**, *30*, 705–712.
37. Vasilenko, S.K.; Rytte, V.C. Isolation of highly purified ribonuclease from cobra (*Naja oxiana*) venom. *Biokhimiia* **1975**, *40*, 578–583.
38. Favorova, O.O.; Fasiolo, F.; Keith, G.; Vassilenko, S.K.; Ebel, J.P. Partial Digestion of tRNA—Aminoacyl-tRNA Synthetase Complexes with Cobra Venom Ribonuclease. *Biochemistry* **1981**, *20*, 1006–1011. [[CrossRef](#)]
39. Desai, N.A.; Shankar, V. Single-Strand-Specific Nucleases. *FEMS Microbiol. Rev.* **2003**, *26*, 457–491. [[CrossRef](#)]
40. Mitchell, D.; Ritchey, L.E.; Park, H.; Babbitzke, P.; Assmann, S.M.; Bevilacqua, P.C. Glyoxals as in Vivo RNA Structural Probes of Guanine Base-Pairing. *RNA* **2018**, *24*, 114–124. [[CrossRef](#)]
41. Staehelin, M. Inactivation of Virus Nucleic Acid with Glyoxal Derivatives. *Biochim. Biophys. Acta* **1959**, *31*, 448–454. [[CrossRef](#)]
42. Litt, M. Structural Studies on Transfer Ribonucleic Acid. I. Labeling of Exposed Guanine Sites in Yeast Phenylalanine Transfer Ribonucleic Acid with Kethoxal. *Biochemistry* **1969**, *8*, 3249–3253. [[CrossRef](#)]
43. Oberg, B. Biochemical and Biological Characteristics of Carboxylated Poliovirus and Viral RNA. *Biochim. Biophys. Acta* **1970**, *204*, 430–440. [[CrossRef](#)]
44. Leonard, N.J.; McDonald, J.J.; Henderson, R.E.L.; Reichmann, M.E. Reaction of Diethyl Pyrocarbonate with Nucleic Acid Components. Adenosine. *Biochemistry* **1971**, *10*, 3335–3342. [[CrossRef](#)] [[PubMed](#)]
45. Peattie, D.A.; Gilbert, W. Chemical Probes for Higher-Order Structure in RNA. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 4679–4682. [[CrossRef](#)]
46. Stern, S.; Wilson, R.C.; Noller, H.F. Localization of the Binding Site for Protein S4 on 16 S Ribosomal RNA by Chemical and Enzymatic Probing and Primer Extension. *J. Mol. Biol.* **1986**, *192*, 101–110. [[CrossRef](#)]
47. Inoue, T.; Cech, T.R. Secondary Structure of the Circular Form of the Tetrahymena rRNA Intervening Sequence: A Technique for RNA Structure Analysis Using Chemical Probes and Reverse Transcriptase. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 648–652. [[CrossRef](#)]
48. Guo, J.U.; Bartel, D.P. RNA G-Quadruplexes Are Globally Unfolded in Eukaryotic Cells and Depleted in Bacteria. *Science* **2016**, *353*, aaf5371. [[CrossRef](#)]
49. Singer, B. All Oxygens in Nucleic Acids React with Carcinogenic Ethylating Agents. *Nature* **1976**, *264*, 333–339. [[CrossRef](#)]
50. Vlassov, V.V.; Giege, R.; Ebel, J.P. The Tertiary Structure of Yeast tRNA<sup>Phe</sup> in Solution Studied by Phosphodiester Bond Modification with Ethylnitrosourea. *FEBS Lett.* **1980**, *120*, 12–16. [[CrossRef](#)]

51. Wang, X.D.; Padgett, R.A. Hydroxyl Radical “Footprinting” of RNA: Application to Pre-mRNA Splicing Complexes. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 7795–7799. [[CrossRef](#)]
52. Vary, C.P.; Vournakis, J.N. RNA Structure Analysis Using Methidiumpropyl-EDTA.Fe(II): A Base-Pair-Specific RNA Structure Probe. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 6978–6982. [[CrossRef](#)]
53. Zamir, A.; Holley, R.W.; Marquisee, M. Evidence for the Occurrence of a Common Pentanucleotide Sequence in the Structures of Transfer Ribonucleic Acids. *J. Biol. Chem.* **1965**, *240*, 1267–1273. [[CrossRef](#)]
54. Wintermeyer, W.; Zachau, H.G. Tertiary Structure Interactions of 7-Methylguanosine in Yeast tRNA Phe as Studied by Borohydride Reduction. *FEBS Lett.* **1975**, *58*, 306–309. [[CrossRef](#)]
55. Ehresmann, C.; Baudin, F.; Mougél, M.; Romby, P.; Ebel, J.P.; Ehresmann, B. Probing the Structure of RNAs in Solution. *Nucleic Acids Res.* **1987**, *15*, 9109–9128. [[CrossRef](#)]
56. Peattie, D.A. Direct Chemical Method for Sequencing RNA. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 1760–1764. [[CrossRef](#)]
57. Kuśmierk, J.T.; Singer, B. Sites of Alkylation of Poly(U) by Agents of Varying Carcinogenicity and Stability of Products. *Biochim. Biophys. Acta* **1976**, *442*, 420–431. [[CrossRef](#)]
58. Markham, R.; Smith, J.D. The Structure of Ribonucleic Acid. I. Cyclic Nucleotides Produced by Ribonuclease and by Alkaline Hydrolysis. *Biochem. J.* **1952**, *552*–557. [[CrossRef](#)]
59. Rosenstein, S.P.; Been, M.D. Evidence That Genomic and Antigenomic RNA Self-Cleaving Elements from Hepatitis Delta Virus Have Similar Secondary Structures. *Nucleic Acids Res.* **1991**, *19*, 5409–5416. [[CrossRef](#)]
60. Florentz, C.; Briand, J.P.; Romby, P.; Hirth, L.; Ebel, J.P.; Glegé, R. The tRNA-like Structure of Turnip Yellow Mosaic Virus RNA: Structural Organization of the Last 159 Nucleotides from the 3′ OH Terminus. *EMBO J.* **1982**, *1*, 269–276. [[CrossRef](#)]
61. Rietveld, K.; Van Poelgeest, R.; Pleij, C.W.; Van Boom, J.H.; Bosch, L. The tRNA-like Structure at the 3′ Terminus of Turnip Yellow Mosaic Virus RNA. Differences and Similarities with Canonical tRNA. *Nucleic Acids Res.* **1982**, *10*, 1929–1946. [[CrossRef](#)] [[PubMed](#)]
62. Nishikawa, K.; Takemura, S. Structure and Function of 5S Ribosomal Ribonucleic Acid from *Torulopsis Utilis*. *J. Biochem.* **1977**, *81*, 995–1003. [[CrossRef](#)]
63. Ahlquist, P.; Dasgupta, R.; Kaesberg, P. Near Identity of 3′ RNA Secondary Structure in Bromoviruses and Cucumber Mosaic Virus. *Cell* **1981**, *23*, 183–189. [[CrossRef](#)]
64. Troutt, A.; Savin, T.J.; Curtiss, W.C.; Celentano, J.; Vournakis, J.N. Secondary Structure of Bombyx Mori and Dictyostelium Discoideum 5S rRNA from S1 Nuclease and Cobra Venom Ribonuclease Susceptibility, and Computer Assisted Analysis. *Nucleic Acids Res.* **1982**, *10*, 653–664. [[CrossRef](#)] [[PubMed](#)]
65. Skinner, M.A.; Racaniello, V.R.; Dunn, G.; Cooper, J.; Minor, P.D.; Almond, J.W. New Model for the Secondary Structure of the 5′ Non-Coding RNA of Poliovirus Is Supported by Biochemical and Genetic Data That Also Show That RNA Secondary Structure Is Important in Neurovirulence. *J. Mol. Biol.* **1989**, *207*, 379–392. [[CrossRef](#)]
66. Garret, M.; Labouesse, B.; Litvak, S.; Romby, P.; Ebel, J.P.; Giegé, R. Tertiary Structure of Animal tRNA<sup>Trp</sup> in Solution and Interaction of tRNA<sup>Trp</sup> with Tryptophanyl-tRNA Synthetase. *Eur. J. Biochem.* **1984**, *138*, 67–75. [[CrossRef](#)] [[PubMed](#)]
67. Isel, C.; Marquet, R.; Keith, G.; Ehresmann, C.; Ehresmann, B. Modified Nucleotides of tRNA(3Lys) Modulate Primer/Template Loop-Loop Interaction in the Initiation Complex of HIV-1 Reverse Transcription. *J. Biol. Chem.* **1993**, *268*, 25269–25272. [[CrossRef](#)]
68. Mougél, M.; Philippe, C.; Ebel, J.P.; Ehresmann, B.; Ehresmann, C. The E. Coli 16S rRNA Binding Site of Ribosomal Protein S15: Higher-Order Structure in the Absence and in the Presence of the Protein. *Nucleic Acids Res.* **1988**, *16*, 2825–2839. [[CrossRef](#)]
69. Glickman, J.N.; Howe, J.G.; Steitz, J.A. Structural Analyses of EBER1 and EBER2 Ribonucleoprotein Particles Present in Epstein-Barr Virus-Infected Cells. *J. Virol.* **1988**, *62*, 902–911. [[CrossRef](#)]
70. Hertzberg, R.P.; Dervan, P.B. Cleavage of Double Helical DNA by Methidium-Propyl-EDTA-Iron(II). *J. Am. Chem. Soc.* **1982**, *104*, 313–315. [[CrossRef](#)]
71. Lusvardi, S.; Sztuba-Solinska, J.; Purzycka, K.J.; Pauly, G.T.; Rausch, J.W.; Grice, S.F.J.L. The HIV-2 Rev-Response Element: Determining Secondary Structure and Defining Folding Intermediates. *Nucleic Acids Res.* **2013**, *41*, 6637–6649. [[CrossRef](#)]
72. Tullius, T.D.; Dombroski, B.A. Hydroxyl Radical “Footprinting”: High-Resolution Information about DNA-Protein Contacts and Application to Lambda Repressor and Cro Protein. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 5469–5473. [[CrossRef](#)]
73. Isel, C.; Westhof, E.; Massire, C.; Le Grice, S.F.J.; Ehresmann, B.; Ehresmann, C.; Marquet, R. Structural Basis for the Specificity of the Initiation of HIV-1 Reverse Transcription. *EMBO J.* **1999**, *18*, 1038–1048. [[CrossRef](#)]
74. Werner, C.; Krebs, B.; Keith, G.; Dirheimer, G. Specific Cleavages of Pure tRNAs by Plumbous Ions. *Biochim. Biophys. Acta* **1976**, *432*, 161–175. [[CrossRef](#)]
75. Gornicki, P.; Baudin, F.; Romby, P.; Wiewiorowski, M.; Kryzosiak, W.; Ebel, J.P.; Ehresmann, C.; Ehresmann, B. Use of Lead(II) to Probe the Structure of Large RNA’s. Conformation of the 3′ Terminal Domain of E. Coli 16S rRNA and Its Involvement in Building the tRNA Binding Sites. *J. Biomol. Struct. Dynam.* **1989**, *6*, 971–984. [[CrossRef](#)]
76. Paillart, J.C.; Westhof, E.; Ehresmann, C.; Ehresmann, B.; Marquet, R. Non-Canonical Interactions in a Kissing Loop Complex: The Dimerization Initiation Site of HIV-1 Genomic RNA. *J. Mol. Biol.* **1997**, *270*, 36–49. [[CrossRef](#)]
77. Qu, H.L.; Michot, B.; Bachelier, J.P. Improved Methods for Structure Probing in Large RNAs: A Rapid “heterologous” Sequencing Approach Is Coupled to the Direct Mapping of Nuclease Accessible Sites. Application to the 5′ Terminal Domain of Eukaryotic 28S rRNA. *Nucleic Acids Res.* **1983**, *11*, 5903–5920. [[CrossRef](#)]

78. Moazed, D.; Stern, S.; Noller, H.F. Rapid Chemical Probing of Conformation in 16 S Ribosomal RNA and 30 S Ribosomal Subunits Using Primer Extension. *J. Mol. Biol.* **1986**, *187*, 399–416. [[CrossRef](#)]
79. Pinck, M.; Yot, P.; Chapeville, F.; Duranton, H.M. Enzymatic Binding of Valine to the 3' End of TYMV-RNA. *Nature* **1970**, *226*, 954–956. [[CrossRef](#)]
80. Yamakawa, M.; Shatkin, A.J.; Furuichi, Y. Chemical Methylation of RNA and DNA Viral Genomes as a Probe of in Situ Structure. *J. Virol.* **1981**, *40*, 482–490. [[CrossRef](#)]
81. Joshi, R.L.; Joshi, S.; Chapeville, F.; Haenni, A.L. tRNA-like Structures of Plant Viral RNAs: Conformational Requirements for Adenylation and Aminoacylation. *EMBO J.* **1983**, *2*, 1123–1127. [[CrossRef](#)] [[PubMed](#)]
82. Langeberg, C.J.; Sherlock, M.E.; MacFadden, A.; Kieft, J.S. An Expanded Class of Histidine-Accepting Viral tRNA-like Structures. *RNA* **2021**, *27*, 653–664. [[CrossRef](#)] [[PubMed](#)]
83. Hammond, J.A.; Rambo, R.P.; Filbin, M.E.; Kieft, J.S. Comparison and Functional Implications of the 3D Architectures of Viral tRNA-like Structures. *RNA* **2009**, *15*, 294–307. [[CrossRef](#)] [[PubMed](#)]
84. Nicholson, B.L.; White, K.A. Context-Influenced Cap-Independent Translation of Tombusvirus MRNAs in Vitro. *Virology* **2008**, *380*, 203–212. [[CrossRef](#)]
85. Wu, B.; White, K.A. A Primary Determinant of Cap-Independent Translation Is Located in the 3'-Proximal Region of the Tomato Bushy Stunt Virus Genome. *J. Virol.* **1999**, *73*, 8982–8988. [[CrossRef](#)]
86. Danthinne, X.; Seurinck, J.; Meulewaeter, F.; Van Montagu, M.; Cornelissen, M. The 3' untranslated Region of Satellite Tobacco Necrosis Virus RNA Stimulates Translation in Vitro. *Mol. Cell. Biol.* **1993**, *13*, 3340–3349. [[CrossRef](#)]
87. Miller, W.A.; Wang, Z.; Treder, K. The Amazing Diversity of Cap-Independent Translation Elements in the 3'-Untranslated Regions of Plant Viral RNAs. *Biochem. Soc. Trans.* **2007**, *35*, 1629–1633. [[CrossRef](#)]
88. Fabian, M.R.; White, K.A. 5'-3' RNA-RNA Interaction Facilitates Cap- and Poly(A) Tail-Independent Translation of Tomato Bushy Stunt Virus Mrna: A Potential Common Mechanism for Tombusviridae. *J. Biol. Chem.* **2004**, *279*, 28862–28872. [[CrossRef](#)]
89. Pelletier, J.; Sonenberg, N. Internal Initiation of Translation of Eukaryotic mRNA Directed by a Sequence Derived from Poliovirus RNA. *Nature* **1988**, *334*, 320–325. [[CrossRef](#)]
90. Jang, S.K.; Kräusslich, H.G.; Nicklin, M.J.; Duke, G.M.; Palmenberg, A.C.; Wimmer, E. A Segment of the 5' Nontranslated Region of Encephalomyocarditis Virus RNA Directs Internal Entry of Ribosomes during in Vitro Translation. *J. Virol.* **1988**, *62*, 2636–2643. [[CrossRef](#)]
91. Pilipenko, E.V.; Blinov, V.M.; Chernov, B.K.; Dmitrieva, T.M.; Agol, V.I. Conservation of the Secondary Structure Elements of the 5'-Untranslated Region of Cardio- and Aphthovirus RNAs. *Nucleic Acids Res.* **1989**, *17*, 5701–5711. [[CrossRef](#)]
92. Mailliot, J.; Martin, F. Viral Internal Ribosomal Entry Sites: Four Classes for One Goal. *WIREs RNA* **2017**, *9*. [[CrossRef](#)]
93. Evans, D.M.A.; Dunn, G.; Minor, P.D.; Schild, G.C.; Cann, A.J.; Stanway, G.; Almond, J.W.; Currey, K.; Maizel, J.V. Increased Neurovirulence Associated with a Single Nucleotide Change in a Noncoding Region of the Sabin Type 3 Poliovaccine Genome. *Nature* **1985**, *314*, 548–550. [[CrossRef](#)]
94. Harrison, G.P.; Lever, A.M. The Human Immunodeficiency Virus Type 1 Packaging Signal and Major Splice Donor Region Have a Conserved Stable Secondary Structure. *J. Virol.* **1992**, *66*, 4144–4153. [[CrossRef](#)]
95. Baudin, F.; Marquet, R.; Isel, C.; Darlix, J.-L.; Ehresmann, B.; Ehresmann, C. Functional Sites in the 5' Region of Human Immunodeficiency Virus Type 1 RNA Form Defined Structural Domains. *J. Mol. Biol.* **1993**, *229*, 382–397. [[CrossRef](#)]
96. Isel, C.; Ehresmann, C.; Keith, G.; Ehresmann, B.; Marquet, R. Initiation of Reverse Transcription of HIV-1: Secondary Structure of the HIV-1 RNA/TRNA(3Lys) (Template/Primer). *J. Mol. Biol.* **1995**, *247*, 236–250. [[CrossRef](#)]
97. Ennifar, E.; Paillart, J.-C.; Bodlenner, A.; Walter, P.; Weibel, J.-M.; Aubertin, A.-M.; Pale, P.; Dumas, P.; Marquet, R. Targeting the Dimerization Initiation Site of HIV-1 RNA with Aminoglycosides: From Crystal to Cell. *Nucleic Acids Res.* **2006**, *34*, 2328–2339. [[CrossRef](#)]
98. Abbink, T.E.M.; Berkhout, B. A Novel Long Distance Base-Pairing Interaction in Human Immunodeficiency Virus Type 1 RNA Occludes the Gag Start Codon\*. *J. Biol. Chem.* **2003**, *278*, 11601–11611. [[CrossRef](#)]
99. Huthoff, H.; Berkhout, B. Two Alternating Structures of the HIV-1 Leader RNA. *RNA* **2001**, *7*, 143–157. [[CrossRef](#)]
100. Damgaard, C.K.; Andersen, E.S.; Knudsen, B.; Gorodkin, J.; Kjems, J. RNA Interactions in the 5' Region of the HIV-1 Genome. *J. Mol. Biol.* **2004**, *336*, 369–379. [[CrossRef](#)]
101. Symons, R.H. Self-Cleavage of RNA in the Replication of Small Pathogens of Plants and Animals. *Trends Biochem. Sci.* **1989**, *14*, 445–450. [[CrossRef](#)]
102. Kruger, K.; Grabowski, P.J.; Zaug, A.J.; Sands, J.; Gottschling, D.E.; Cech, T.R. Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell* **1982**, *31*, 147–157. [[CrossRef](#)]
103. Wu, H.N.; Lin, Y.J.; Lin, F.P.; Makino, S.; Chang, M.F.; Lai, M.M. Human Hepatitis Delta Virus RNA Subfragments Contain an Autocleavage Activity. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 1831–1835. [[CrossRef](#)] [[PubMed](#)]
104. Strobel, B.; Spöring, M.; Klein, H.; Blazevic, D.; Rust, W.; Sayols, S.; Hartig, J.S.; Kreuz, S. High-Throughput Identification of Synthetic Riboswitches by Barcode-Free Amplicon-Sequencing in Human Cells. *Nat. Commun.* **2020**, *11*, 714. [[CrossRef](#)]
105. Merino, E.J.; Wilkinson, K.A.; Coughlan, J.L.; Weeks, K.M. RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.* **2005**, *127*, 4223–4231. [[CrossRef](#)]
106. Chamberlin, S.I.; Weeks, K.M. Differential Helix Stabilities and Sites Pre-Organized for Tertiary Interactions Revealed by Monitoring Local Nucleotide Flexibility in the B15 Group I Intron RNA. *Biochemistry* **2003**, *42*, 901–909. [[CrossRef](#)]

107. Wilkinson, K.A.; Merino, E.J.; Weeks, K.M. Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE): Quantitative RNA Structure Analysis at Single Nucleotide Resolution. *Nat. Protoc.* **2006**, *1*, 1610–1616. [[CrossRef](#)]
108. Wilkinson, K.A.; Merino, E.J.; Weeks, K.M. RNA SHAPE Chemistry Reveals Nonhierarchical Interactions Dominate Equilibrium Structural Transitions in tRNA(Asp) Transcripts. *J. Am. Chem. Soc.* **2005**, *127*, 4659–4667. [[CrossRef](#)]
109. Mortimer, S.A.; Weeks, K.M. A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry. *J. Am. Chem. Soc.* **2007**, *129*, 4144–4145. [[CrossRef](#)]
110. Gherghe, C.M.; Mortimer, S.A.; Krahn, J.M.; Thompson, N.L.; Weeks, K.M. Slow Conformational Dynamics at C2'-Endo Nucleotides in RNA. *J. Am. Chem. Soc.* **2008**, *130*, 8884–8885. [[CrossRef](#)]
111. Mortimer, S.A.; Weeks, K.M. Time-Resolved RNA SHAPE Chemistry. *J. Am. Chem. Soc.* **2008**, *130*, 16178–16180. [[CrossRef](#)]
112. Spitale, R.C.; Crisalli, P.; Flynn, R.A.; Torre, E.A.; Kool, E.T.; Chang, H.Y. RNA SHAPE Analysis in Living Cells. *Nat. Chem. Biol.* **2013**, *9*, 18–20. [[CrossRef](#)]
113. Tyrrell, J.; McGinnis, J.L.; Weeks, K.M.; Pielak, G.J. The Cellular Environment Stabilizes Adenine Riboswitch RNA Structure. *Biochemistry* **2013**, *52*, 8777–8785. [[CrossRef](#)]
114. Vasa, S.M.; Guex, N.; Wilkinson, K.A.; Weeks, K.M.; Giddings, M.C. ShapeFinder: A Software System for High-Throughput Quantitative Analysis of Nucleic Acid Reactivity Information Resolved by Capillary Electrophoresis. *RNA* **2008**, *14*, 1979–1990. [[CrossRef](#)]
115. Mitra, S.; Shcherbakova, I.V.; Altman, R.B.; Brenowitz, M.; Laederach, A. High-Throughput Single-Nucleotide Structural Mapping by Capillary Automated Footprinting Analysis. *Nucleic Acids Res.* **2008**, *36*, e63. [[CrossRef](#)]
116. Yoon, S.; Kim, J.; Hum, J.; Kim, H.; Park, S.; Kladwang, W.; Das, R. HiTRACE: High-Throughput Robust Analysis for Capillary Electrophoresis. *Bioinformatics* **2011**, *27*, 1798–1805. [[CrossRef](#)]
117. Pang, P.S.; Elazar, M.; Pham, E.A.; Glenn, J.S. Simplified RNA Secondary Structure Mapping by Automation of SHAPE Data Analysis. *Nucleic Acids Res.* **2011**, *39*, e151. [[CrossRef](#)]
118. Karabiber, F.; McGinnis, J.L.; Favorov, O.V.; Weeks, K.M. QuShape: Rapid, Accurate, and Best-Practices Quantification of Nucleic Acid Probing Information, Resolved by Capillary Electrophoresis. *RNA* **2013**, *19*, 63–73. [[CrossRef](#)]
119. Cantara, W.A.; Hatterschide, J.; Wu, W.; Musier-Forsyth, K. RiboCAT: A New Capillary Electrophoresis Data Analysis Tool for Nucleic Acid Probing. *RNA* **2017**, *23*, 240–249. [[CrossRef](#)]
120. Gumna, J.; Zok, T.; Figurski, K.; Pachulska-Wieczorek, K.; Szachniuk, M. RNAtor—Fast, Accurate Normalization, Visualization and Statistical Analysis of RNA Probing Data Resolved by Capillary Electrophoresis. *PLoS ONE* **2020**, *15*, e0239287. [[CrossRef](#)]
121. Wirecki, T.K.; Merdas, K.; Bernat, A.; Boniecki, M.J.; Bujnicki, J.M.; Stefaniak, F. RNAProbe: A Web Server for Normalization and Analysis of RNA Structure Probing Data. *Nucleic Acids Res.* **2020**, *48*, W292–W299. [[CrossRef](#)]
122. Poulsen, L.D.; Kielpinski, L.J.; Salama, S.R.; Krogh, A.; Vinther, J. SHAPE Selection (SHAPES) Enrich for RNA Structure Signal in SHAPE Sequencing-Based Probing Data. *RNA* **2015**, *21*, 1042–1052. [[CrossRef](#)]
123. Spitale, R.C.; Flynn, R.A.; Zhang, Q.C.; Crisalli, P.; Lee, B.; Jung, J.-W.; Kuchelmeister, H.Y.; Batista, P.J.; Torre, E.A.; Kool, E.T.; et al. Structural Imprints in Vivo Decode RNA Regulatory Mechanisms. *Nature* **2015**, *519*, 486–490. [[CrossRef](#)]
124. Wang, P.Y.; Sexton, A.N.; Culligan, W.J.; Simon, M.D. Carbodiimide Reagents for the Chemical Probing of RNA Structure in Cells. *RNA* **2019**, *25*, 135–146. [[CrossRef](#)] [[PubMed](#)]
125. Feng, C.; Chan, D.; Joseph, J.; Muuronen, M.; Coldren, W.H.; Dai, N.; Corrêa, I.R.; Furche, F.; Hadad, C.M.; Spitale, R.C. Light-Activated Chemical Probing of Nucleobase Solvent Accessibility Inside Cells. *Nat. Chem. Biol.* **2018**, *14*, 276–283. [[CrossRef](#)]
126. Watts, J.M.; Dang, K.K.; Gorelick, R.J.; Leonard, C.W.; Bess, J.W., Jr.; Swanstrom, R.; Burch, C.L.; Weeks, K.M. Architecture and Secondary Structure of an Entire HIV-1 RNA Genome. *Nature* **2009**, *460*, 711–716. [[CrossRef](#)] [[PubMed](#)]
127. Burrill, C.P.; Westesson, O.; Schulte, M.B.; Strings, V.R.; Segal, M.; Andino, R. Global RNA Structure Analysis of Poliovirus Identifies a Conserved RNA Structure Involved in Viral Replication and Infectivity. *J. Virol.* **2013**, *87*, 11670–11683. [[CrossRef](#)]
128. Pirakitikulr, N.; Kohlway, A.; Lindenbach, B.D.; Pyle, A.M. The Coding Region of the HCV Genome Contains a Network of Regulatory RNA Structures. *Mol. Cell* **2016**, *62*, 111–120. [[CrossRef](#)]
129. Cordero, P.; Kladwang, W.; VanLang, C.C.; Das, R. Quantitative DMS Mapping for Automated RNA Secondary Structure Inference. *Biochemistry* **2012**, *51*, 7037–7039. [[CrossRef](#)]
130. Maurin, T.; Melko, M.; Abekhoukh, S.; Khalfallah, O.; Davidovic, L.; Jarjat, M.; D'Antoni, S.; Catania, M.V.; Moine, H.; Bechara, E.; et al. The FMRP/GRK4 mRNA Interaction Uncovers a New Mode of Binding of the Fragile X Mental Retardation Protein in Cerebellum. *Nucleic Acids Res.* **2015**, *43*, 8540–8550. [[CrossRef](#)] [[PubMed](#)]
131. Alghoul, F.; Eriani, G.; Martin, F. RNA Secondary Structure Study by Chemical Probing Methods Using DMS and CMCT. In *Small Non-Coding RNAs: Methods and Protocols*; Methods in Molecular Biology; Rederstorff, M., Ed.; Springer: New York, NY, USA, 2021; pp. 241–250. ISBN 978-1-07-161386-3.
132. Kertesz, M.; Wan, Y.; Mazor, E.; Rinn, J.L.; Nutter, R.C.; Chang, H.Y.; Segal, E. Genome-Wide Measurement of RNA Secondary Structure in Yeast. *Nature* **2010**, *467*, 103–107. [[CrossRef](#)]
133. Underwood, J.G.; Uzilov, A.V.; Katzman, S.; Onodera, C.S.; Mainzer, J.E.; Mathews, D.H.; Lowe, T.M.; Salama, S.R.; Hausler, D. FragSeq: Transcriptome-Wide RNA Structure Probing Using High-Throughput Sequencing. *Nat. Methods* **2010**, *7*, 995–1001. [[CrossRef](#)]
134. Wan, Y.; Qu, K.; Ouyang, Z.; Chang, H.Y. Genome-Wide Mapping of RNA Structure Using Nuclease Digestion and High-Throughput Sequencing. *Nat. Protoc.* **2013**, *8*, 849–869. [[CrossRef](#)]

135. Wan, Y.; Qu, K.; Ouyang, Z.; Kertesz, M.; Li, J.; Tibshirani, R.; Makino, D.L.; Nutter, R.C.; Segal, E.; Chang, H.Y. Genome-Wide Measurement of RNA Folding Energies. *Mol. Cell* **2012**, *48*, 169–181. [[CrossRef](#)]
136. Saus, E.; Willis, J.R.; Prysycz, L.P.; Hafez, A.; Llorens, C.; Himmelbauer, H.; Gabaldón, T. NextPARS: Parallel Probing of RNA Structures in Illumina. *RNA* **2018**, *24*, 609–619. [[CrossRef](#)]
137. Lucks, J.B.; Mortimer, S.A.; Trapnell, C.; Luo, S.; Aviran, S.; Schroth, G.P.; Pachter, L.; Doudna, J.A.; Arkin, A.P. Multiplexed RNA Structure Characterization with Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension Sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11063–11068. [[CrossRef](#)]
138. Loughrey, D.; Watters, K.E.; Settle, A.H.; Lucks, J.B. SHAPE-Seq 2.0: Systematic Optimization and Extension of High-Throughput Chemical Probing of RNA Secondary Structure with next Generation Sequencing. *Nucleic Acids Res.* **2014**, *42*, e165. [[CrossRef](#)]
139. Li, P.; Zhou, X.; Xu, K.; Zhang, Q.C. RASP: An Atlas of Transcriptome-Wide RNA Secondary Structure Probing Data. *Nucleic Acids Res.* **2021**, *49*, D183–D191. [[CrossRef](#)]
140. Kielpinski, L.J.; Vinther, J. Massive Parallel-Sequencing-Based Hydroxyl Radical Probing of RNA Accessibility. *Nucleic Acids Res.* **2014**, *42*, e70. [[CrossRef](#)]
141. Ding, Y.; Tang, Y.; Kwok, C.K.; Zhang, Y.; Bevilacqua, P.C.; Assmann, S.M. In Vivo Genome-Wide Profiling of RNA Secondary Structure Reveals Novel Regulatory Features. *Nature* **2014**, *505*, 696–700. [[CrossRef](#)]
142. Rouskin, S.; Zubradt, M.; Washietl, S.; Kellis, M.; Weissman, J.S. Genome-Wide Probing of RNA Structure Reveals Active Unfolding of mRNA Structures in Vivo. *Nature* **2014**, *505*, 701–705. [[CrossRef](#)] [[PubMed](#)]
143. Talkish, J.; May, G.; Lin, Y.; Woolford, J.L.; McManus, C.J. Mod-Seq: High-Throughput Sequencing for Chemical Probing of RNA Structure. *RNA* **2014**, *20*, 713–720. [[CrossRef](#)] [[PubMed](#)]
144. Incarnato, D.; Neri, F.; Anselmi, F.; Oliviero, S. Genome-Wide Profiling of Mouse RNA Secondary Structures Reveals Key Features of the Mammalian Transcriptome. *Genome Biol.* **2014**, *15*, 491. [[CrossRef](#)] [[PubMed](#)]
145. Zhou, K.I.; Clark, W.C.; Pan, D.W.; Eckwahl, M.J.; Dai, Q.; Pan, T. Pseudouridines Have Context-Dependent Mutation and Stop Rates in High-Throughput Sequencing. *RNA Biol.* **2018**, *15*, 892–900. [[CrossRef](#)]
146. Mitchell, D.; Renda, A.J.; Douds, C.A.; Babitzke, P.; Assmann, S.M.; Bevilacqua, P.C. In Vivo RNA Structural Probing of Uracil and Guanine Base-Pairing by 1-Ethyl-3-(3-Dimethylaminopropyl)Carbodiimide (EDC). *RNA* **2019**, *25*, 147–157. [[CrossRef](#)]
147. Twittenhoff, C.; Brandenburg, V.B.; Righetti, F.; Nuss, A.M.; Mosig, A.; Dersch, P.; Narberhaus, F. Lead-Seq: Transcriptome-Wide Structure Probing in Vivo Using Lead(II) Ions. *Nucleic Acids Res.* **2020**, *48*, e71. [[CrossRef](#)]
148. Métifiot, M.; Amrane, S.; Litvak, S.; Andreola, M.-L. G-Quadruplexes in Viruses: Function and Potential Therapeutic Applications. *Nucleic Acids Res.* **2014**, *42*, 12352–12366. [[CrossRef](#)]
149. Lavezzo, E.; Berselli, M.; Frasson, I.; Perrone, R.; Palù, G.; Brazzale, A.R.; Richter, S.N.; Toppo, S. G-Quadruplex Forming Sequences in the Genome of All Known Human Viruses: A Comprehensive Guide. *PLoS Comput. Biol.* **2018**, *14*, e1006675. [[CrossRef](#)]
150. Lyu, K.; Chow, E.Y.-C.; Mou, X.; Chan, T.-F.; Kwok, C.K. RNA G-Quadruplexes (RG4s): Genomics and Biological Functions. *Nucleic Acids Res.* **2021**, *49*, 5426–5450. [[CrossRef](#)]
151. Kwok, C.K.; Balasubramanian, S. Targeted Detection of G-Quadruplexes in Cellular RNAs. *Angew Chem. Int. Ed.* **2015**, *54*, 6751–6754. [[CrossRef](#)]
152. Kwok, C.K.; Marsico, G.; Sahakyan, A.B.; Chambers, V.S.; Balasubramanian, S. RG4-Seq Reveals Widespread Formation of G-Quadruplex Structures in the Human Transcriptome. *Nat. Methods* **2016**, *13*, 841–844. [[CrossRef](#)]
153. Yang, X.; Cheema, J.; Zhang, Y.; Deng, H.; Duncan, S.; Umar, M.I.; Zhao, J.; Liu, Q.; Cao, X.; Kwok, C.K.; et al. RNA G-Quadruplex Structures Exist and Function in Vivo in Plants. *Genome Biol.* **2020**, *21*, 226. [[CrossRef](#)]
154. Yang, S.Y.; Lejault, P.; Chevrier, S.; Boidot, R.; Robertson, A.G.; Wong, J.M.Y.; Monchaud, D. Transcriptome-Wide Identification of Transient RNA G-Quadruplexes in Human Cells. *Nat. Commun.* **2018**, *9*, 4730. [[CrossRef](#)]
155. Weng, X.; Gong, J.; Chen, Y.; Wu, T.; Wang, F.; Yang, S.; Yuan, Y.; Luo, G.; Chen, K.; Hu, L.; et al. Keth-Seq for Transcriptome-Wide RNA Structure Mapping. *Nat. Chem. Biol.* **2020**, *16*, 489–492. [[CrossRef](#)]
156. Sun, L.; Li, P.; Ju, X.; Rao, J.; Huang, W.; Ren, L.; Zhang, S.; Xiong, T.; Xu, K.; Zhou, X.; et al. In Vivo Structural Characterization of the SARS-CoV-2 RNA Genome Identifies Host Proteins Vulnerable to Repurposed Drugs. *Cell* **2021**, *184*, 1865–1883.e20. [[CrossRef](#)]
157. Siegfried, N.A.; Busan, S.; Rice, G.M.; Nelson, J.A.E.; Weeks, K.M. RNA Motif Discovery by SHAPE and Mutational Profiling (SHAPE-MaP). *Nat. Methods* **2014**, *11*, 959–965. [[CrossRef](#)]
158. Smola, M.J.; Rice, G.M.; Busan, S.; Siegfried, N.A.; Weeks, K.M. Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension and Mutational Profiling (SHAPE-MaP) for Direct, Versatile and Accurate RNA Structure Analysis. *Nat. Protoc.* **2015**, *10*, 1643–1669. [[CrossRef](#)]
159. Zubradt, M.; Gupta, P.; Persad, S.; Lambowitz, A.M.; Weissman, J.S.; Rouskin, S. DMS-MaPseq for Genome-Wide or Targeted RNA Structure Probing in Vivo. *Nat. Methods* **2017**, *14*, 75–82. [[CrossRef](#)]
160. Lister, R.; Pelizzola, M.; Downen, R.H.; Hawkins, R.D.; Hon, G.; Tonti-Filippini, J.; Nery, J.R.; Lee, L.; Ye, Z.; Ngo, Q.-M.; et al. Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences. *Nature* **2009**, *462*, 315–322. [[CrossRef](#)]
161. Zinshteyn, B.; Chan, D.; England, W.; Feng, C.; Green, R.; Spitale, R.C. Assaying RNA Structure with LASER-Seq. *Nucleic Acids Res.* **2019**, *47*, 43–55. [[CrossRef](#)]
162. Marinus, T.; Fessler, A.B.; Ogle, C.A.; Incarnato, D. A Novel SHAPE Reagent Enables the Analysis of RNA Structure in Living Cells with Unprecedented Accuracy. *Nucleic Acids Res.* **2021**. [[CrossRef](#)] [[PubMed](#)]

163. Homan, P.J.; Favorov, O.V.; Lavender, C.A.; Kursun, O.; Ge, X.; Busan, S.; Dokholyan, N.V.; Weeks, K.M. Single-Molecule Correlated Chemical Probing of RNA. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13858–13863. [[CrossRef](#)] [[PubMed](#)]
164. Krokhotin, A.; Mustoe, A.M.; Weeks, K.M.; Dokholyan, N.V. Direct Identification of Base-Paired RNA Nucleotides by Correlated Chemical Probing. *RNA* **2017**, *23*, 6–13. [[CrossRef](#)] [[PubMed](#)]
165. Mustoe, A.M.; Lama, N.N.; Irving, P.S.; Olson, S.W.; Weeks, K.M. RNA Base-Pairing Complexity in Living Cells Visualized by Correlated Chemical Probing. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 24574–24582. [[CrossRef](#)]
166. Dethoff, E.A.; Boerneke, M.A.; Gokhale, N.S.; Muhire, B.M.; Martin, D.P.; Sacco, M.T.; McFadden, M.J.; Weinstein, J.B.; Messer, W.B.; Horner, S.M.; et al. Pervasive Tertiary Structure in the Dengue Virus RNA Genome. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 11513–11518. [[CrossRef](#)]
167. Sengupta, A.; Rice, G.M.; Weeks, K.M. Single-Molecule Correlated Chemical Probing Reveals Large-Scale Structural Communication in the Ribosome and the Mechanism of the Antibiotic Spectinomycin in Living Cells. *PLoS Biol.* **2019**, *17*, e3000393. [[CrossRef](#)]
168. Ehrhardt, J.E.; Weeks, K.M. Time-Resolved, Single-Molecule, Correlated Chemical Probing of RNA. *J. Am. Chem. Soc.* **2020**, *142*, 18735–18740. [[CrossRef](#)]
169. Tomezsko, P.J.; Corbin, V.D.A.; Gupta, P.; Swaminathan, H.; Glasgow, M.; Persad, S.; Edwards, M.D.; Mcintosh, L.; Papenfuss, A.T.; Emery, A.; et al. Determination of RNA Structural Diversity and Its Role in HIV-1 RNA Splicing. *Nature* **2020**, *582*, 438–442. [[CrossRef](#)]
170. Kladwang, W.; Cordero, P.; Das, R. A Mutate-and-Map Strategy Accurately Infers the Base Pairs of a 35-Nucleotide Model RNA. *RNA* **2011**, *17*, 522–534. [[CrossRef](#)]
171. Kladwang, W.; VanLang, C.C.; Cordero, P.; Das, R. A Two-Dimensional Mutate-and-Map Strategy for Non-Coding RNA Structure. *Nat. Chem.* **2011**, *3*, 954–962. [[CrossRef](#)]
172. Cheng, C.Y.; Kladwang, W.; Yesselman, J.D.; Das, R. RNA Structure Inference through Chemical Mapping after Accidental or Intentional Mutations. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 9876–9881. [[CrossRef](#)]
173. Weidmann, C.A.; Mustoe, A.M.; Jariwala, P.B.; Calabrese, J.M.; Weeks, K.M. Analysis of RNA-Protein Networks with RNP-MaP Defines Functional Hubs on RNA. *Nat. Biotechnol.* **2021**, *39*, 347–356. [[CrossRef](#)]
174. Dadonaite, B.; Gilbertson, B.; Knight, M.L.; Trifkovic, S.; Rockman, S.; Laederach, A.; Brown, L.E.; Fodor, E.; Bauer, D.L.V. The Structure of the Influenza A Virus Genome. *Nat. Microbiol.* **2019**, *4*, 1781–1789. [[CrossRef](#)]
175. Simon, L.M.; Morandi, E.; Luganini, A.; Gribaudo, G.; Martinez-Sobrido, L.; Turner, D.H.; Oliviero, S.; Incarnato, D. In Vivo Analysis of Influenza A mRNA Secondary Structures Identifies Critical Regulatory Motifs. *Nucleic Acids Res.* **2019**, *47*, 7003–7017. [[CrossRef](#)]
176. Madden, E.A.; Plante, K.S.; Morrison, C.R.; Kutchko, K.M.; Sanders, W.; Long, K.M.; Taft-Benz, S.; Cruz Cisneros, M.C.; White, A.M.; Sarkar, S.; et al. Using SHAPE-MaP To Model RNA Secondary Structure and Identify 3'UTR Variation in Chikungunya Virus. *J. Virol.* **2020**, *94*, e00701-20. [[CrossRef](#)]
177. Manfredonia, I.; Nithin, C.; Ponce-Salvatierra, A.; Ghosh, P.; Wirecki, T.K.; Marinus, T.; Ogando, N.S.; Snijder, E.J.; van Hemert, M.J.; Bujnicki, J.M.; et al. Genome-Wide Mapping of SARS-CoV-2 RNA Structures Identifies Therapeutically-Relevant Elements. *Nucleic Acids Res.* **2020**, *48*, 12436–12452. [[CrossRef](#)]
178. Huston, N.C.; Wan, H.; Strine, M.S.; de Cesaris Araujo Tavares, R.; Wilen, C.B.; Pyle, A.M. Comprehensive in Vivo Secondary Structure of the SARS-CoV-2 Genome Reveals Novel Regulatory Motifs and Mechanisms. *Mol. Cell* **2021**, *81*, 584–598.e5. [[CrossRef](#)]
179. Lan, T.C.; Allan, M.F.; Malsick, L.; Khandwala, S.; Nyeo, S.S.; Sun, Y.; Guo, J.U.; Bathe, M.; Griffiths, A.; Rouskin, S. Insights into the Secondary Structural Ensembles of the Full SARS-CoV-2 RNA Genome in Infected Cells. *Biorxiv* **2021**. [[CrossRef](#)]
180. Zhao, J.; Qiu, J.; Aryal, S.; Hackett, J.L.; Wang, J. The RNA Architecture of the SARS-CoV-2 3'-Untranslated Region. *Viruses* **2020**, *12*, 1473. [[CrossRef](#)]
181. Christy, T.W.; Giannetti, C.A.; Houlihan, G.; Smola, M.J.; Rice, G.M.; Wang, J.; Dokholyan, N.V.; Laederach, A.; Holliger, P.; Weeks, K.M. Direct Mapping of Higher-Order RNA Interactions by SHAPE-JuMP. *Biochemistry* **2021**, *60*, 1971–1982. [[CrossRef](#)]
182. Sharma, E.; Sterne-Weiler, T.; O'Hanlon, D.; Blencowe, B.J. Global Mapping of Human RNA-RNA Interactions. *Mol. Cell* **2016**, *62*, 618–626. [[CrossRef](#)] [[PubMed](#)]
183. Lu, Z.; Gong, J.; Zhang, Q.C. PARIS: Psoralen Analysis of RNA Interactions and Structures with High Throughput and Resolution. *Methods Mol. Biol.* **2018**, *1649*, 59–84. [[CrossRef](#)] [[PubMed](#)]
184. Ziv, O.; Gabryelska, M.M.; Lun, A.T.L.; Gebert, L.F.R.; Sheu-Gruttadauria, J.; Meredith, L.W.; Liu, Z.-Y.; Kwok, C.K.; Qin, C.-F.; MacRae, I.J.; et al. COMRADES Determines in Vivo RNA Structures and Interactions. *Nat. Methods* **2018**, *15*, 785–788. [[CrossRef](#)]
185. Cheng, C.Y.; Chou, F.-C.; Kladwang, W.; Tian, S.; Cordero, P.; Das, R. Consistent Global Structures of Complex RNA States through Multidimensional Chemical Mapping. *eLife* **2015**, *4*, e07600. [[CrossRef](#)]
186. Aw, J.G.A.; Lim, S.W.; Wang, J.X.; Lambert, F.R.P.; Tan, W.T.; Shen, Y.; Zhang, Y.; Kaewsapsak, P.; Li, C.; Ng, S.B.; et al. Determination of Isoform-Specific RNA Structure with Nanopore Long Reads. *Nat. Biotechnol.* **2021**, *39*, 336–346. [[CrossRef](#)] [[PubMed](#)]
187. Cubi, R.; Bouhedda, F.; Collot, M.; Klymchenko, A.S.; Ryckelynck, M. MIVC-Useq: A Microfluidic-Assisted High-Throughput Functional Screening in Tandem with next-Generation Sequencing and Artificial Neural Network to Rapidly Characterize RNA Molecules. *RNA* **2021**, *27*, 841–853. [[CrossRef](#)]

### IV/ Evolution des modèles structuraux rapportés dans la littérature

Dans la partie précédente, nous avons pu voir différentes approches permettant l'étude de la structure des ARN ainsi que leurs évolutions qui ont été suscitées notamment par l'intérêt grandissant pour ces structures depuis le milieu du 20<sup>ème</sup> siècle. Ainsi, au moment de la découverte du VIH-1 en 1983, différents modèles rétroviraux avaient déjà été analysés et avaient permis de mettre en évidence la présence de motifs structuraux qui se sont avérés importants pour les cycles répliatifs viraux. En effet, les séquences répétées présentes aux extrémités 5' et 3', la structure fixant l'ARNt cellulaire nécessaire à la transcription inverse, la région permettant la dimérisation et celle permettant l'encapsidation avaient déjà été identifiées chez différents rétrovirus, comme le RSV ou le Mo-MuLV, sans pour autant avoir été caractérisées (141–144).

Il n'est donc pas étonnant que les premières recherches structurales impliquant l'ARNg du VIH-1 se soient intéressées aux domaines homologues déjà étudiés chez les autres rétrovirus comme les transcrits des LTR, permettant l'identification et la caractérisation des motifs TAR (145) et PolyA (146), puis le motif d'export cytoplasmique RRE (46,147,148) ou encore les régions PBS interagissant avec l'ARNt pour initier la transcription inverse (149). Cependant, l'identification d'autres régions comme le site de dimérisation (*Dimer Linkage Sequence* ou DLS) ou le signal d'encapsidation (*Packaging signal* ou Psi) ont nécessité plus de temps (150,151). Après l'analyse et la caractérisation de chacun de ces domaines séparément, l'évolution des méthodes structurales a permis d'étudier des domaines de plus en plus larges, culminant avec l'analyse structurale de la totalité des 9,2 kb de l'ARNg du VIH-1 (152).

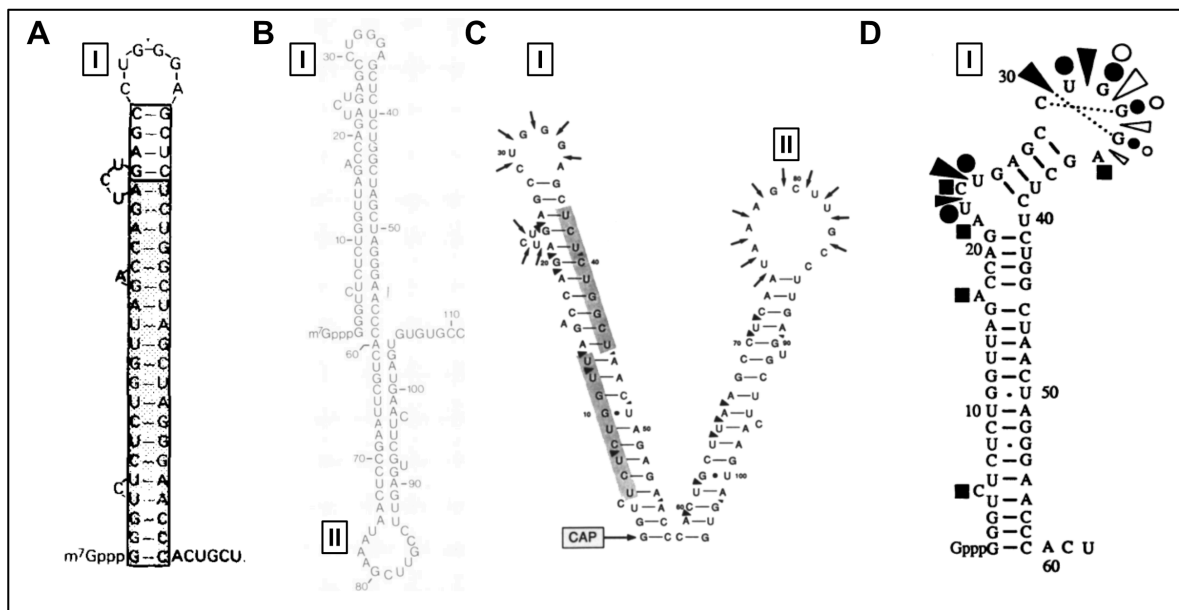
Finalement, l'objectif de la biologie structurale autour de l'ARNg du VIH-1 a changé de paradigme au cours des années. Alors que les premières études s'intéressaient à identifier quelles structures jouaient un rôle dans le cycle viral, il s'est avéré de plus en plus important de comprendre comment cette fonction était menée. Cette transition s'est opérée avec la mise en évidence des capacités de différents partenaires cellulaires ou viraux à affecter les repliements adoptés par l'ARNg et les mécanismes impliqués dans le cycle.

## INTRODUCTION

### IV.1/ Les motifs de l'ARNg du VIH-1

#### IV.1.1/ La tige-boucle TAR

En 1986, une des premières structures prédites dans le génome du VIH-1 correspond à la structure en tige-boucle du *trans-activation response* (TAR) *element* (145). Ce modèle (**Figure 9A**) a été déterminé en suivant uniquement les règles établies par Tinoco *et al.*, (1973) qui se basent sur l'énergie libre de chaque doublet de paires de bases, plutôt que celle des paires de bases prises indépendamment de leurs voisines, et qui considère les boucles internes comme étant plus stable que les boucles apicales (153). Ainsi, une structure présentant une boucle apicale de motif CUGGGA, une boucle interne de 3 pyrimidines et l'exclusion des nucléotides C5 et A17 de l'hélice a été décrite. Des expériences postérieures ont montré que la destruction de cette structure par des délétions inactivait la fonction portée par le motif TAR, permettant ainsi de relier l'étape de trans-activation à la présence de ce domaine structural en tige-boucle dans la région répétée (R) du 5'-UTR (154).



**Figure 9** : Premiers modèles structuraux des tiges-boucles TAR et PolyA de l'ARNg du VIH-1. Ces modèles ont été prédits par (A) Okamoto et Wong-Stall (1986) et (B) Parkin *et al.*, (1988) ou (C) obtenus par cartographie enzymatique par Muesing *et al.*, (1987) ou (D) cartographie chimique et enzymatique par Colvin et Garcia-Blanco, (1992) (145,146,155,156). La structure globale de la tige-boucle TAR (I) fait consensus dans les différents modèles rapportés dans la littérature. Dans le cadre de la structure de la tige-boucle PolyA (II), cette dernière est plus rarement retrouvée, mais en général sous la forme présentée ici.



## INTRODUCTION

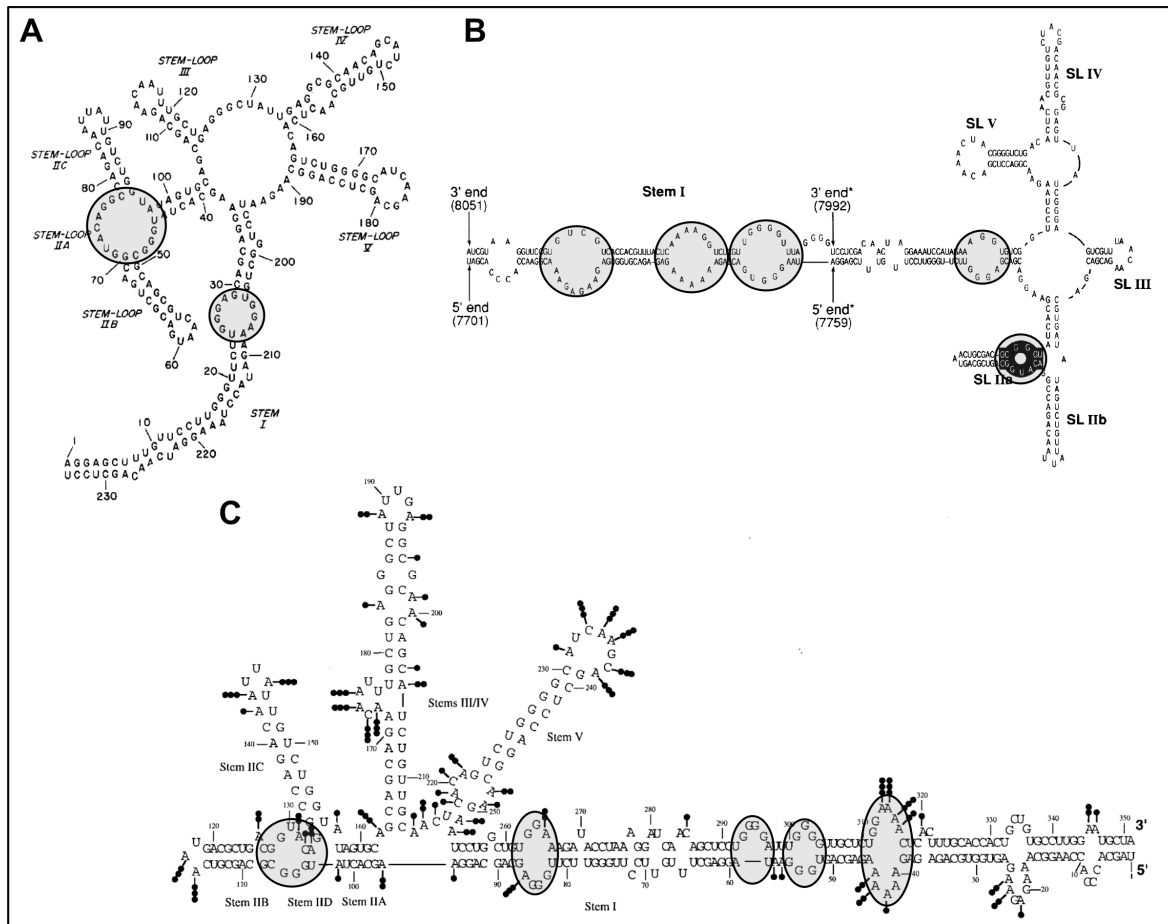
Une seconde étude, s'intéressant aux effets du domaine 5'-UTR dans la traduction (146), avait également modélisé les motifs TAR et PolyA (**Figure 9B**) en se servant des mêmes règles de Tinoco *et al.*, (1973) (153). De façon intéressante, la partie TAR du modèle structural prédit n'est pas affectée par l'ajout des 55 nucléotides de la séquence du motif PolyA. Cette prédiction, mise en parallèle d'une étude de cartographie enzymatique (**Figure 9C**) présentait déjà un modèle structural valide pour la région 5'-UTR du VIH-1 (155). Par la suite, de nombreuses expériences de cartographie chimique et enzymatique (157,158), de mutagenèse (159) et de RMN (160–162) ont permis de confirmer et d'affiner la structure de la tige-boucle TAR (156) mais aussi d'améliorer la compréhension du mécanisme de reconnaissance avec la protéine Tat (78) (**Figure 9D**).

### IV.1.2/ Le domaine RRE

Dans les années suivant l'identification de TAR, la région impliquée dans le transport cytoplasmique des ARN viraux partiellement ou non épissés a été découverte dans l'ARNg du HTLV-I (148) avant d'être également identifiée dans l'ARNg du VIH-1 (46). La structure de cette région, reconnue par la protéine Rev et nommée *Rev-response element* (RRE), a été prédite en suivant les règles décrites par Salser (1978) et Cech *et al.*, (1983) (163–165) (**Figure 10A**). Ces règles sont basées sur celles définies par Tinoco *et al.*, (1973) et Borer *et al.*, (1974) avec les valeurs d'énergies arrondies à deux chiffres significatifs et des valeurs énergétiques différentes pour les boucles (153,166). Ce modèle présente une structure composée de 5 tiges-boucles (nommées « *Stem-loop* »), arrangées autour d'une jonction à 5 voies (« *5-ways junction* ») et de courts domaines riches en purine dans les stems I et IIB (**Figure 10A**). Cette conformation en « 5 tiges-boucles » ou 5L a également été prédite par des expériences de cartographie enzymatique et avec un ARN possédant des extrémités 5' et 3' plus longues (167) même si la jonction à 5 voies n'est pas retrouvée (**Figure 10B**). Des études ultérieures ont rapporté que des interactions non Watson-Crick étaient impliquées dans la fixation des protéines Rev sur ces régions riches en purines (168,169).

Il faut attendre 1994 pour qu'une structure alternative au RRE 5L soit obtenue par cartographie enzymatique (47). Celle-ci présente un RRE avec les domaines stem III et IV fusionnés en une tige-boucle III/IV, mais également une tige I allongée avec de nombreuses boucles internes riches en purine pour l'ancrage de protéines Rev (170) (**Figure 10C**).

## INTRODUCTION



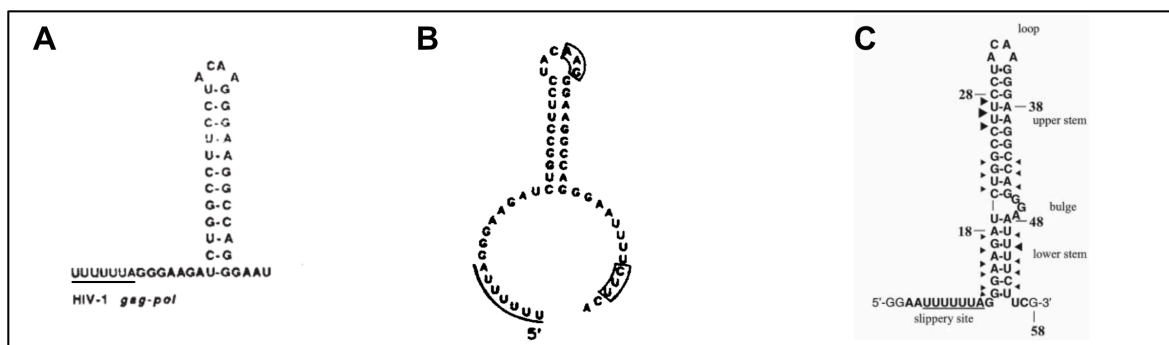
**Figure 10** : Premiers modèles structuraux de la région (RRE) présente dans le gène *env*. (A) La première représentation montre la région RRE prédite par Malim *et al.*, (1990) sous la forme de 5-ways junction et confirmée expérimentalement (Holland *et al.*, 1990) (165,171). (B) Le modèle prédit de Kimura et Ohya, (1994) est similaire à (A) avec la présentation de la structure en 4-ways junction et d'une tige SL1 allongée (167). Elle montre également le site de fixation primaire de Rev en noir où des interactions non canoniques purines-purines ont lieu (168). (C) Le dernier modèle, rapporté par Charpentier *et al.*, (1997) et obtenu par des expériences de cartographie chimique, montre une structure en forme de 4-ways junction où les tiges-boucles SL3 et SL4 sont fusionnées (170). Les boucles internes entourées correspondent aux régions riches en purines.

### IV.1.3/ Le domaine de « Frameshift »

Une troisième spécificité des rétrovirus réside dans l'expression des gènes *gag* et *pol*, régulée par un mécanisme de décalage du cadre de lecture ou « frameshift ». Pour la traduction de la protéine Pr160<sup>Gag-Pol</sup> du VIH-1, ce mécanisme repose sur une séquence « glissante » heptanucléotidique ainsi qu'une structure en tige-boucle. La région impliquée dans ce frameshift a été identifiée et sa structure prédite dès 1988 (172) (**Figure 11A**). Ce

## INTRODUCTION

modèle présente le motif glissant UUUUUUA permettant le décalage d'un nucléotide en amont (-1) sur le cadre de lecture ainsi qu'une tige-boucle avec un motif apical ACCA. Cette même tige boucle a été prédite par une autre approche ainsi qu'un modèle légèrement différent présentant un pseudo-nœud et une boucle apicale plus grande (173,174) (**Figure 11B**). Cependant, ces structures ont été prédites sans support expérimental, ni prise en compte des séquences flanquant la région étudiée. Il faut donc attendre 2002 et les travaux de Dulude *et al.*, (2002) pour obtenir des données expérimentales issues de cartographie enzymatique qui permettent de corriger les modèles prédits (175) (**Figure 11C**). Ces résultats confirment la présence de la tige-boucle prédite par Jacks *et al.*, 1988 et Parkin *et al.*, 1992 mais ils affinent un peu plus le modèle structural du motif et suggèrent que la structure exposerait une boucle interne de trois purines impliquées dans le mécanisme de frameshifting (172,173).



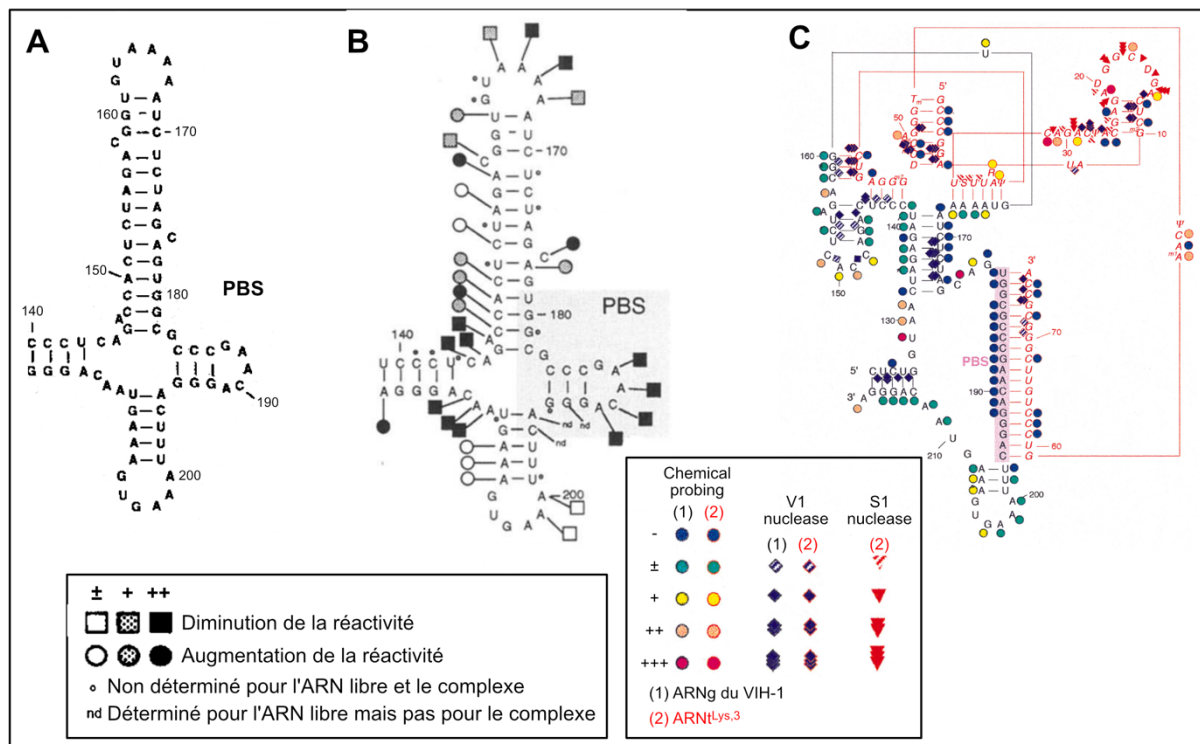
**Figure 11** : Premiers modèles structuraux pour la région du frameshift de Gag-Pol. Les deux premières structures sont issues de prédictions réalisées par (A) Jacks *et al.*, (1988) et (B) Le *et al.*, (1991) (172,174). (B) Le modèle diffère par la présence d'un pseudo-nœud, encadré sur la structure. (C) La structure de Dulude *et al.*, (2002) est la seule à reposer sur des données de cartographie enzymatique à la RNase V1 où chaque flèche représente l'intensité de la coupure par l'enzyme (175). La séquence soulignée correspond au motif glissant permettant le frameshift -1.

### IV.1.4/ Le domaine PBS

L'initiation de la transcription inverse qui conduit à la synthèse du premier brin d'ADNc viral est une étape clé dans le cycle répliatif des rétrovirus. Dans le cas du VIH-1, l'ARN<sup>t<sub>Lys,3</sub></sup> cellulaire a été identifié comme étant l'amorce à cette synthèse par la transcriptase inverse (RT) et son site de fixation à l'ARNg est nommé « *Primer Binding Site* » (PBS) (149). Plusieurs études se sont intéressées à la structure adoptée par la région de l'ARNg qui sert de matrice à l'ARN<sup>t<sub>Lys,3</sub></sup>, en présence ou en absence de ce dernier et de l'enzyme de RT. Le premier modèle

## INTRODUCTION

du PBS a été obtenu par des approches de cartographie chimique (**Figure 12A**) alors que l'entièreté de la région 5' de l'isolat MAL était analysée (158). Il est important de noter que cet isolat présente une insertion de 19 nucléotides en aval du PBS, en faisant un cas particulier. Ce modèle structural présente un domaine PBS de 78 nts avec une jonction à 4 voies (*4 ways junction*). L'effet de la fixation de l'ARNt<sup>Lys,3</sup> a ensuite été étudié par cartographie chimique et enzymatique (176,177). Ces expériences ont permis de mettre en évidence les différentes zones d'interactions entre le domaine PBS de l'ARNg et l'ARNt, mais également les nombreux changements induits lors de la formation du complexe binaire PBS/ARNt (177) (**Figure 12B**). Ainsi, la boucle apicale riche en A présente dans le domaine PBS a été identifiée comme un site d'interaction avec la boucle anticodon de l'ARNt<sup>Lys,3</sup> (176). De plus, d'autres interactions impliquées dans le maintien du complexe binaire ont pu être identifiées et confirmées par une étude phylogénétique qui prouve la conservation du modèle proposé (177). Ces différents résultats ont été confirmés par des expériences de crosslinking au psoralène (178) et valident le modèle du complexe binaire PBS/ARNt<sup>Lys,3</sup> (**Figure 12C**).

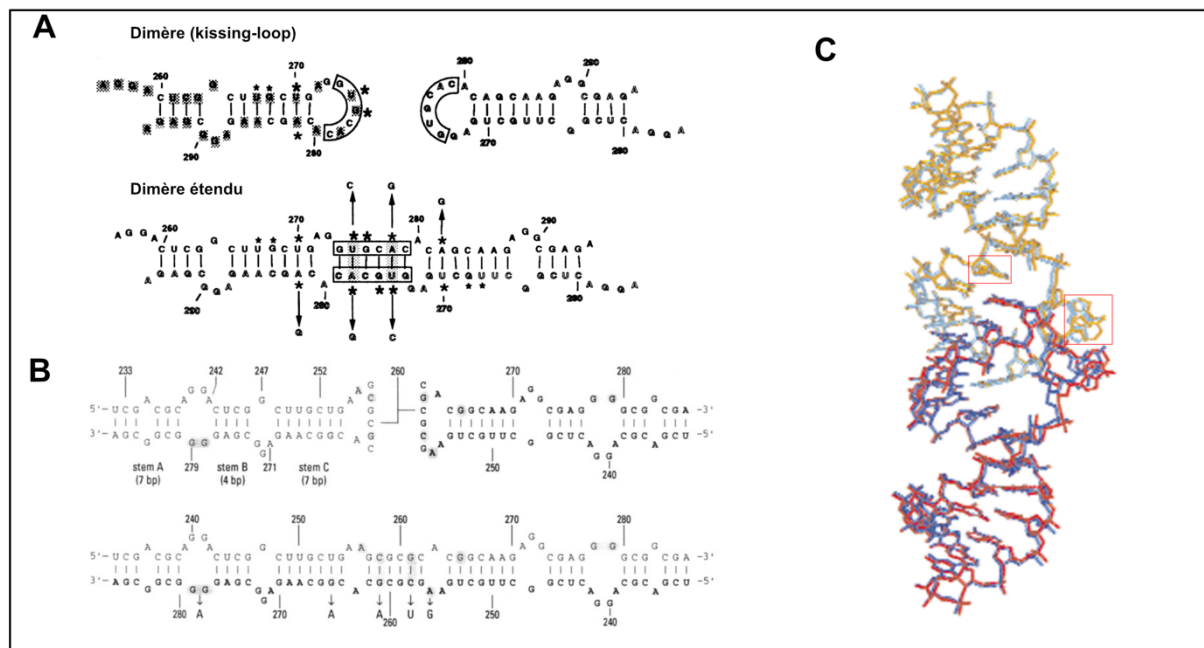


**Figure 12** : Représentation du domaine PBS (A) seul ou (B et C) en complexe avec l'ARNt<sup>Lys,3</sup>. (A) La structure est issue d'expériences de cartographie chimique et adaptée de Baudin *et al.*, (1993) (158). (B) Le modèle représente les effets de la fixation de l'ARNt<sup>Lys,3</sup> sur les réactivités du domaine PBS d'après les expériences de Isel *et al.*, (1993) (176). (C) Le complexe binaire entre le PBS et l'ARNt<sup>Lys,3</sup> a été modélisé par Isel *et al.*, (1995) (177). Le PBS et l'ARNt<sup>Lys,3</sup> sont représentés en noir et en rouge, respectivement.

## INTRODUCTION

### IV.1.5/ Le site d'initiation de la dimérisation

Le site de dimérisation du VIH-1 a d'abord été supposé localisé dans la région en aval du site donneur majeur d'épissage et formée par des quadruplexes de guanines (179,180). Cependant, en 1994, Skripkin *et al.*, ont identifié et caractérisé par cartographie chimique la région palindromique qui permet l'initiation de la dimérisation, le *dimerisation initiation site* (DIS) de l'ARNg du VIH-1 (110) (**Figure 13A**). Dans cette publication, les auteurs représentent le mode d'interaction non covalent par *kissing-loop* entre les boucles apicales d'une tige-boucle dans chacun des deux brins d'ARNg et, ce faisant, postulent également l'existence d'une forme de dimère « étendue » qui sera reprise et nommée plus tard même si son existence n'a pas encore été prouvée expérimentalement dans des virus (181–183) (**Figure 13B**). Les différentes approches mises en œuvre pour définir la structure adoptée par la tige-boucle contenant le site d'initiation de la dimérisation DIS sont rapidement arrivées à un consensus structural suffisamment robuste pour supporter les différentes séquences des souches étudiées (**Figure 13A** et **13B**). Ce modèle de dimérisation a également été validé par des expériences de cristallographie dont les résultats mettent en évidence une identité structurale des motifs entre les isolats MAL et LAI (**Figure 13C**) qui est conservée jusqu'à « l'exclusion » des purines flaquant la séquence palindromique (184).



**Figure 13** : Premiers modèles de la tige-boucle SL1 du domaine Psi. (A) Structure de la tige-boucle SL1 (isolat MAL) sous forme de dimère (*kissing-loop*) ou de dimère étendu issu de données de cartographie chimique (110). (B) Version étendue de la tige-boucle SL1 (isolat LAI) obtenue par cartographie chimique (181). (C) Structures

## INTRODUCTION

obtenues par cristallographie avec superposition des structures des isolats MAL (bleu et bleu clair) et LAI (rouge et jaune) (184). Les rectangles rouges soulignent les purines exclues du *kissing-loop* dans le dimère.

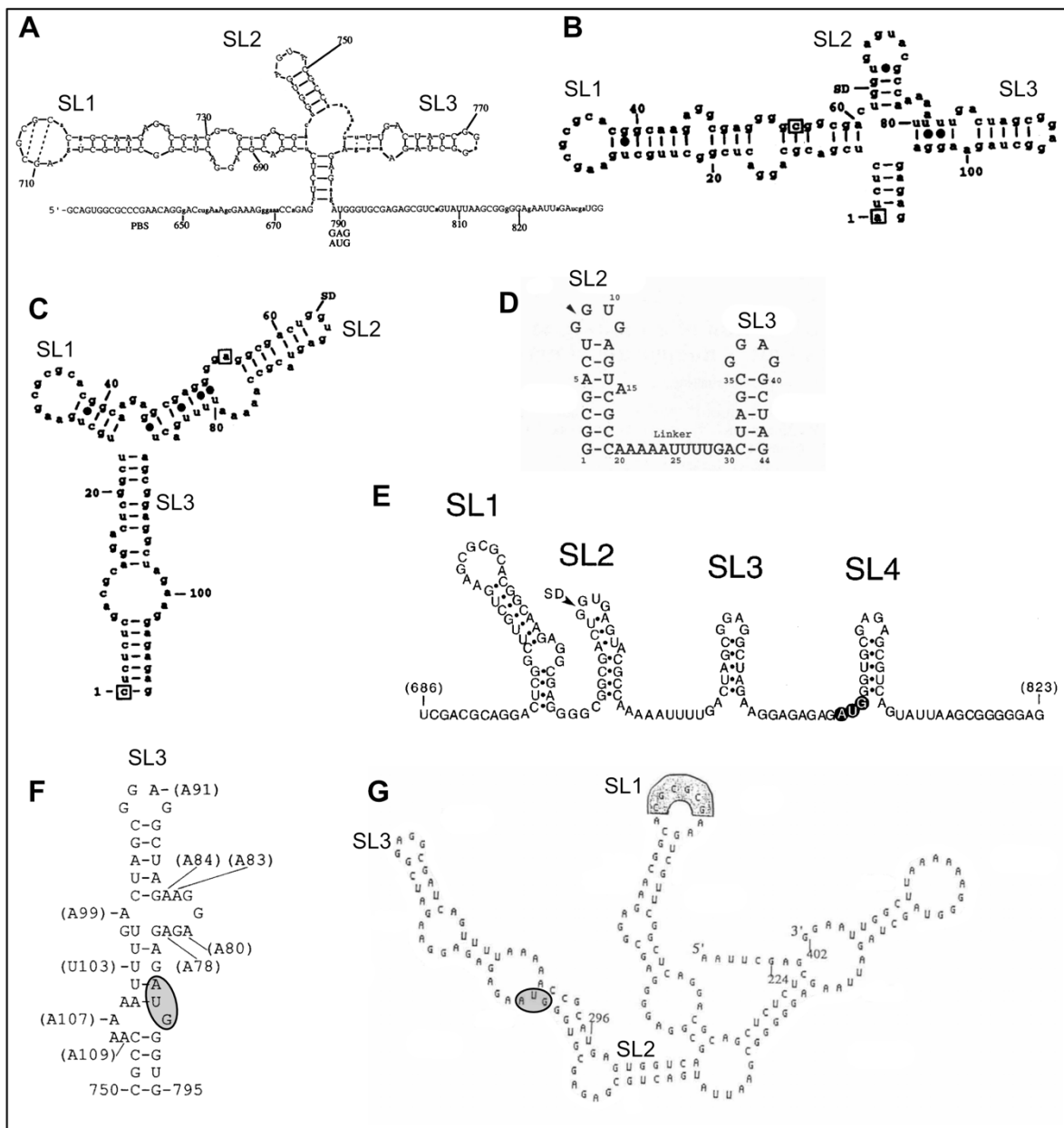
### IV.1.6/ Le signal d'encapsidation

Les études structurales se sont également intéressées à la région permettant l'encapsidation de l'ARNg, appelée *packaging signal* (Psi ou «  $\Psi$  ») (185–188). Ces régions Psi avaient déjà été documentées chez certains rétrovirus comme ayant des structures reconnues spécifiquement par le domaine de nucléocapside des protéines virales structurales Gag (189,190). Dans le cas du VIH-1, une expérience de délétion entre le site donneur majeur d'épissage et le codon AUG du gène *gag* avait permis, dans un premier temps, de situer la région contenant les signaux d'encapsidation dans l'extrémité 5' du génome, en aval du site donneur principal d'épissage (151). En 1992, Harrison et Lever ont proposé le premier modèle structural expérimental d'un grand domaine allant du PBS à la séquence codante de *gag* (51) (**Figure 14A**). Ce modèle a été réalisé à partir d'une combinaison d'expériences de cartographie chimique et enzymatique, d'analyses phylogénétiques et de modèles prédictifs utilisant le minimum d'énergie libre. La structure présente une jonction à 4 voies fermée par une interaction à longue distance CU:GA (191,192), contenant [1] la tige-boucle SL1 dont la boucle apicale expose le DIS et qui possède 2 boucles internes riches en purines, [2] la tige-boucle SL2 qui expose le site donneur majeur d'épissage dans sa boucle apicale et [3] la tige-boucle SL3, supposée contenir le signal principal d'encapsidation. Ce modèle a été conforté par une approche prédictive mise en place par Han et Kim, (1993) pour l'ARNg du clone moléculaire NL4-3 (193) (**Figure 14B**). Cependant, cette dernière étude a surtout mis en avant la sensibilité abusive des approches prédictives par énergie libre minimale dont les prédictions sont hautement influencées par la présence de mutations ponctuelles (**Figure 14C**). D'autres expériences de cartographie réalisées par Sakaguchi *et al.*, (1993) (**Figure 14D**) ou Clever *et al.*, (1995) (**Figure 14E**) ont permis de proposer un modèle alternatif de la structure de Psi (194,195). Dans ce modèle, les repliements des tiges-boucles SL2 et SL3 postulés par Harrison et Lever adoptent une conformation moins allongée. Cela permet le repliement d'une tige-boucle SL4, contenant le codon AUG de *gag*. Une structure alternative pour ces 3 tiges-boucles qui engloberait les régions SL2, SL3 et le codon AUG a également été obtenue par une combinaison d'approches de cartographie enzymatique et chimique (52,196) (**Figure 14F**).

## INTRODUCTION

Malgré l'absence des régions 5' et 3' du domaine Psi, ce modèle ne souffre pas la comparaison avec les autres structures proposées dans la littérature. En effet, des résultats issus d'approches prédictives (182) supportent également ce modèle en tige-boucle (**Figure 14G**).

Cependant, même si les modèles obtenus divergent, la majorité conduit tout de même à un léger consensus quant aux régions SL1 et SL3 qui exposent, respectivement, la séquence palindromique permettant la dimérisation avec au moins une boucle interne et un motif apical GGAG identique.



**Figure 14** : Premiers modèles de la région contenant le domaine d'encapsidation Psi. (A) Modèle de Harrison et Lever allant du domaine PBS au début de la séquence codante de *gag* (Adapté de 51). (B et C) Prédictions de Han et Kim, (1993) des signaux d'encapsidation (nucléotides 227 – 335) du VIH-1 (B) LAI et (C) NL4-3 (193).

## INTRODUCTION

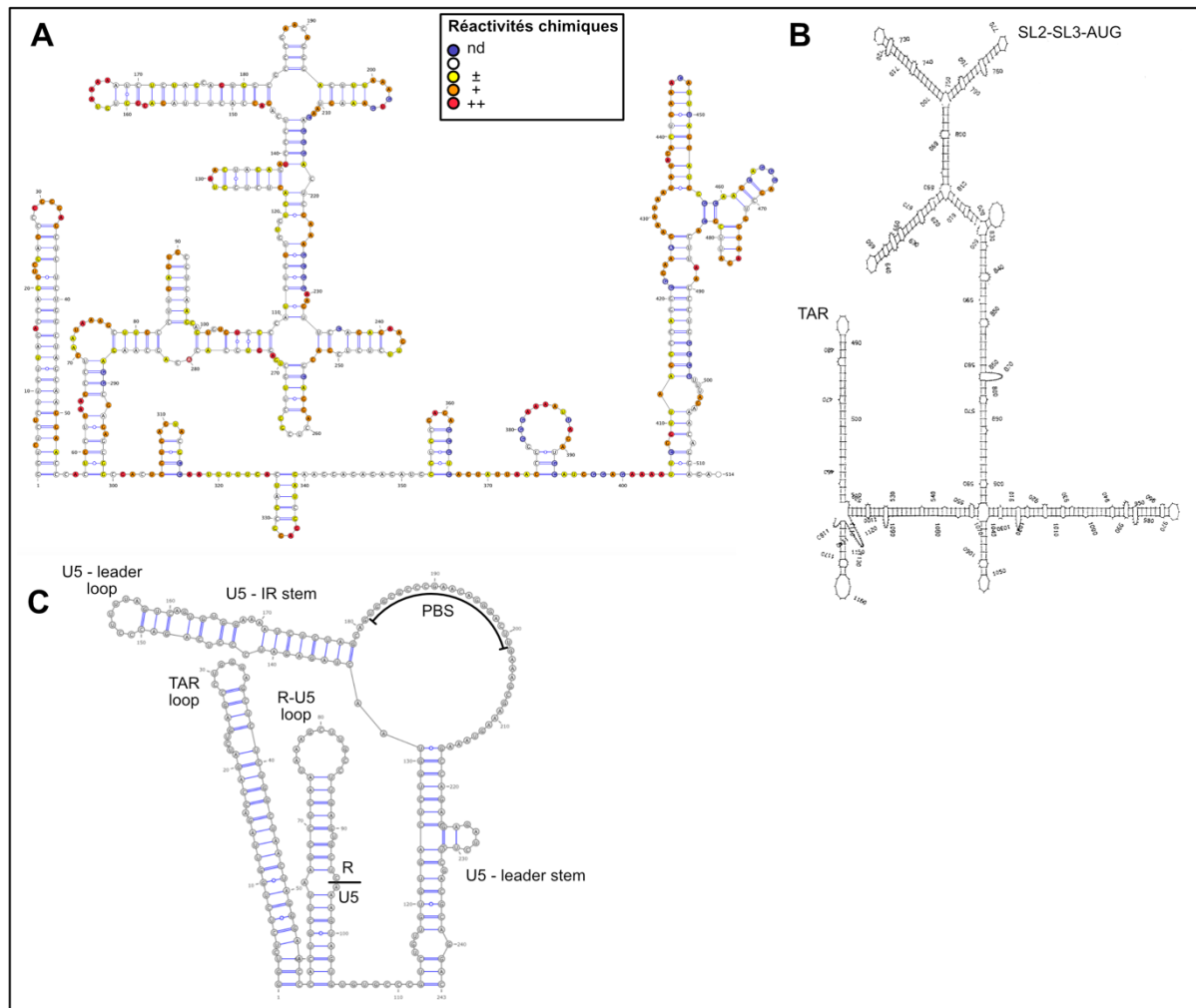
(D) Représentation du modèle SL2-SL3 de Psi par Sakaguchi *et al.*, (1993) (194). (E) Représentation du modèle de Psi par Clever *et al.*, (1995) (195). (F) Représentation de SL3 prédite par Hayashi *et al.*, (1992) (52). (G) Représentation du modèle de Psi par Muriaux *et al.*, (1995) (182). Le codon d'initiation de *gag* est entouré.

### IV.2/ La région 5' terminale de l'ARNg du VIH-1 dans sa globalité

En s'intéressant aux différentes structures portant les fonctions nécessaires au bon déroulement du cycle viral, il est rapidement devenu évident que les repliements des domaines structuraux pouvaient être affectés par la présence des régions voisines à ces domaines. La première étude à s'être intéressée à la globalité de la région 5' terminale de l'ARNg du VIH-1 (Isolat MAL) par cartographie chimique et enzymatique subdivise les 510 premiers nucléotides en 3 domaines structuraux distincts : TAR, PBS et DIS-Psi-AUG (158) (**Figure 15A**). Ce modèle prend en compte les variations possibles dans la région Psi et représente les différentes structures associées. Un second modèle, également réalisé par des approches de cartographie (52) offre une représentation dont les seules régions communes avec le précédent modèle sont celles du TAR et de la structure alternative à SL2-SL3 et SL4 (**Figure 15B**). Cependant, l'analyse détaillée du modèle a posteriori est rendue complexe par le manque de précision de sa représentation dans la figure de l'article. En 1993, un troisième modèle, issu de prédictions, a été proposé (197) (**Figure 15C**). Ce modèle est construit par comparaison structurale avec le modèle de l'ARNg du SIV et force la formation des structures TAR et PolyA. Comparé aux autres modèles rapportés dans la littérature de l'époque, une structure altérée du PBS est proposée, proche de celle du SIV, en absence de la région Psi. Parmi ces trois premiers modèles, il est intéressant de noter que le modèle comparatif et prédictif de Rizvi et Panganiban, (1993) est le plus proche des structures actuellement mises en avant dans la littérature (**Figures 15C vs 16 et 17 pour comparaison**).



## INTRODUCTION

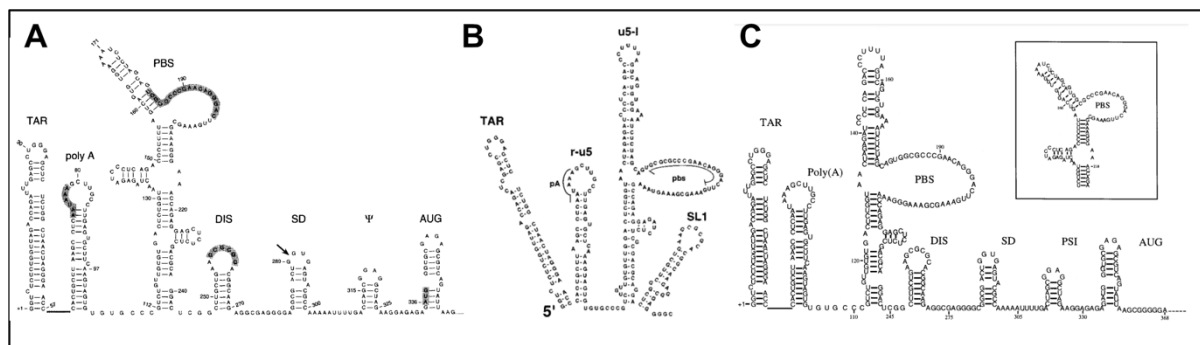


**Figure 15** : Premières structures de la région 5' terminale de l'ARNg du VIH-1. (A et B) Modèles réalisés à partir de données expérimentales issues de cartographie au (A) DMS, CMCT et DEPC (Baudin *et al.*, 1993) ou (B) au DEP et à la nucléase S1 (Hayashi *et al.*, 1993) (158,196). (C) Modèle prédictif par Rizvi et Panganiban (1993) (197). Les représentations A et C ont été reprises et dessinées sur VARNA (Darty *et al.*, 2009) pour améliorer la qualité des structures proposées (198). Le modèle B n'a pas pu être modifié par manque d'informations. Aucun consensus n'est identifié sur la région 5' terminale.

En 1996, une structure de la région 5' terminale a été publiée par Ben Berkhout (199) (**Figure 16A**). Cette représentation rassemble la majeure partie des observations réalisées sur les domaines isolés lors de différentes études (voir figures partie IV.1/ pour comparaison) et s'est basée sur des analyses phylogéniques lorsque les données recueillies ne permettaient pas de trancher. En revanche, l'auteur précise que des boucles plus grandes auraient pu être modélisées par l'ajout de boucles internes ou d'autres éléments déstabilisants. L'année suivante, dans une étude sur les *cis element* pour l'encapsidation, McBride *et al.*, (1997) ont proposé une structure dont la région TAR-polyA est similaire à celle du modèle de Berkhout

## INTRODUCTION

(1996), mais dont les domaines PBS et Psi diffèrent légèrement (200) (**Figure 16B**). Dans leur représentation, la tige-boucle SL1 étendue est retrouvée, de même que le domaine PBS dont la structure est altérée pour permettre une structure plus longue, comme représentée par Rizvi et Panganiban (1993). En 1998, une étude s'intéressant à la capacité des protéines contenant le domaine de nucléocapside (NC) à reconnaître certaines régions de l'ARNg du VIH-1 a rapporté une structure issue de données de cartographie enzymatique (201) (**Figure 16C**). Le modèle reprend des fragments des deux structures précédentes, avec un domaine PBS altéré et une tige-boucle SL1 identique à celle présentée par Berkhout (1996). Les données structurales obtenues favorisent la représentation de Berkhout (1996) à l'exception du domaine PBS qui est plus proche de celui proposé par Rizvi et Panganiban, 1993. Ainsi, à la fin du 2<sup>ème</sup> millénaire, différents modèles structuraux, issus de prédictions, d'analyses phylogéniques et de résultats expérimentaux de cartographie chimique et enzymatique co-existent dans la littérature. Ces modèles, parfois relativement proches les uns des autres, sont difficilement rejetables du fait des données expérimentales les accompagnant. L'explication des divergences est alors hypothétisée sur l'existence de différentes conformations de l'ARNg du VIH-1, portant chacune une ou plusieurs des fonctions que l'ARNg doit remplir (202).

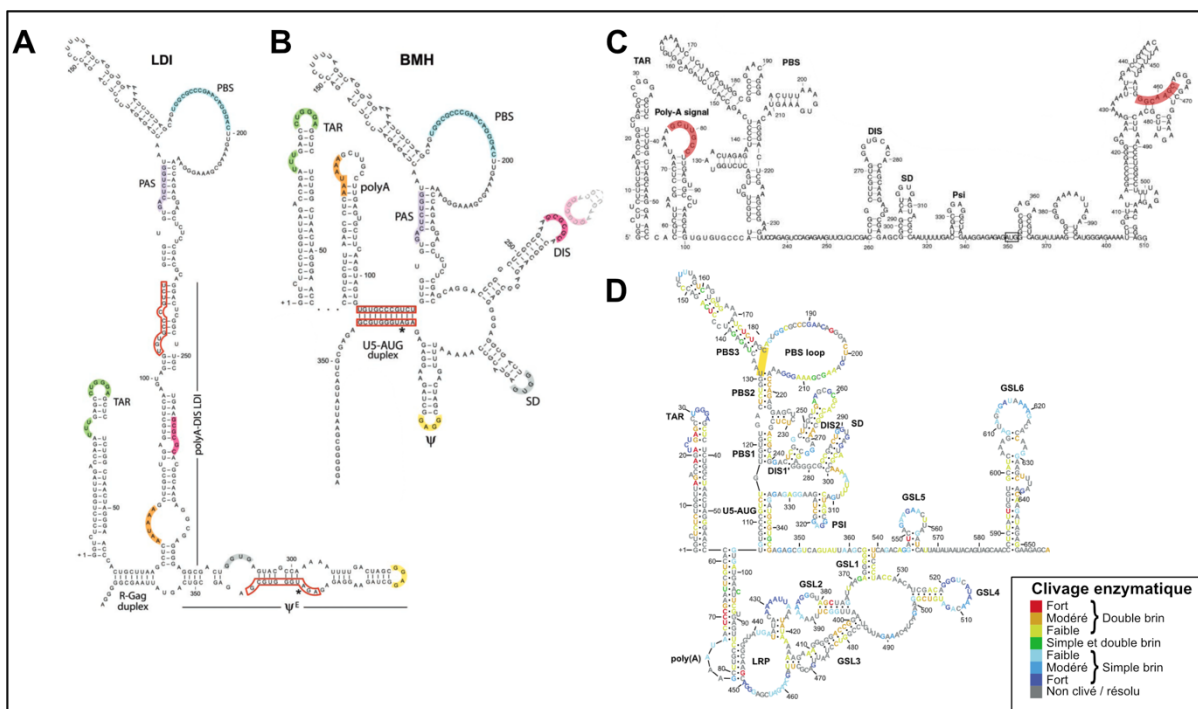


**Figure 16** : Évolution des modèles structuraux de la région 5' terminale de l'ARNg du VIH-1. (A) Représentation des 361 premiers nucléotides par Berkhout (1996) (199). (B) Représentation des 281 premiers nucléotides par McBride *et al.*, (1997) (200). Adaptation du modèle de Damgaard *et al.*, (1998), relativement proche du modèle de Maddison *et al.*, (1998) (55,201). Ces différentes représentations de la région 5' terminale commencent à trouver un consensus sur les régions R, U5 et PBS. La tige-boucle SL1 est modélisée (A et C) courte ou (B) étendue.

# INTRODUCTION

## IV.2.1/ Le polymorphisme de la région 5' terminale et l'activité biologique

L'ARNg du VIH-1 s'acquiesse du rôle de matrice pour la synthèse des protéines virales structurales et enzymatiques (Pr55<sup>Gag</sup> et Pr160<sup>Gag-Pol</sup>), mais également de génome à encapsider sous forme dimérique dans les nouveaux virions. L'idée de conformères capables de réaliser l'une ou l'autre de ces fonctions apparait avec les modèles « *Long Distance Interaction* » ou LDI et « *Branched Multiple Hairpins* » ou BMH de Huthoff et Berkhout, (2001) (**Figure 17A et B**). Le modèle « LDI » se replie en une longue tige-boucle éliminant les motifs PolyA et Psi qui interagissent ensemble. En conséquence, le motif palindromique du DIS est séquestré dans une hélice empêchant ainsi la dimérisation et favorisant la traduction. Le modèle « BMH », quant à lui, adopte une structure plus compacte et qui permet l'exposition de la région palindromique dans une boucle apicale (203). Dans ce modèle, l'ARN devient compétent pour l'étape de dimérisation, et donc pour l'encapsidation et par la suite, la transcription inverse (204).



**Figure 17** : Représentation des premiers modèles de conformères et des interactions à longue distance. (A et B) Modèles des structures (A) « *Long Distance Interaction* » ou (B) « *Branched Multiple Hairpins* » adapté de Abbink et Berkhout (2003) (205). Les séquences impliquées dans l'interaction à longue distance U5:AUG sont entourées en rouge. (C) Modèle de la région 5' terminale de l'isolat MAL avec l'interaction à longue distance PolyA:gag représentée en rouge (adapté de 206). (D) Modèle de la région 5' terminale avec les interactions U5:AUG représentées (selon 207).

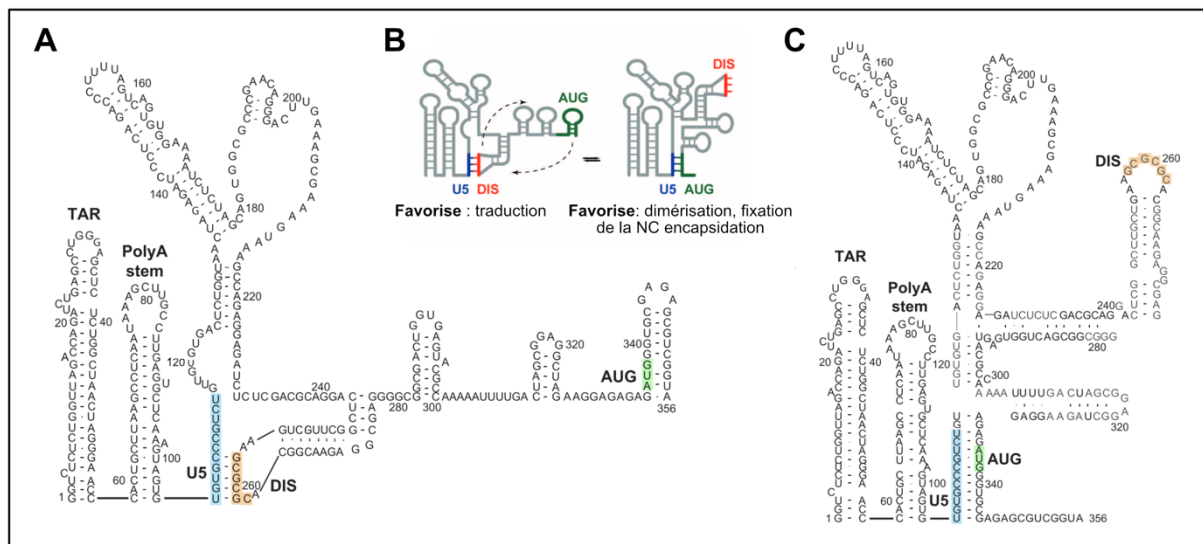
## INTRODUCTION

Pourtant, ce modèle d'interrupteur structural LDI – BMH a rapidement été rejeté. En effet, différents éléments abondant contre l'existence de la structure LDI sont apparus dans la littérature dans les années suivant sa publication. Dans un premier temps, une publication de Paillart *et al.*, (2002) a identifié une interaction à longue distance entre la région PolyA et la séquence codante de *gag*, dans le domaine de matrice (206) (**Figure 17C**). De plus, l'année suivante, une seconde interaction à longue distance entre les régions U5 et AUG (U5:AUG), permettant la fermeture du domaine Psi, a été identifiée (205) (**Figure 17B**). Enfin en 2004, des expériences de cartographie *in vitro* ont permis à Damgaard *et al.*, de proposer un modèle pour la structure des 744 premiers nucléotides de la région 5' terminale (207) qui est cohérent avec l'existence des différentes interactions à longue distance précédemment identifiées (**Figure 17D**). L'observation de ces différentes interactions est incompatible avec le modèle LDI qui séquestre les régions PolyA et U5 ce qui a conduit au rejet de ce modèle. En revanche, elles peuvent exister dans le modèle BMH et supporteraient alors ce dernier modèle ainsi que l'hypothèse de la présence d'un interrupteur structural, qui impliquerait une structure alternative autre que la structure LDI. De plus, des mutations visant à perturber la structure LDI et donc, l'activité de traduction associée, n'ont pas donné de résultats concluants (208). Finalement, jusqu'à très récemment, aucune étude externe à l'équipe de Berkhout n'avait réussi à mettre en évidence l'existence de la conformation LDI *in vitro* ou *in vivo*. Cependant, des travaux récents (209) appliquant une approche nouvellement développée de « *Functional analysis of RNA structure-sequencing* » (FARS-seq) ont permis d'obtenir des résultats mettant en avant une structure de la région 5' terminale monomérique qui n'est pas sans rappeler le conformère LDI, également prédit par Kasprzak *et al.*, (2005) (205,210).

Malgré cette nouvelle étude, les éléments précédemment présentés avaient conduit la communauté scientifique, à la fin des années 2000, à rejeter le modèle LDI au profit d'un second proposé par le groupe de Michael Summers et basé sur une approche de RMN, plus fiable dans le cas d'études de la structure de petits ARN (211). Dans leur modèle d'interrupteur structural, les auteurs développent une idée similaire à celle du groupe de Berkhout avec la présence d'une structure favorisant la traduction (U5:DIS) (**Figure 18A**) ou bien la dimérisation et l'encapsidation (U5:AUG) (**Figure 18C**) via des interactions à longue distance (**Figure 18B**). Dans ce cas, le remodelage structural induit par l'activation de l'interrupteur serait bien moins

## INTRODUCTION

brutal comparé à celui proposé précédemment par B. Berkhout (**Figure 17A et 17B**) et reposerait sur la reconnaissance de l'ARNg par le domaine de nucléocapside de Pr55<sup>Gag</sup> (212). En outre, la comparaison des modèles structuraux proposés (**Figure 18A et 18C**) avec le modèle BMH montre des similarités remarquables entre les domaines TAR, PolyA, PBS, SL1, SL2 et SL3. Les différences les plus importantes se situent au niveau des interactions à longue distance, de la longueur des tiges-boucles dans le domaine Psi, mais également au niveau de l'organisation de ce domaine sous forme d'une *4-ways junction* dans le modèle de U5:AUG, qui contraste avec la *5-ways junction* proposée dans le modèle BMH. À la suite de cette proposition de modèles reposant sur des interactions à longues distances impliquant U5, d'autres études ont également apporté des éléments de preuves confirmant l'existence d'interactions dans cette région (113,213,214).

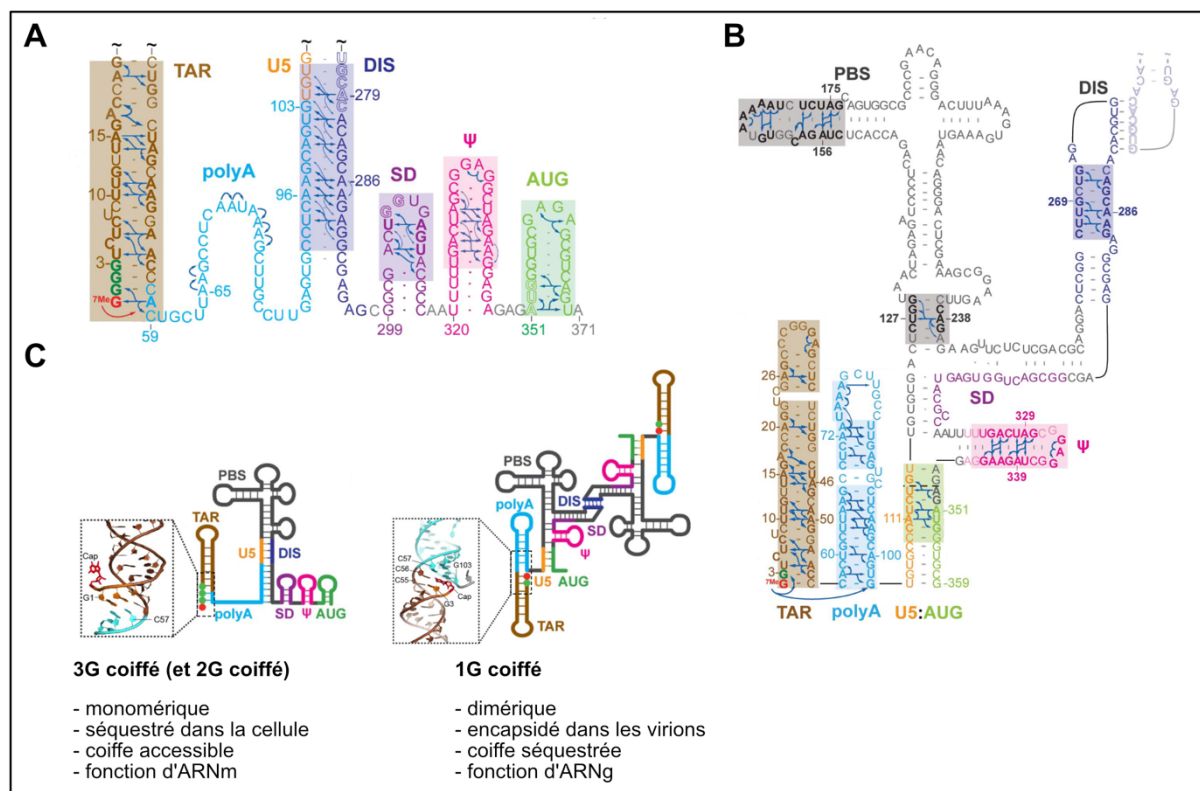


**Figure 18** : Représentation des modèles de conformères proposés par l'équipe de Summers (211). (A et C) Les modèles (A) favorisant la traduction de l'ARNg ou (C) favorisant sa dimérisation, la fixation du domaine de NC et l'encapsidation sont adaptés de Boeras *et al.*, (2017) (106). (B) Représentation schématique du changement de conformation entre U5:DIS et U5:AUG.

Un dernier modèle permettant de décrire le double rôle de l'ARNg du VIH-1 a été proposé dans la littérature en 2016 (215). Ici, les auteurs mettent en évidence la présence d'une hétérogénéité de site pour le début de la transcription qui conduit à la synthèse d'ARNg commençant avec 1, 2 ou 3 guanosines coiffées (1G, 2G ou 3G, respectivement). D'après cette étude, les ARNg 1G auraient une tendance supérieure à la dimérisation (**Figure 19B**), tandis que les ARNg 2G/3G resteraient monomériques (**Figure 19A**). Cette différence fonderait le

## INTRODUCTION

devenir des transcrits à être dirigés vers la traduction d'une part ou la dimérisation et l'encapsidation d'autre part (**Figure 19C**). Le paradigme avancé pour expliquer ces importantes différences de fonctions se base sur l'analyse des structures secondaires. La présence des 3G favorise un repliement différentiel de la tige-boucle PolyA, stabilisant par la même occasion la structure monomérique (209,216). A l'inverse, l'absence de multiples guanosines à l'extrémité 5' allonge d'une paire de base la tige-boucle PolyA, empêchant par la même occasion la stabilisation de l'interaction U5:DIS et favorisant la dimérisation. Ces résultats ont été confirmés par des expériences de mutagenèse sur la partie inférieure de la tige-boucle PolyA. La stabilité de cette dernière exerce une influence inversement proportionnelle à la monomérisation (106,215). De plus, le nombre de guanosines à l'extrémité 5' de l'ARN semble être impliqué dans l'exposition ou au contraire la séquestration de la coiffe (216–218), modulant ainsi la reconnaissance de cette dernière par les protéines impliquées dans la traduction.



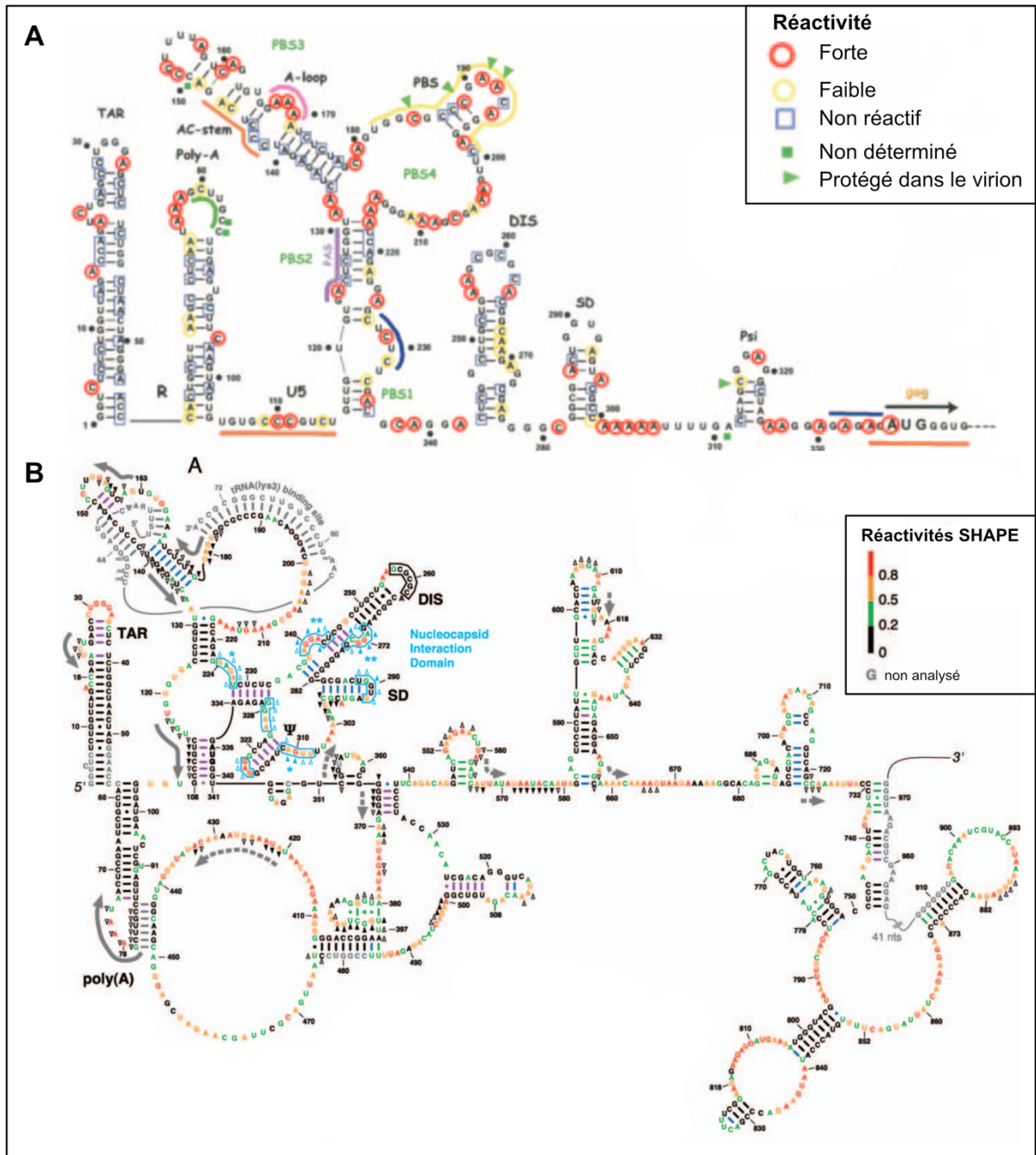
**Figure 19** : Représentation du modèle d'hétérotranscription proposé par Kharytonchuk *et al.*, (2016) (215). (A) Modèle de la structure secondaire d'une partie de la région 5' terminale des ARN 3G et 2G coiffés (216). (B) Modèle de la structure secondaire de la région 5' terminale des ARN 1G coiffés (216). (C) Représentation schématique du modèle d'hétérotranscription (218).

## INTRODUCTION

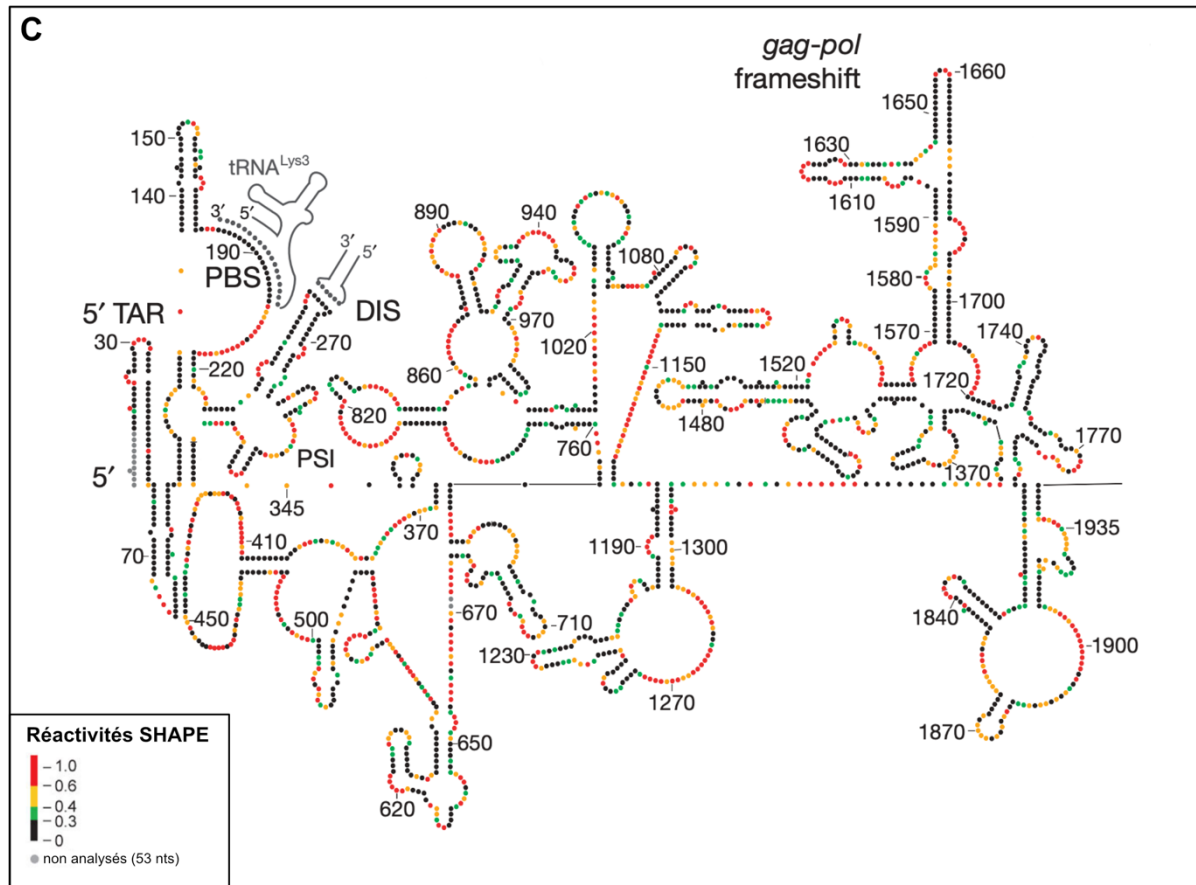
### IV.2.2/ Les modèles structuraux *in virio*

Pour compléter cette présentation des modèles de structures de la région 5' terminale de l'ARNg du VIH-1, il me paraît important de discuter les structures rapportées dans la littérature dans les années 2000 et obtenues à partir de particules virales. En effet, les études *in virio* présentent l'avantage de prendre en compte la totalité de l'ARNg et d'inclure directement tous les partenaires impliqués dans le cycle viral. En 2004, Paillart *et al.*, présentent pour la première fois des résultats de cartographie chimique sur un ARNg issu de cellules infectées ou bien de particules virales (219) (**Figure 20A**). Les modèles obtenus ne montrent que quelques différences mineures entre les deux conditions, émettant l'hypothèse d'une structure globale similaire entre ARNg avant et après encapsidation dans les particules virales. Cependant, il est intéressant de noter que certaines régions semblent protégées dans le virion. En effet, le PBS, qui est réactif à la modification dans les cellules, ne l'est plus dans la particule virale. De même, la tige-boucle SL3 est protégée dans le virion. Ces différences proviennent probablement de la fixation de l'ARNt<sub>Lys,3</sub> et de protéines NCp7. Finalement, le modèle présenté ici rappelle celui du BMH proposé par Berkhout (**Figure 17B**) à la différence que celui-ci ne propose pas l'interaction à longue distance permettant la fermeture du domaine Psi dans sa représentation. Quoiqu'il en soit cette ressemblance entre les données issues d'expériences de cartographie *in vitro* et *in virio* semble suggérer une similarité de la structure globale entre ces deux conditions.

# INTRODUCTION







**Figure 20** : Représentation des modèles structuraux issus de données de cartographie *in virio*. (A) Premier modèle de la région 5' terminale de l'ARNg du VIH-1 issu de particules virales (219). (B) Modèle des 1000 premiers nucléotides de l'ARNg du VIH-1 issus de particules virales matures par une approche SHAPE (220). (C) Premier modèle structural de la totalité de l'ARNg du VIH-1 (152). Seuls les 1950 premiers nucléotides ont été représentés pour des raisons de clarté.

En 2008, une nouvelle étude sur la structure de l'ARNg du VIH-1 *in virio* a été publiée. Celle-ci s'intéresse aux 1000 premiers nucléotides par une approche de hSHAPE (220) dans 4 conditions (**Figure 20B**) : [1] dans la particule virale (*in virio*) et hors de la particule virale après déprotéinisation à l'aide [2] d'une enzyme (*ex virio*) ou [3] d'un agent chimique perturbant l'interaction entre l'ARNg et la nucléocapside (AT-2) et [4] un ARN synthétisé *in vitro* en absence de tout composant viral (*in vitro*). La structure proposée est la même pour les quatre conditions d'intérêt (*ex virio*, *in virio*, AT-2 et *in vitro*), ce qui renforce l'idée de correspondance entre les structures obtenues *in vitro* et *in virio*. Les différences issues de la fixation de l'ARN<sup>t<sup>Lys,3</sup></sup> sont encore une fois identifiées en comparant les conditions *ex virio* et *in vitro*. De plus, la structure adoptée par le PBS rappelle celle proposée par Rizvi et Panganiban, (1993) ou, plus récemment, celle de Damgaard *et al.*, (2004) (**Figure 17D**). Des interactions à longue

## INTRODUCTION

distance sont également représentées (CU:GA, U5:AUG et PolyA:Gag) en accord avec les réactivités attendues. En revanche, le domaine en aval du codon d'initiation AUG qui est globalement moins structuré est différent de celui proposé dans les premières structures (158). En effet, la séquence codante de *gag* ressemble plus à celle proposée par Damgaard *et al.*, (2004) (207) (**Figure 17D**).

Le dernier modèle présenté ici est celui de Watts *et al.*, (2009) où la structure de la totalité de l'ARNg a été cartographiée par hSHAPE (152) (**Figure 20C**). Les deux modèles, proposés à une année d'écart par la même équipe, sont très semblables, tant sur les repliements qu'au niveau des réactivités associées. Il est également intéressant de noter que la dernière structure a été modélisée par un logiciel prenant en compte la totalité de l'ARNg du VIH-1, et que cela n'a pas d'effet sur le repliement final obtenu. Cependant, cette structure semble aussi avoir de nombreux défauts. En effet, une étude a mis en avant les biais introduits par les méthodologies SHAPE et montre que la structure modélisée par Watts *et al.*, (2009) est incertaine sur plus de 50% des hélices modélisées (221). Concernant la région 5' terminale, les résultats des deux études semblent tout de même aller dans le sens d'une structure globalement cohérente (152,221).

Finalement, les derniers modèles présentés actuellement dans la littérature ne diffèrent plus beaucoup de ceux décrits et modélisés au début des années 2000 (57,84,209,222). En effet, un consensus semble avoir été atteint autour des structures adoptées par les tiges-boucles SL1 et SL3 même si ce n'est pas encore le cas pour la structure du domaine Psi ni celle de la région 5' terminale de l'ARNg qui restent sujets à débat. Cependant, ce désaccord structural révèle l'importance du développement de nouveaux outils méthodologiques toujours plus poussés permettant d'étudier les structures secondaires et tertiaires adoptées par les différents conformères, avec ou sans leur(s) partenaire(s).

## INTRODUCTION

### IV.3/ Implication des protéines structurales virales sur la structure et le rôle de l'ARNg

Comme nous venons de le voir, la littérature scientifique s'intéressant aux structures adoptées par les différentes régions de l'ARNg du VIH-1 est abondante et soumise à discussion, mais une sphère de complexité supplémentaire apparaît lorsque ces études sont réalisées en présence de partenaires protéiques cellulaires et/ou viraux (176,220). Pourtant ces études sont nécessaires puisqu'elles permettent de comprendre les mécanismes moléculaires qui dirigent certaines étapes du cycle réplcatif viral en influençant la structure adoptée par l'ARNg. De ce fait, plusieurs études ont été réalisées pour tenter de modéliser les sites de fixation de différents partenaires ainsi que les effets de leur fixation sur la structure de l'ARN. Pour plus de clarté, je ne m'intéresserai ici qu'aux études ayant ciblé les protéines virales contenant le domaine de nucléocapside (NC), responsables des étapes de sélection et d'encapsidation avec la protéine structurale Pr55<sup>Gag</sup> et de la maturation avec les protéines Pr55<sup>Gag</sup>, NCp15, NCp9 et NCp7 qui font l'objet de mon travail de thèse (pour revue, voir 58,223,224).

#### IV.3.1/ La protéine Pr55<sup>Gag</sup>

Les protéines virales contenant le domaine NC sont impliquées dans les étapes tardives du cycle viral, en particulier par leur capacité à se lier spécifiquement à l'ARNg viral et à médier son encapsidation (189,190,195). Initialement, il était supposé que la reconnaissance spécifique du domaine NC de la protéine Pr55<sup>Gag</sup> s'effectuait dans une région de l'ARN située en aval du site donneur majeur de l'épissage (SD). Cette hypothèse était soutenue par plusieurs éléments. Tout d'abord, l'absence de cette région dans les ARN viraux épissés permettait d'expliquer leur non-sélection par les protéines virales (pour revue, voir 112). De plus, des sites de fortes affinités à la NC avaient été identifiés sur les tiges-boucles SL2 et SL3 (195,225,226) alors que seuls des sites de basses affinités étaient identifiés sur SL1 et SL4 (150,227) ce qui soutenait un modèle de sélection initié sur la région SL2-SL3 du domaine Psi. Cependant, il faut préciser que ces expériences ont été réalisées avec la protéine NCp7 et qu'un second modèle de sélection est apparu dans la littérature lorsque la présence d'un site primaire de fixation de la protéine Pr55<sup>Gag</sup> entière a été démontré dans SL1 (118). En

## INTRODUCTION

effet, des expériences complémentaires de cartographie et d'empreintes par la méthode de SHAPE et de « *filter binding* » sur la région 5' terminale d'ARN sauvage et muté ont permis d'identifier l'hélice inférieure et la boucle interne de SL1 comme la région primaire de fixation de Pr55<sup>Gag</sup>. Cette étude a également détecté les sites de fixations de la protéine sur SL2 et SL3, en accord avec les données obtenues par des approches de *Cross-linking immunoprecipitation* (CLIP) (228).

Ces deux modèles de sélection par Pr55<sup>Gag</sup> (fixation initiale sur SL2-SL3 ou sur SL1) sont relativement proches et cohérents avec différentes observations expérimentales. En effet, Kutluay *et al.*, (2014) ont montré une préférence pour la reconnaissance des guanosines simples brins dans le domaine Psi par la protéine Pr55<sup>Gag</sup> cytosolique et membranaire (228). Ces données ont été corroborées par des expériences avec NCp7 (229) ou plus récemment avec Pr55<sup>Gag</sup> (119,230). La mutagenèse dirigée de ces guanosines montre une hiérarchisation des sites de fixation pour l'encapsidation avec la boucle interne de SL1 comme étant un site de haute importance (49,119), suivi par la tige-boucle SL3. Cependant, il est important de noter que la mutation de ces deux sites simultanément ou bien de l'ensemble des guanosines simples brins présentent dans le Psi ne permettent pas une inhibition totale de l'encapsidation (92,3% et 97,1%, respectivement) (119), suggérant une certaine plasticité de la fixation de Pr55<sup>Gag</sup> à l'ARNg du VIH-1. Finalement, l'ensemble de ces observations rejoint l'idée qu'une nucléation de Pr55<sup>Gag</sup> se fait au niveau du domaine Psi lors de l'assemblage (231–234) et concorde avec le modèle de sélection des ARNg au dépend des ARN viraux épissés proposé par Abd El-Wahab *et al.*, (2014). Ce mécanisme repose sur la présence d'un signal d'inhibition de la sélection en amont du SD, qui serait levé par un signal présent en aval de ce même SD, potentiellement par la mise en place d'interactions à longue distance induites par la dimérisation et/ou la fixation des protéines Pr55<sup>Gag</sup> (118).

Pour finir, il est intéressant de noter que la reconnaissance de l'ARN viral par Pr55<sup>Gag</sup> ne semble pas avoir d'effet de remodelage majeur sur la structure secondaire de l'ARNg (235,236). Cela explique pourquoi les modèles de sélection proposés dans la littérature pour la région 5' terminale de l'ARNg sont similaires à ceux de l'ARNg en absence de la protéine Pr55<sup>Gag</sup>.

## INTRODUCTION

### IV.3.2/ Les protéines NCp15 et NCp9

Les protéines NCp15 et NCp9 ont été relativement peu étudiées en raison notamment de leur présence uniquement transitoire lors de la maturation (67). Cependant, ces protéines semblent jouer un rôle dans la maturation génomique de la particule virale (64,237). En effet, des études ont mis en évidence la capacité de ces protéines à se lier aux acides nucléiques (195,238). De plus, NCp9 semble capable d'affecter suffisamment le génome pour que la particule virale acquière son infectivité (239) et que la reverse transcription se déroule correctement. En revanche, la propriété chaperon d'ARN de NCp15 est aussi faible que celui de Pr55<sup>Gag</sup> (240) et ne permet pas de stabiliser suffisamment le dimère d'ARNg dans la particule virale pour que celle-ci devienne infectieuse (64,237).

Ces résultats montrent que le domaine p6, qui est présent dans NCp15 mais absent de NCp9, module la capacité du domaine NC à lier les acides nucléiques (241,242) et que l'étape de maturation protéolytique est nécessaire au bon déroulement de la maturation génomique. Malgré ces différentes observations, il n'y a, à ma connaissance, aucune étude s'étant intéressée à la structure de la région 5' de l'ARNg en présence des protéines NCp15 et NCp9.

### IV.3.3/ La protéine NCp7

Dans le cas des études impliquant NCp7, de nombreux sites de fixation ont été identifiés sur l'ARNg (120). De plus, il a été montré que comme pour NCp9, la fixation de NCp7 à l'ARNg permet un ou des changements nécessaires aux étapes de transcription inverse (243) et d'intégration (244) lors du cycle viral suivant. En effet, des travaux de FRET ont mis en évidence la capacité de la protéine NCp7 à favoriser la formation du dimère étendu, en présence ou en absence de l'ARNt<sup>Lys,3</sup>, suggérant que ces facteurs puissent jouer un rôle dans le remodelage de l'ARNg (245). Une avancée importante dans la compréhension du mécanisme d'action de NCp7 a été obtenue grâce à des travaux de 2008 par Wilkinson *et al.*, (2008) sur l'ARNg du VIH-1 issu de particules virales (220). Dans cette étude, les auteurs mettent en évidence 7 des sites de fixation de la protéine NCp7 (**Figure 20B**) déjà identifiés dans la littérature au niveau des tiges-boucles SL1, SL2 et SL3 (53,225,227,246). Ces résultats sont particulièrement intéressants lorsqu'ils sont mis en parallèle d'autres résultats obtenus à l'aide de Pr55<sup>Gag</sup> (ou Pr55<sup>GagΔp6</sup>) par des approches de cartographie SHAPE *in vitro* (118,247). En effet, en comparant les approches de Wilkinson *et al.*, (2008) et de Abd El-Wahab *et al.*,

## INTRODUCTION

(2014), de nombreux sites de fixation de Pr55<sup>Gag</sup> sont identiques à ceux de NCp7. De plus, cette comparaison révèle également un site de fixation primaire commun aux deux protéines sur la boucle SL1, malgré une différence au niveau du matériel utilisé. Ces résultats indiquent donc une redondance du mode de reconnaissance de l'ARN par les deux domaines en doigts de zinc (ZF). Cependant, cette fixation de NCp7 à l'ARN n'est pas équivalente à celle de Pr55<sup>Gag</sup>. En effet, il a été montré que la fixation de NCp7 est moins spécifique que celle de son précurseur, même si elle induit des déstabilisations structurales plus importantes (236,240). Cette différence fonctionnelle peut trouver une explication dans l'étude de la maturation protéolytique. En effet, lors du clivage du domaine p6, l'activité modulatrice de fixation des acides nucléiques associée est perdue (241). De plus, le taux de zinc fixé diffère d'une protéine à l'autre avec 2 Zn<sup>2+</sup> fixés sur Pr55<sup>Gag</sup> et 1 sur NCp7 (248). Il a été proposé que le domaine ZF1 pourrait favoriser la déstabilisation des régions doubles brins tandis que la fixation des régions simples brins contenant des guanosines serait médiée par le domaine ZF2 (229). Ces différences pourraient expliquer la variation fonctionnelle observée entre les protéines Pr55<sup>Gag</sup> et NCp7 (248,249).

Pour conclure cet état des lieux sur les motifs structuraux et leurs rôles biologiques, nous venons de voir que de très nombreuses études ont été réalisées sur les caractéristiques biochimiques et biophysiques des différentes protéines virales contenant le domaine de nucléocapside. C'est ainsi que leur capacité à assurer la maturation de l'ARNg du VIH-1 pour permettre son passage d'un dimère instable à un dimère stable et compact a été dévoilée (66,67,250). Cependant, bien que plusieurs études se soient attachées à étudier l'effet chaperon d'ARN de ces protéines, il existe très peu d'informations sur les structures adoptées par l'ARNg lors des étapes clés du cycle viral que sont la sélection, l'encapsidation et la maturation. Cela a pour conséquence que ces étapes ne trouvent pas de consensus mécanistique dans la littérature actuelle, ou sont même une véritable boîte noire en ce qui concerne la maturation.

# Problématique du sujet de recherche

Le génome du VIH-1 est composé d'un homodimère d'ARN de polarité positive dont les deux molécules sont liées entre elles par des liaisons non-covalentes. Suite à la transcription de cet ARN génomique par la machinerie cellulaire, ce dernier peut suivre deux voies différentes dans la cellule pour assurer le bon déroulement du cycle répliatif. Il peut tout d'abord être pris en charge par la machinerie traductionnelle et servir de matrice pour la traduction des protéines structurales Pr55<sup>Gag</sup> et enzymatiques Pr160<sup>Gag-Pol</sup> par un mécanisme de changement de cadre de lecture. L'ARNg peut également être spécifiquement sélectionné dans le cytoplasme par les protéines virales structurales et encapsidé sous forme de dimère dans les particules virales bourgeonnantes. Le génome sera alors mûré concomitamment à la particule virale. Cette maturation génomique est une étape cruciale à l'acquisition du potentiel infectieux et elle permettra le bon déroulement de l'étape de transcription inverse médiée par la polymérase virale lors du cycle viral suivant.

De nombreuses études ont pu mettre en lumière l'importance de la région 5' non traduite de l'ARNg dans la régulation des étapes tardives du cycle viral, notamment par la présence de domaines structuraux permettant le recrutement de partenaires cellulaires et viraux. Ces différents domaines ont été étudiés en détail par des approches biophysiques et biochimiques, en présence ou non de leurs partenaires, afin de comprendre les mécanismes qui sous-tendent leur rôle dans le cycle viral. Ainsi, plusieurs domaines structuraux ont pu être associés à des fonctions précises, mettant en place des mécanismes de régulation variés. Cependant, la significativité biologique de ces structures n'a pas toujours été validée et un consensus n'a pas été atteint quant aux structures globales adoptées par l'ARNg. Ainsi, à ce jour, différents modèles pouvant expliquer les étapes tardives impliquant l'ARNg du VIH-1 co-existent dans la littérature. De plus, il n'existe que très peu d'études concernant les structures des ARNg dimériques encapsidés dans les particules virales, avant, pendant ou après la maturation.

Depuis de nombreuses années, notre équipe s'implique dans la recherche fondamentale sur le VIH-1, permettant par exemple l'identification de la séquence palindromique responsable de la dimérisation du génome viral, mais également l'existence

## PROBLEMATIQUE

d'une interaction à longue distance entre la région PolyA et la séquence codante de gag. L'expertise de l'équipe dans les approches de cartographies chimiques et enzymatiques, couplée à des approches biophysiques, a permis de mettre en avant la présence de sites de fixation de Pr55<sup>Gag</sup> dans la région Psi, avec un site de haute affinité dans la partie basale de la tige-boucle SL1. Dans son objectif de mieux comprendre les mécanismes moléculaires qui entourent les ARNg viraux, le laboratoire a également développé une méthode novatrice qui permet d'étudier à l'échelle du nucléotide, l'importance de la séquence primaire et de la structure secondaire d'un ARN dans le maintien de sa fonction par *Mutational interference mapping experiment* (MIME).

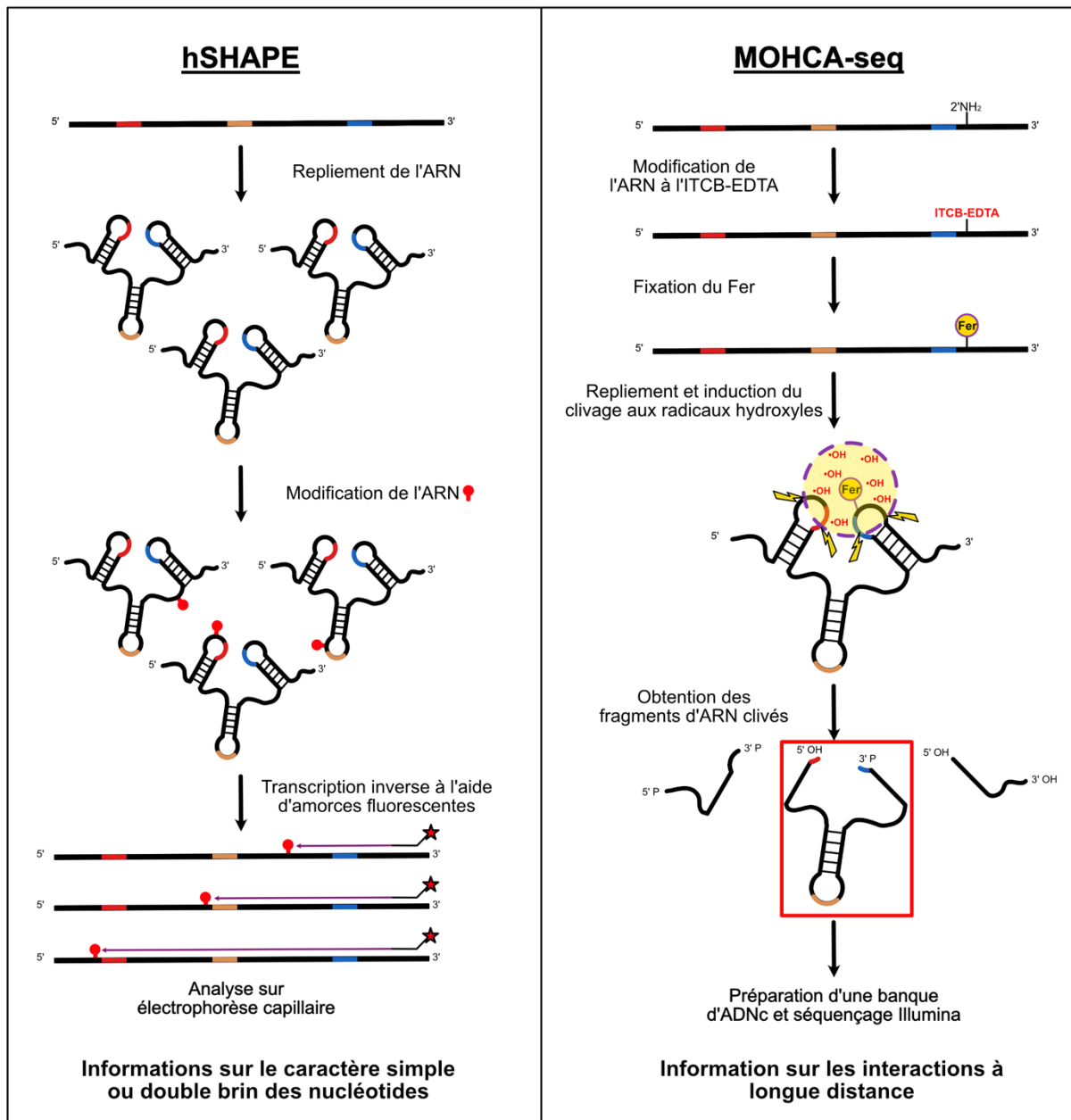
Ainsi, dans le cadre des études précédemment menées par l'équipe, mon projet de thèse s'est axé autour des différentes structures adoptées par la région 5' de l'ARNg du VIH-1, en présence ou non de différentes protéines virales, à l'aide d'approches de cartographies chimiques en solution. Ce projet a été scindé en deux parties. Premièrement, je me suis intéressé à résoudre la structure des 600 premiers nucléotides de l'ARNg, *in vitro*, avant de caractériser les sites de fixation des protéines contenant le domaine de nucléocapside (NC), d'identifier leurs effets sur la structure secondaire de l'ARN et d'analyser l'importance de leurs domaines en doigt de zinc pour le maintien de leur interaction avec l'ARN. Dans un second temps, j'ai souhaité identifier les interactions à longue distance impliquées dans le maintien de la structure tridimensionnelle des 600 premiers nucléotides de l'ARNg afin de produire un modèle plus robuste incorporant des données expérimentales 3D.

Pour ce faire, j'ai mis en place deux approches de cartographies chimiques en solution, le *high-throughput Selective 2'Hydroxyl Acylation analysed by Primer Extension* (hSHAPE) et le *Multiplexed •OH Cleavage Analysis with paired-end sequencing* (MOHCA-seq) (**Figure 21**). La première approche consiste à modifier les 2'OH des nucléotides flexibles (non appariés et/ou non protégés par des interactions protéiques) à l'aide d'un agent acylant. Ces modifications sont alors détectées par une étape de transcription inverse et analysées par électrophorèse capillaire. Dans le cas du MOHCA-seq, cette approche de cartographie chimique multidimensionnelle repose sur le clivage simultané de deux régions d'un ARN lorsque celles-ci se trouvent à proximité immédiate dans l'espace 3D. Pour cela, un groupement chimique qui permet l'induction de radicaux hydroxyles est incorporé de façon statistique dans l'ARN étudié. Les radicaux hydroxyles induisent un clivage de l'ARN dans la proximité immédiate à



## PROBLEMATIQUE

leur production. Les deux extrémités d'un même fragment d'ARN issus d'un clivage sont détectées et analysés par séquençage à haut-débit en mode *paired-end*.



**Figure 21 :** Approches méthodologiques employées pour l'étude de la région 5' de l'ARNg du VIH-1. Les 604 premiers nucléotides de l'ARNg du VIH-1 sont étudiés par hSHAPE et MOHCA-seq afin de recueillir des informations sur leurs caractères simple brin ou double brin, ainsi que les partenaires de ces interactions.



## Axe 1

Étude de la structure de la région 5' de  
l'ARNg du VIH-1

Réarrangements structuraux induits par la  
fixation des protéines contenant le  
domaine de nucléocapside (NC)  
au cours de la maturation



## AXE 1

Lors du cycle infectieux du VIH-1, l'ARNg interagit avec de nombreux partenaires ayant la capacité de moduler les différents mécanismes nécessaires à sa réplication comme la transcription, l'épissage, l'export nucléaire, la traduction, la sélection ou encore la maturation. La fixation de ces partenaires à l'ARNg peut parfois affecter sa structure et ainsi influencer les fonctions qu'il porte. C'est le cas de la protéine Pr55<sup>Gag</sup> qui assure la sélection et l'encapsidation de l'ARNg, mais également des autres protéines contenant le domaine de nucléocapside et issues de la cascade protéolytique, qui sont impliquées dans l'étape de maturation du virion. L'activité chaperon d'ARN de ces différentes protéines a été caractérisée (251) et leur responsabilité dans la maturation génomique a également été mise en évidence (64,66). Cependant, les réarrangements structuraux du dimère d'ARNg induits par ces différentes protéines n'ont que très peu été étudiés.

Afin d'étudier les remodelages de la structure du dimère d'ARNg après la fixation des protéines Pr55<sup>Gag</sup>, NCp15, NCp9 et NCp7 auxquels je me suis intéressé lors de la première partie de mon travail de thèse, j'ai mis en place des protocoles de hSHAPE permettant l'étude des 604 premiers nucléotides de l'ARNg seul ou en complexe avec ces différentes protéines. Différentes raisons nous ont poussé à nous intéresser à cette région particulière de l'ARNg. [1] La région 5' terminale de l'ARNg contient de nombreuses structures fonctionnellement impliquées dans différentes étapes du cycle répliatif viral. [2] De nombreux sites de fixation, d'affinités variables, ont été identifiés dans les premiers nucléotides de l'ARNg (49,228,247). [3] Notre laboratoire a montré qu'une région située au centre de cet ARN, entre les nucléotides 227-337 correspond au noyau de fixation des protéines Pr55<sup>Gag</sup> lors de la sélection (57,70). (4) Lors de l'étude structurale d'une région particulière d'un ARN, il est nécessaire de considérer la potentielle dépendance de cette région vis-à-vis des domaines structuraux externes, impliqués dans des interactions à longue distance (206). En ce qui concerne les complexes ribonucléoprotéiques, nous avons choisi de les former *in vitro* avec un ratio de 1 ARN pour 60 protéines, qui est approximativement le ratio retrouvé dans les particules virales. De plus, nous avons établi 4 protocoles différents afin de pouvoir étudier les effets induits par la présence des protéines. Les 4 conditions sont les suivantes :

- **NoProt** : correspond à l'ARN replié en l'absence de toute protéine. La cartographie hSHAPE fourni des informations sur la structure secondaire de l'ARN.
- **Complexe** : l'ARN est replié en présence des protéines et la modification chimique est réalisée directement sur le complexe ARN-protéines. L'analyse de cette expérience

permet d'identifier les sites de fixation des protéines et les remodelages de la structure qu'elles induisent.

- **Protéinase K** : Après la formation du complexe, les protéines sont éliminées par un traitement à la protéinase K et l'ARN déprotéinisé est modifié par le réactif hSHAPE pour la cartographie. Les modifications structurales stables induites par la fixation des protéines sont mises en évidence.
- **AT-2** : Après la formation du complexe, celui-ci est traité à l'aldrithiol-2 (AT-2) puis la modification chimique permettant le hSHAPE est réalisée. L'AT-2 est utilisé pour modifier les cystéines, perturbant ainsi l'interaction avec les ions  $Zn^{2+}$  et par conséquent la formation des doigts de zinc du domaine NC (252). Les protéines fixées à l'ARN uniquement à l'aide de leur doigt de zinc sont retirées du complexe et il devient possible d'estimer le mode de liaison des protéines pour chaque site sur l'ARN (220,253).

# Matériel & Méthodes :

## I/ Production de l'ARN *in vitro*

Le plasmide pNL4-3 clone 159 utilisé comme matrice pour les transcriptions *in vitro* dérive du pUC19, il était déjà disponible à mon arrivée au laboratoire. Les 604 premiers nucléotides du plasmides pNL4-3, correspondant à la région 5' UTR de l'ARNg du VIH-1 et les 300 premiers nucléotides du gène *gag*, ont été ligués sous le contrôle d'un promoteur T7. L'insert est flanqué du site de restriction *EcoRI* en amont du promoteur T7 et du site *PvuII* à l'extrémité 3' pour permettre un *run-off* transcriptionnel après la synthèse des 604 nucléotides.

### I.1/ Linéarisation du plasmide pNL4-3 clone 159

Cent µg de plasmide sont digérés par 100 U de *PvuII* en présence de 10 µL de tampon Green 10x (10 mM Tris-HCl (pH 7,5 à 37°C) ; 10 mM MgCl<sub>2</sub> ; 50 mM NaCl ; 0,1 mg/mL BSA) (Thermo Scientific) dans un volume final de 100 µL. Le mélange réactionnel est incubé 1 h à 37°C et les fragments de plasmides obtenus sont purifiés par une extraction (vol/vol) de phénol-chloroforme ADN (Roti®-Phenol/Chloroform/isoamyl alcohol 25 :24 :1) puis précipités à l'éthanol. Brièvement, les acides nucléiques sont précipités par l'ajout de 0,1 volume d'acétate de sodium 3 M (pH 5), 3 volumes d'éthanol 100% et 0,5 µL de GlycoBLUE® et incubation 30 min à – 80°C. Après une centrifugation de 30 min à 21 000 g à 4°C, le culot d'acides nucléiques est lavé deux fois à l'éthanol 70% froid suivi d'une centrifugation de 5 min à 21 000 g et 4°C. Le culot est séché à l'air libre et à température ambiante avant d'être repris dans un volume d'eau MQ. Un aliquot est prélevé et déposé sur un gel d'agarose 0,8%, TBE 1x supplémenté de Bet pour vérifier la digestion complète du plasmide pNL4-3 clone 159.

### 1.2/ Transcription *in vitro* de l'ARNg 1-604

L'étape de transcription *in vitro* pour la synthèse des ARN a été faite selon deux protocoles. Le premier est décrit dans l'article Gilmer *et al.*, 2022 en suivant le protocole du kit MEGAscript® (ThermoFisher) (72). La seconde approche est décrite brièvement plus bas.

La transcription *in vitro* est réalisée à l'aide d'une ARN polymérase ADN dépendante produite et purifiée au laboratoire. Ainsi, 25 µg d'ADN linéarisés sont mélangés à un milieu réactionnel de 300 µL contenant 30 µL de tampon T7 10x (400 mM de Tris-HCl (pH 8,0 à 37°C), 150 mM de MgCl<sub>2</sub>, 500 mM de NaCl), 4 mM de chaque NTPs, 0,01% de Triton X-100 (Sigma), 1 mM de spermidine (Sigma), 5 mM de DTT (Sigma), 160 U de RNasine® (Promega), 0,05 mg/mL de BSA, 1 µg/mL de pyrophosphatase (Roche) et 3 µL d'ARN polymérase T7. Après 3 h de transcription à 37°C, les échantillons sont traités avec 50 µL d'un mélange comprenant 35 U de DNase I (Sigma) et 35 µL de tampon de DNase I (400 mM Tris-HCl (pH 8,0), 100 mM NaCl, 60 mM MgCl<sub>2</sub>, 10 mM CaCl<sub>2</sub>). La réaction de dégradation de la matrice est réalisée à 37°C pendant 1 h avant d'être stoppée par l'ajout de 150 µL d'EDTA à 100 mM. Les ARN sont purifiés par extraction phénol-chloroforme (vol/vol) et précipité à l'éthanol avant d'être repris dans 500 µL H<sub>2</sub>O MQ. La qualité des ARN est contrôlée sur un gel dénaturant à 8% de polyacrylamide et 8 M d'urée.

### 1.3/ Purification de l'ARN sur colonne TSK G4000

Après la précipitation à l'éthanol, les ARN produits sont purifiés par une approche de chromatographie d'exclusion de taille sur une colonne TSKgel® G4000SW (Tosoh Bioscience) avec un système ÄKTApurifier (GE Healthcare). L'ensemble des solutions utilisées sont filtrées sur des membranes de celluloses de 0,22 µm (GE Healthcare). La colonne est équilibrée dans 120 mL de tampon de 0,2 M d'acétate de sodium (pH 5,0) et 1% (vol/vol) d'éthanol à un débit de 1 mL/min. Les échantillons transcrits *in vitro* sont ensuite dénaturés à 90°C pendant 2 min avant d'être injectés dans la colonne puis élués à un débit de 2 mL/min. Des fractions de 1,0 mL sont collectées et celles contenant des acides nucléiques sont rassemblées en fonction des pics d'éluion détectés à 254 nm. Les acides nucléiques purifiés sont précipités à l'éthanol. Les échantillons sont repris dans 100 µL H<sub>2</sub>O MQ. La qualité de la purification est vérifiée par électrophorèse sur un gel dénaturant à 8% de polyacrylamide et 8 M d'urée.



## II/ Formation des complexes ARN-protéines

Les protéines Pr55<sup>Gag</sup> (NL4-3) étiquetées par 6 histines (6His-tag) en C-terminal ont été produites, purifiées et caractérisées par le laboratoire de Johnson Mak (Institute for glycomics, Griffith University, Queensland, Australia), selon le protocole de McKinstry *et al.*, (2014) (254). Les protéines NCp15, NCp9 et NCp7 ont été exprimées et purifiées par le laboratoire de Carine Tisné (UMR 8261 Expression génétique microbienne, CNRS, Université Paris-Diderot, Paris 7), selon les protocoles de Lee *et al.*, (1998) ; De Guzman *et al.*, (1998) et Mouhand *et al.*, (2018) (53,255,256). La qualité et la pureté de chacun des lots protéiques a été confirmé par SDS-PAGE et caractérisé par des analyses en DLS (« Dynamic Light Scattering »).

Dans un premier temps, 4 pmoles d'ARN 1-604 purifiés (0,8 µg) sont dénaturées à 90°C pendant 2 min puis refroidies sur glace pendant 2 min. L'ARN est ensuite incubé 30 min à 37°C dans un volume final de 20 µL contenant du tampon de dimérisation 1x (30 mM HEPES (pH 8,0), 300 mM KCl, 5 mM MgCl<sub>2</sub>), supplémenté de 1 µg d'ARNt totaux de levure et 8 U de RNasine® (Promega). En parallèle, jusqu'à 240 pmoles de protéines (Pr55<sup>Gag</sup>, NCp15, NCp9 et NCp7) sont préparées dans un mélange de 100 µL de tampon de repliement 1x (30 mM HEPES (pH 8,0), 5 mM MgCl<sub>2</sub>), 0,01% Triton X-100, 0,01 M de DTT, 4 µg de BSA et une concentration finale de KCl de 300 mM prenant en compte la concentration de 1 M de NaCl apportée par les solutions de protéines. Dans le cas des protéines Pr55<sup>Gag</sup>, le mélange est également supplémenté de 10% de glycérol pour empêcher l'agrégation des protéines avant d'être incubé 15 min sur glace.

Les complexes ribonucléoprotéiques sont formés en ajoutant 100 µL du mélange de protéines aux 20 µL de la solution d'ARN. Pour la condition « NoProt », le mélange est réalisé de façon similaire mais la solution de protéine est substituée par H<sub>2</sub>O MQ. Les complexes sont formés à 37°C pendant 30 min puis stabilisés par une incubation de 15 min sur glace. Les 120 µL de complexes ribonucléoprotéiques sont séparés en deux volumes équivalents pour être traités avec un réactif SHAPE (+) ou avec du DMSO, pour le contrôle négatif (-).

### III/ Modifications SHAPE à l'aide de NMIA

#### III.1/ Modification de l'ARN seul ou en complexe

Les 60 µL d'ARN 1-604 repliés seuls (condition « NoProt ») ou en complexes avec les protéines (condition « Complex ») sont traités avec 12 µL de NMIA (Sigma-Aldrich) à 20 mM dans du DMSO anhydre (Sigma-Aldrich) (+) ou avec 12 µL de DMSO anhydre (-). Après 50 min d'incubation à température ambiante, 128 µL de H<sub>2</sub>O MQ sont ajoutés au mélange pour interrompre la réaction avant de précipiter les complexes ribonucléoprotéiques à l'éthanol. Les culots sont repris dans 15 µL d'eau MQ supplémentés de 4 µL de tampon de dimérisation 5x et de 1 µL de Proteinase K (Roche). La digestion enzymatique est réalisée à 37°C pendant 30 min avant l'ajout de 80 µL d'eau MQ. Les ARN modifiés ou non sont purifiés par extraction phénol-chloroforme (vol/vol, pH 7,5) suivie d'une précipitation à l'éthanol. Les culots sont repris dans 6 µL d'eau MQ.

#### III.2/ Modification des complexes traités à l'AT-2

Dans le cas des conditions traitées à l'AT-2 (condition « AT-2 »), deux échantillons de 60 µL de complexes ribonucléoprotéiques sont supplémentés de 2 µL d'AT-2 (Sigma-Aldrich) à 30 mM dans du DMSO (Sigma-Aldrich) et incubés pendant 1 h à 37°C. L'ARN des complexes traités à l'AT-2 est ensuite modifié soit par le NMIA soit traité par du DMSO pour l'échantillon contrôle selon le protocole présenté précédemment.

#### III.3/ Modification des complexes traités à la protéinase K

Après la formation des complexes ribonucléoprotéiques décrite plus haut, les protéines sont éliminées par un traitement à la Protéinase K avant leur modification au NMIA. Pour cela, deux échantillons de 60 µL de complexes sont traités avec 1 µL de Protéinase K (Roche) et incubés à 37°C pendant 30 min. Les ARN déprotéinisés sont modifiés soit par le NMIA soit traités par du DMSO selon le protocole présenté précédemment.

## IV/ Transcription inverse des ADNc et séparation par électrophorèse capillaire

Les échantillons d'ARN modifiés au NMIA (+) ou traités par le DMSO (-) sont mélangés avec 2 µL de primer AS 1 (**Tableau 2**) ou de primer AS 2 à 1 mM (**Tableau 2**) couplés au fluorophore Vic (Life Technologies SAS, France). Le mélange est dénaturé à 90°C pendant 2 min puis refroidi sur glace pendant 2 min. Après l'ajout de 2 µL de tampon RT AMV 5x (125 mM Tris-HCl pH 8,3 ; 250 mM KCl ; 10 mM DTT ; 25 mM MgCl<sub>2</sub>) et 10 min d'incubation à température ambiante, la réaction de transcription inverse est débutée par l'ajout de 2 µL de tampon RT AMV 5x, 6 µL d'un mélange des 4 dNTP à 2,5 mM (Invitrogen), 2 U de RT AMV (Life Sciences) dans un volume final de 20 µL. L'élongation est réalisée par une incubation à 42°C pendant 20 min, suivie de 30 min à 50°C. L'enzyme est ensuite inactivée thermiquement à 60°C pendant 10 min.

Simultanément à la transcription inverse des échantillons (+) et (-), deux réactions de séquençage sont réalisées à l'aide de 2 pmoles d'ARN non modifié et de 2 µL de primer AS 1 ou AS 2 à 2 mM, couplés au fluorophore Ned (Life Technologies SAS, France). La transcription inverse des réactions de séquençage suit la même procédure que celle des ARN (+) et (-) à l'exception de l'ajout d'un mélange réactionnel composé de 2 µL de tampon RT AMV 5x, 6 µL de G10 (0,25 mM dGTP, 1 mM dCTP, 1 mM dATP, 1 mM dTTP), 2 µL de ddGTP à 100 µM et 2 U de RT AMV (Life Sciences).

A la fin de l'inactivation thermique des échantillons (+), (-) et des 2 réactions de séquençage, les volumes réactionnels sont ajustés à 100 µL avec de H<sub>2</sub>O MQ et les ADNc néosynthétisés sont purifiés par extraction au phénol-chloroforme (vol/vol, pH 7,5). Pour chaque expérience, les ADNc issus des ARN modifiés (+) et non modifiés (-) sont chacun rassemblés avec une réaction de séquençage avant d'être précipités à l'éthanol. Les ADNc sont ensuite repris dans 10 µL de HiDi Formamide (Applied Biosystem), dénaturés 5 min à 90 ° avant d'être refroidis 5 min sur glace. Une courte étape de centrifugation (5 min) à 6 000 g est réalisée avant de déposer les 10 µL d'échantillons sur une plaque 96 puits. Les plaques sont chargées sur un appareil d'électrophorèse capillaire (ABI 3130XL Genetic Analyser (Applied Biosystem)) équipé d'un faisceau de 16 capillaires de 50 cm. Les électrophérogrammes obtenus sont analysés à l'aide du logiciel QuSHAPE (257).

AXE 1 – MATERIEL & METHODES

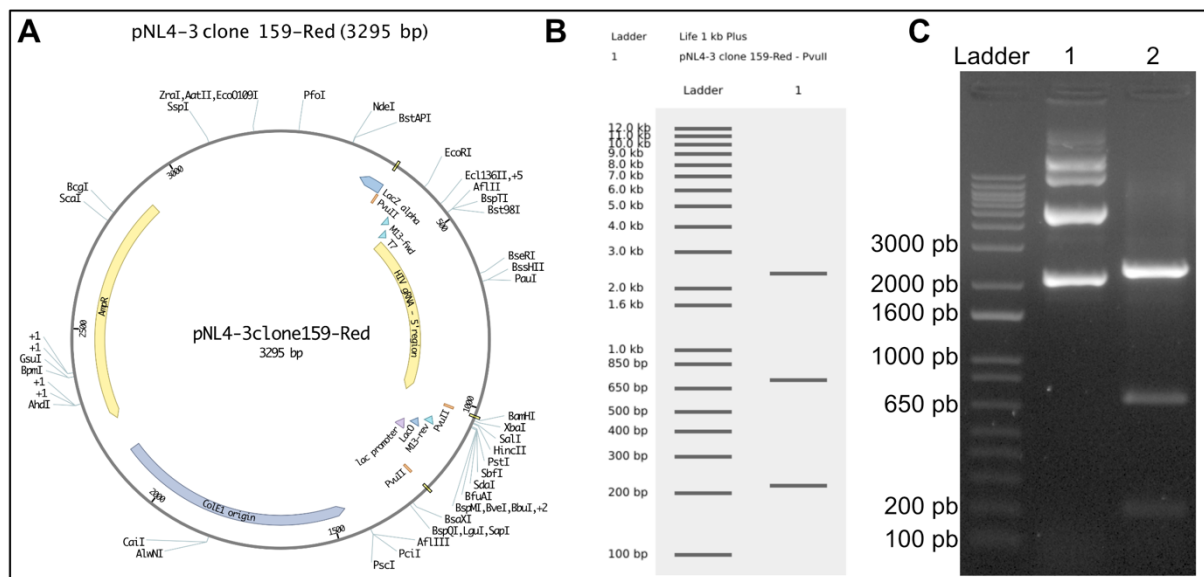
<b>Amorce</b>	<b>Séquence</b>	<b>Région d'hybridation</b>	<b>Modification</b>
<b>AS1 – Vic</b>	5'- dAGC TCC CTG CTT GCC CAT ACT A-3'	436 - 457	5' Vic
<b>AS2 - Vic</b>	5'- dCTT CTG ATC CTG TCT GAA GG -3'	536 - 555	5' Vic
<b>AS1 - Ned</b>	5'- dAGC TCC CTG CTT GCC CAT ACT A-3'	436 - 457	5' Ned
<b>AS2 - Ned</b>	5'- dCTT CTG ATC CTG TCT GAA GG -3'	536 - 555	5' Ned

**Tableau 2** : Séquences des amorces utilisées en hSHAPE.

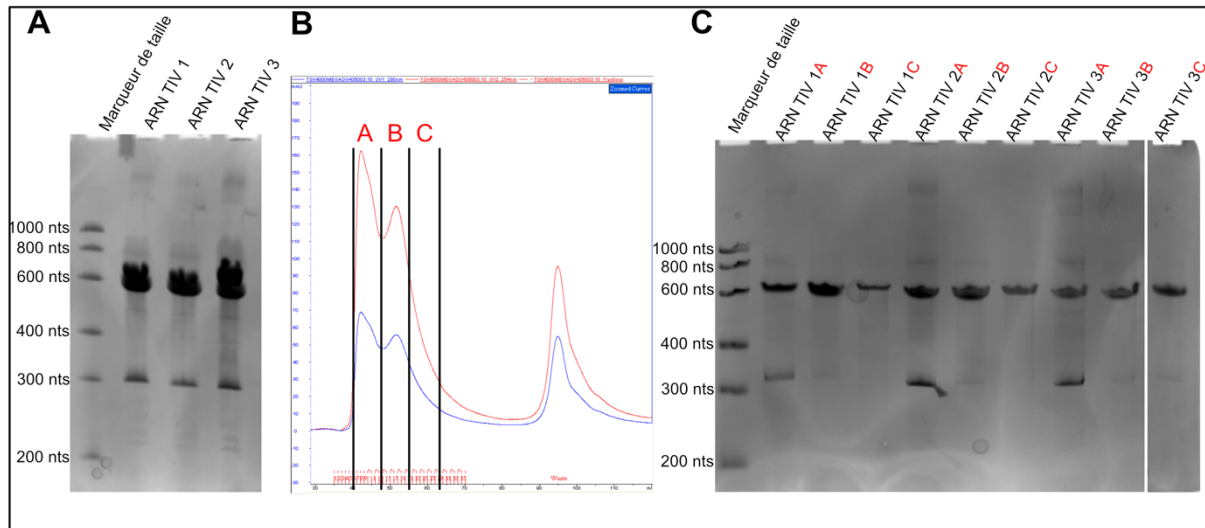
# Résultats

## I/ Production et purification de l'ARN 1-604

Dans le but d'étudier la structure de la région 5' terminale de l'ARNg du VIH-1, il a d'abord été nécessaire de produire et purifier cet ARN. Pour ce faire, le plasmide pNL4-3 clone 159 contenant la séquence des 604 premiers nucléotides de l'ARNg du VIH-1 (**Figure 22A**) a été digéré par l'enzyme de restriction *PvuII*. Ce plasmide contient 3 sites de restriction reconnus par l'enzyme. Le profil théorique de la digestion (**Figure 22B**) est comparable à celui obtenu expérimentalement (**Figure 22C**). Le fragment intermédiaire contient la séquence d'intérêt sous contrôle du promoteur T7 où la synthèse est interrompue par un *run-off* transcriptionnel (**Figure 23A**). Les ARN produits *in vitro* sont ensuite purifiés par une étape de chromatographie d'exclusion de taille (**Figure 23B**) et la qualité des transcrits purifiés est analysée (**Figure 23C**). La comparaison des deux gels (**Figures 23A et 23C**) permet de mettre en évidence l'amélioration de la qualité ainsi que la pureté des ARN synthétisés.



**Figure 22** : Représentation de la carte plasmidique de pNL4-3 clone 159 et des profils de digestion avec *PvuII*. (A) La carte présente le plasmide pNL4-3 clone 159 annoté servant à la synthèse de l'ARN contenant les 604 premiers nucléotides du VIH-1. (B) La digestion virtuelle montre le profil attendu pour une digestion de pNL4-3 clone 159 à l'aide de l'enzyme *PvuII* et (C) le gel montre le profil obtenu expérimentalement [1] avant et [2] après digestion avec *PvuII*. La carte ainsi que la digestion virtuelle ont été représentées sur Benchling.



**Figure 23** : Purification des ARN transcrits *in vitro* sur colonne d'exclusion de taille. Les 3 produits de transcription *in vitro* sont contrôlés par électrophorèses sur gels de polyacrylamide dénaturant (A) avant et (C) après purification. Lors de la purification, chaque produit est séparé en 3 groupes de fractions (A, B ou C) selon (B) le profil d'élution sur la colonne TSKgel G4000SW.

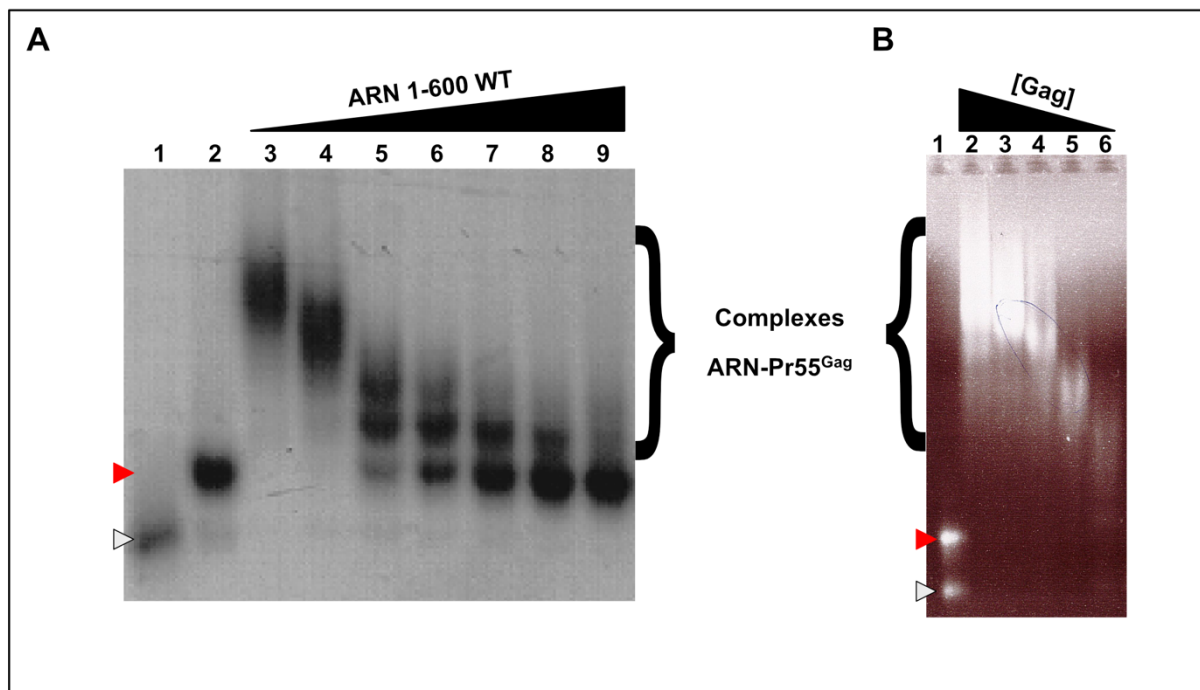
## II/ Optimisation du protocole hSHAPE

### II.1/ Protocole initial

#### II.1.1/ Formation des complexes ARN-Pr55<sup>Gag</sup>

Le laboratoire s'intéresse depuis longtemps à l'ARNg du VIH-1 et aux interactions avec des protéines qui se produisent lors de son cycle viral (49,60,118,258,259). Ainsi, ces études *in vitro* ont nécessité la mise en place de divers protocoles permettant la formation des complexes ribonucléoprotéiques, dont l'un pour l'étude de la région 5' terminale de l'ARNg en présence de la protéine Pr55<sup>Gag</sup> (118) (**Figure 24A**). Ce protocole qui utilise un tampon de cacodylate de sodium à pH 7,5, avec 0,1  $\mu$ M d'ARN et 0 - 6  $\mu$ M de Pr55<sup>Gag</sup> a donc été employé en premier lieu pour réaliser mes travaux et les résultats obtenus ont été comparés à ceux déjà publiés (**Figure 24B**). Ainsi, les résultats précédemment obtenus par le laboratoire (**Figure 24A**) montrent que l'ARNg peut adopter des structures monomérique et dimérique en fonction des conditions dans lesquelles il est replié. De plus, les conditions expérimentales utilisées sont favorables à la formation de complexes ribonucléoprotéiques. En effet, Pr55<sup>Gag</sup> reconnaît et se fixe à la région 5' terminale de l'ARNg comme le montre les résultats de retard de migration sur gel lors d'expériences de compétition entre un ARN radiomarqué et un ARN froid (**Figure 24A**). Ces observations ont également pu être réalisées dans des conditions

légèrement différentes (**Figure 24B**), lors de la formation directe des complexes avec des concentrations en ARN et en protéines supérieures, et en absence d'ARN compétiteur. Il est intéressant de noter que le retard de migration induit par la fixation des protéines Pr55<sup>Gag</sup> est dépendant du ratio ARN / protéines. Ces résultats indiquent que la protéine Pr55<sup>Gag</sup> reconnaît différents sites de fixation sur l'ARNg (49). Sur la **Figure 24B**, le retard de migration semble atteindre un plateau sur les pistes 2 et 3. Dans ces conditions, le ratio ARN:protéines est de 1:50, et 1:37,5, soit une protéine Pr55<sup>Gag</sup> pour 12 ou 16 nucléotides, respectivement. Ces conditions s'approchent de la stœchiométrie trouvée dans les particules virales immatures, où un dimère d'ARNg est encapsidé avec approximativement 2000 protéines Pr55<sup>Gag</sup> – Pr160<sup>Gag-Pol</sup>, soit en moyenne une protéine pour 9-10 nucléotides et il ne semble donc pas nécessaire d'excéder le ratio de 1:60 pour nos expériences.



**Figure 24** : Vérification de la formation des complexes ribonucléoprotéiques entre l'ARN et Pr55<sup>Gag</sup> par retard sur gel d'agarose, TBM 0,5x à 4°C. (A) Le premier gel est adapté de Abd El-Wahab *et al.*, (2014) (118). Trois nM d'ARN 1-600 radiomarqué à l'ATP [<sup>32</sup>P] sont incubés avec 100 nM de Pr55<sup>Gag</sup> et des quantités croissantes d'ARN 1-600 non marqué (piste 3 à 9). Les pistes 1 et 2 correspondent à l'ARN 1-600 radiomarqué en condition dénaturante ou renaturante, respectivement. La concentration en ARN compétiteur sur les pistes 3 à 9 est 0 nM, 10 nM, 50 nM, 100 nM, 200 nM, 400 nM et 800 nM, respectivement. (B) Le second gel a été réalisé avec de l'ARN 1-604 non marqué et mis en présence de concentration décroissante de Pr55<sup>Gag</sup>. La piste 1 correspond à 0,066 μM d'ARN 1-604 en condition favorisant la dimérisation. Les pistes 2 à 6 correspondent à l'ARN 1-604 en présence de 3,3 μM, 2,5 μM, 1,67 μM, 0,83 μM et 0,33 μM de Pr55<sup>Gag</sup>, respectivement. Les conformations monomériques et dimériques sont indiquées par les flèches grises et rouges, respectivement.

II.1.2/ Cartographie chimique *in vitro* : hSHAPE

En suivant les protocoles de formation de complexes et de cartographie chimique *in vitro* publiés par le laboratoire, j'ai étudié l'effet de la concentration en protéine sur la structure des ARN 1-604. Pour ce faire, 2 pmoles d'ARN ont été mises en présence de concentrations croissantes de protéines Pr55<sup>Gag</sup> allant de 10 pmoles (5 Gag / ARN) à 120 pmoles (60 Gag / ARN). Une condition en absence de protéine (NoProt) a également été réalisée afin de servir d'élément de comparaison entre les différentes conditions en complexe avec les protéines. L'analyse des résultats par le logiciel QuSHAPE (257) a permis d'obtenir les réactivités SHAPE pour chaque condition, au moins en triplicata. Ces dernières ont été comparées 2 à 2 et les réplicas dont le coefficient de corrélation était supérieur à 0,70 ont été conservés et moyennés. Dans le cas des conditions « 5 Gag / ARN » et « 20 Gag / ARN », la variabilité entre les réplicas a été jugée trop importante du fait d'au moins un des coefficients de corrélation inférieur à 0,65. Ces conditions ont donc été exclues de l'analyse pour éviter de la fausser. Les moyennes de réactivités SHAPE des conditions restantes (NoProt, 10 Gag, 40 Gag et 60 Gag) ont ensuite été comparées entre elles pour mettre en évidence l'impact de l'augmentation de la concentration en protéine Pr55<sup>Gag</sup> sur la structure de l'ARN (**Tableau 3** et **Annexe 1**).

Condition	NoProt	10 Gag	40 Gag	60 Gag
NoProt	1	0,915	0,686	0,860
10 Gag		1	0,784	0,925
40 Gag			1	0,844
60 Gag				1

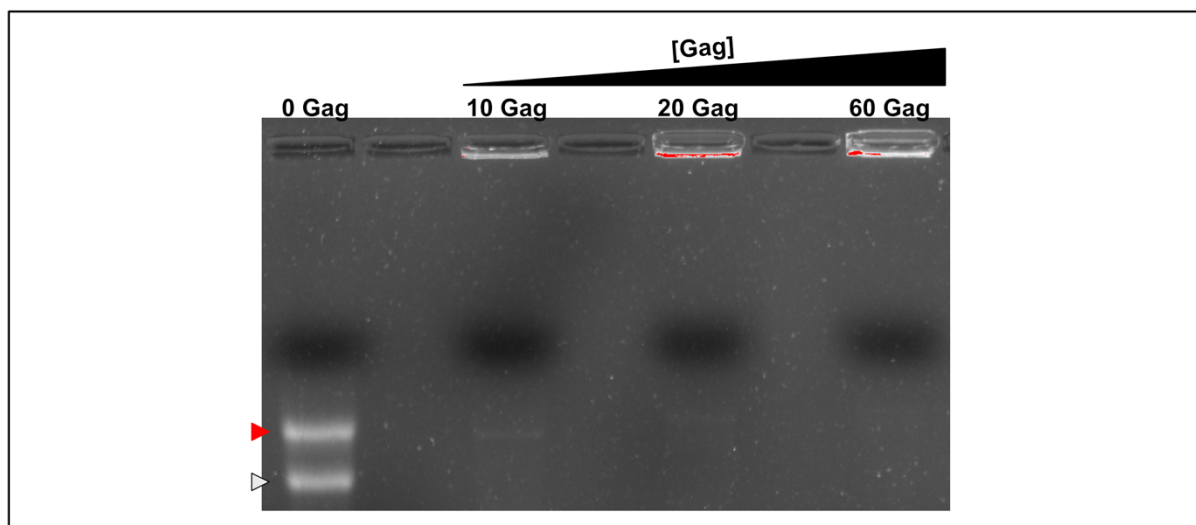
**Tableau 3** : Coefficients de corrélation entre les moyennes de réactivités SHAPE des conditions en absence (NoProt) et présence (10 Gag, 40 Gag et 60 Gag) de la protéine Pr55<sup>Gag</sup>. Les coefficients de corrélation supérieurs à 0,9, compris entre 0,9 et 0,8 et inférieurs à 0,8 sont représentés en vert, jaune et rouge, respectivement.

D'après les résultats obtenus lors des expériences de retard sur gel (**Figure 24B**), les complexes ribonucléoprotéiques formés dans les conditions « 40 Gag / ARN » et « 60 Gag / ARN » devraient être relativement semblables. Cette supposition se vérifie par un coefficient de corrélation de 0,844 entre les deux conditions (**Tableau 3**). Cependant, la condition « 60 Gag » corrèle plus fortement avec les conditions « NoProt » et « 10 Gag » que la condition « 40 Gag ». Cette inconsistance dans les résultats obtenus se confirme en



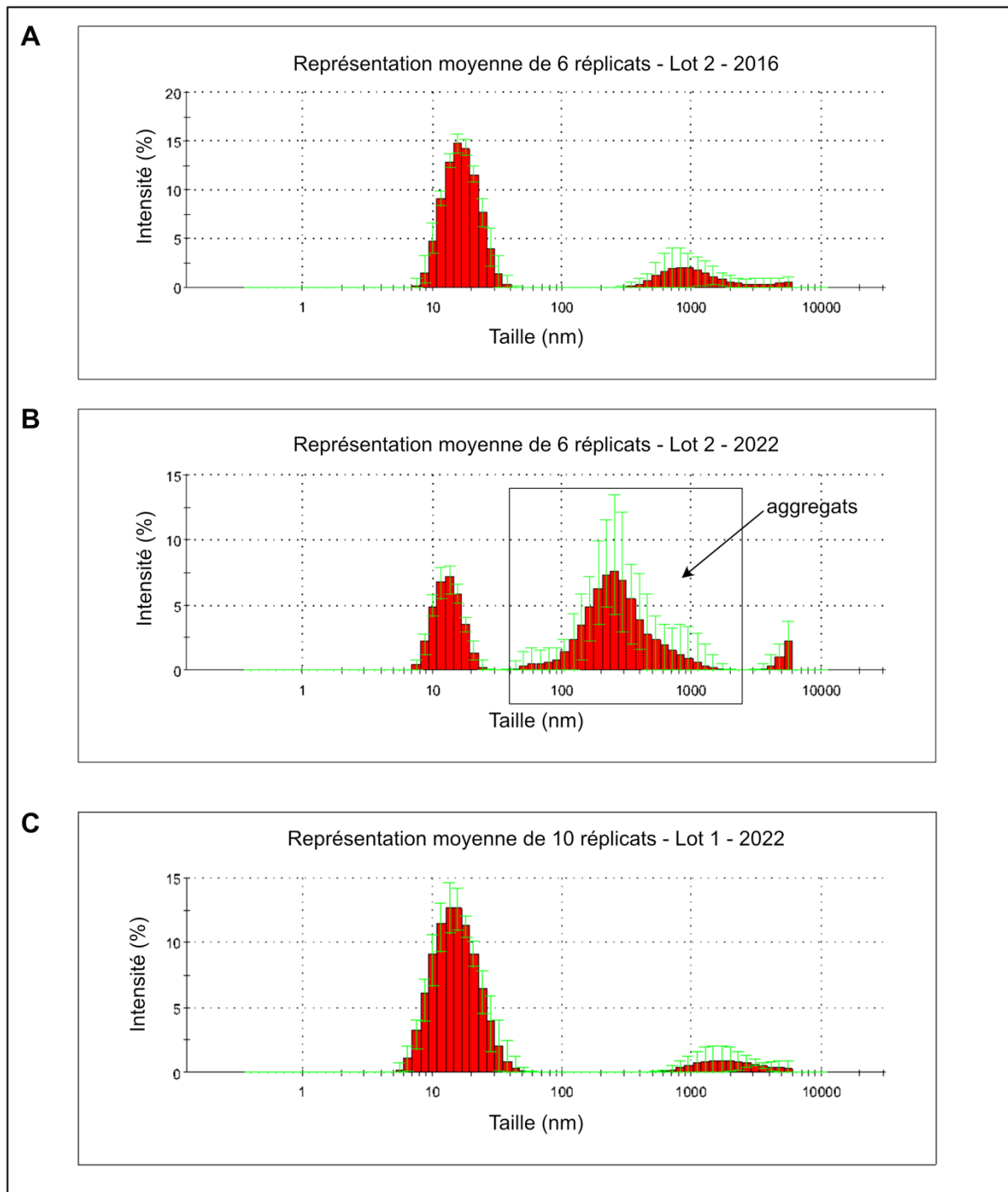
## AXE 1 – RESULTATS

comparant les réactivités une à une (**Annexe 1**). En effet, les différences de réactivités de la condition « 40 Gag » par rapport au contrôle « NoProt » sont régulièrement plus marquées que la condition « 60 Gag ». De plus, en comparant nos résultats avec ceux précédemment obtenus par notre laboratoire (118), nous pouvons noter que seule la condition « 40 Gag » permet d'identifier les sites de protection situés aux nucléotides 240-242, 247, 271-273 et 317-320 (**Annexe 1**) alors que l'on s'attendrait à retrouver ces sites de protection également pour la condition « 60 Gag ». Ces résultats semblent pointer vers un problème dans la formation des complexes ribonucléoprotéiques. Cette hypothèse a été confirmée sur gel (**Figure 25**) et en DLS (**Figure 26**).



**Figure 25** : Vérification de la qualité des complexes ribonucléoprotéiques formés avec le nouveau lot de Pr55<sup>Gag</sup>. Les ARN sont repliés dans un tampon cacodylate de sodium (pH 7,5). Les flèches grise et rouge indiquent les conformations monomérique et dimérique de l'ARNg, respectivement.

Une différence majeure entre le début des études de formation des complexes ribonucléoprotéiques et les résultats finaux de hSHAPE utilisant différentes concentrations de Pr55<sup>Gag</sup> a été l'utilisation d'un nouveau lot de protéines. Ce lot qui avait été contrôlé lors de sa réception au laboratoire (**Figure 26A**) s'est révélé bien moins stable que le lot précédent (**Figure 26C**) et une tendance accrue à l'agrégation a été observée (**Figure 26B**). Ainsi, dans des conditions identiques, des agrégats protéiques englobant les acides nucléiques étaient formés avec le deuxième lot (**Figure 26**) à l'instar des complexes ribonucléoprotéiques observés précédemment avec le premier lot (**Figure 24B**).



**Figure 26** : Vérification de l'agrégation des protéines Pr55<sup>Gag</sup> par une approche de DLS. Les profils présentés ici récapitulent les résultats de multiples mesures indépendantes pour (A et B) le nouveau et (C) l'ancien lot. Les mesures ont été prises (A) lors de la réception la protéine en 2016 ou après 6 ans à -80°C (B et C).

### II.2/ Mise au point

#### II.2.1/ Formation des complexes

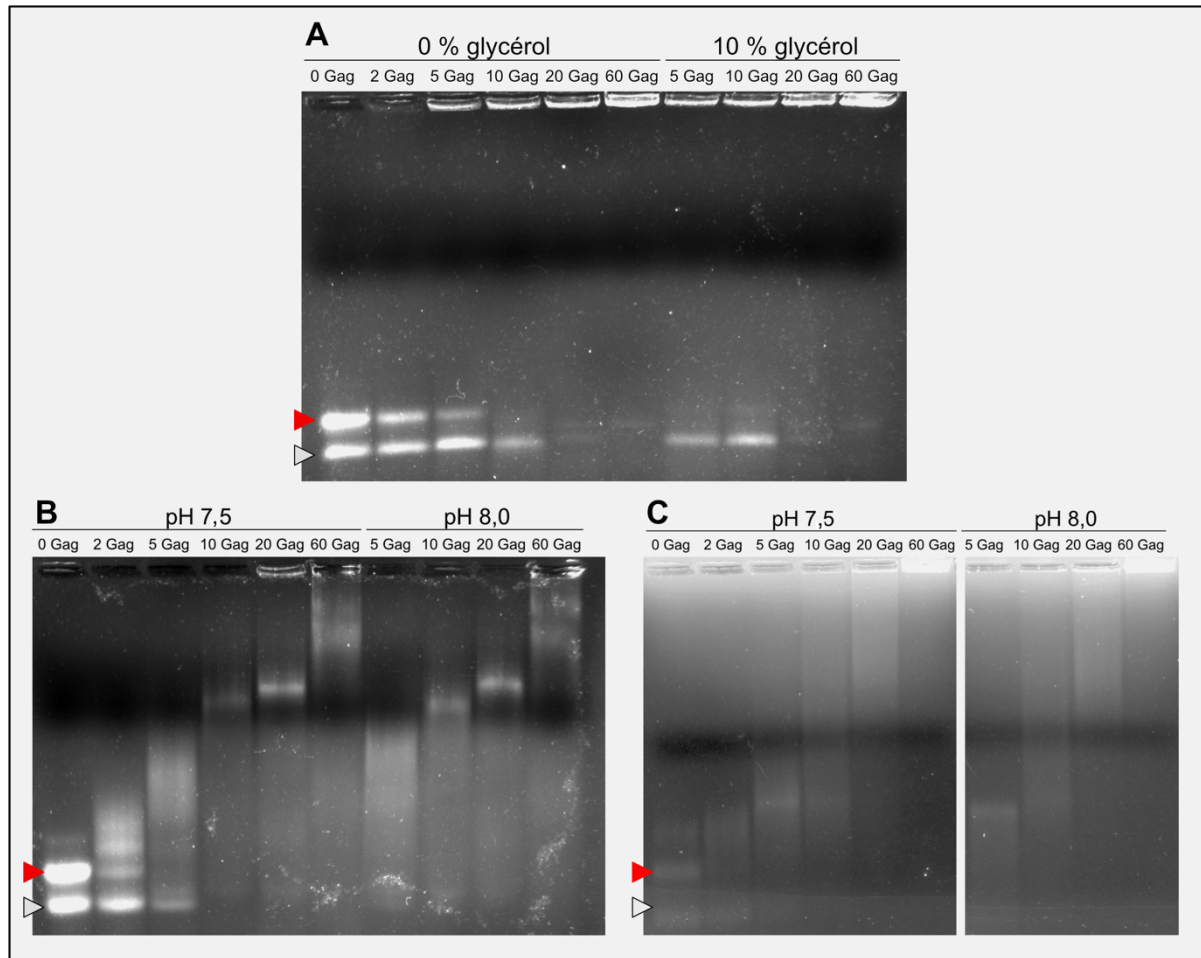
A la suite de l'obtention de ces résultats, il a été nécessaire de revoir les conditions de formation des complexes ribonucléoprotéiques. Pour ce faire, différents paramètres pouvant affecter leur formation ont été modifiés et leur impact a été évalué. Ces paramètres sont :

- Le tampon utilisé : Cacodylate de sodium, Tris-HCl ou HEPES
- Le pH : 7,2, 7,5 ou 8,0
- La température d'incubation : 25°C ou 37°C
- La concentration en glycérol : 0 à 10%
- La concentration en ARN : 33 à 133 nM

Dans un premier temps, les protéines agrégées ont été éliminées par centrifugation à 20 000 g pendant 1 h à 4°C. L'effet de chaque paramètre sur la formation des complexes a ensuite été analysé par des expériences de retard sur gel. Les résultats obtenus ont permis de conclure que les paramètres favorisant la formation correcte des complexes étaient indépendants de la concentration en ARN à cette échelle. Dans le cas des molécules tampons utilisées, le cacodylate de sodium (NaCaco) conduit systématiquement à la formation d'agrégats ribonucléoprotéiques, même en présence de 10% de glycérol (**Figure 27A**). Cependant, l'utilisation de Tris-HCl ou d'HEPES, en présence de 10% de glycérol, permet de retrouver le profil précédemment obtenu en cas de formation de complexe (**Figure 27B et 27C**). Ainsi, les paramètres permettant de former des complexes ribonucléoprotéiques non agrégés sont :

- Tris-HCl ou HEPES
- pH 7,5 ou 8,0
- Incubation à 37°C
- 10% de glycérol

Finalement, l'utilisation du Tris-HCl comme tampon n'a pas été retenue à cause des variations de pH importantes induites par la température ( $\Delta 33^\circ\text{C} \Leftrightarrow 1$  unité de pH). De plus, le choix de l'HEPES a permis de garder une cohérence entre les expériences de hSHAPE présentées ici et les expériences de cartographie tridimensionnelle que je présenterai dans l'axe 2 et qui utilisent également un tampon à base d'HEPES à pH 8,0 (260).



**Figure 27** : Expériences de formation de complexes entre l'ARNg 1-604 et les protéines Pr55<sup>Gag</sup>. Les complexes sont formés avec 0, 2, 5, 10, 20 ou 60 protéines par ARN. (A) La formation de complexes est réalisée dans un tampon au cacodylate de sodium (pH 7,5), avec 0% ou 10% de glycérol, ou bien dans un tampon (B) Tris ou (C) HEPES, à pH 7,5 ou 8,0, avec 10% de glycérol. Les conformations monomériques et dimériques sont indiquées par les flèches grises et rouges, respectivement.

## II.2.2/ Validation des conditions expérimentales sur l'ARN nu

Dans un second temps, j'ai mesuré l'effet de ces changements de conditions sur la structure de l'ARN 1-604 en absence de protéine. Pour ce faire, j'ai réalisé le repliement de l'ARN 1-604 dans un tampon HEPES à pH 7,5 ou 8,0, en présence ou non de 10% de glycérol. J'ai également testé l'efficacité de la sonde SHAPE utilisée pour la modification des ARN. Pour ce faire, les ARN repliés dans les différents tampons ont été modifiés à l'aide de deux sondes SHAPE, le NMIA (261) ou le BzCN (262). Les réactivités obtenues pour chaque condition ont été comparées 2 à 2 et les résultats sont présentés dans le **Tableau 4**.

AXE 1 – RESULTATS

A			0% Glycérol				10% Glycérol			
			NMIA		BzCN		NMIA		BzCN	
			HEPES pH 7,5	HEPES pH 8,0	HEPES pH 7,5	HEPES pH 8,0	HEPES pH 7,5	HEPES pH 8,0	HEPES pH 7,5	HEPES pH 8,0
			0,8 - 1		0,6 - 0,8		≤ 0,6			
0% Glycérol	NMIA	HEPES pH 7,5	1	0,8	0,42	0,6	0,92	0,86	0,45	0,52
		HEPES pH 8,0		1	0,66	0,88	0,71	0,96	0,69	0,79
	BzCN	HEPES pH 7,5			1	0,82	0,36	0,57	0,84	0,84
		HEPES pH 8,0				1	0,54	0,82	0,86	0,92
10% Glycérol	NMIA	HEPES pH 7,5				1	0,8	0,38	0,44	
		HEPES pH 8,0					1	0,62	0,72	
	BzCN	HEPES pH 7,5						1	0,96	
		HEPES pH 8,0							1	

B			0% Glycérol				10% Glycérol			
			NMIA		BzCN		NMIA		BzCN	
			HEPES pH 7,5	HEPES pH 8,0	HEPES pH 7,5	HEPES pH 8,0	HEPES pH 7,5	HEPES pH 8,0	HEPES pH 7,5	HEPES pH 8,0
			p-value ≤ 0,05							
0% Glycérol	NMIA	HEPES pH 7,5	1	0,2144	0,7004	0,7222	0,00775	0,5488	0,6687	0,3133
		HEPES pH 8,0		1	0,3044	0,2448	7,95E-05	0,457	0,06411	0,7809
	BzCN	HEPES pH 7,5			1	0,9024	0,002514	0,7741	0,406	0,4709
		HEPES pH 8,0				1	0,001858	0,7193	0,4568	0,3782
10% Glycérol	NMIA	HEPES pH 7,5				1	0,001079	0,0237	0,0001728	
		HEPES pH 8,0					1	0,2617	0,6533	
	BzCN	HEPES pH 7,5						1	0,1184	
		HEPES pH 8,0							1	

**Tableau 4** : Comparaison 2 à 2 des données de réactivités pour les conditions avec ou sans glycérol, repliées dans un tampon HEPES (pH 7,5 ou 8,0) et modifiées au NMIA ou au BzCN. (A) Les coefficients de corrélation entre chaque condition sont rapportés. Les cases encadrées en noir ou en rouge correspondent aux corrélations entre les pH 7,5 et 8,0, ou à celles entre 0 et 10% de glycérol, respectivement. (B) Les p-values obtenues après un test de Wilcoxon bilatéral sont rapportées. Les cases rouges encadrées correspondent aux p-values significatives.

Les coefficients de corrélation entre les conditions 0% et 10% de glycérol (**Tableau 4A** – encadrés rouges) sont relativement proches de 1, quel que soit le tampon ou l’agent modifiant utilisé, ce qui semblerait indiquer une absence d’effet du glycérol sur la distribution des réactivités. De même, les coefficients de corrélation entre les conditions HEPES pH 7,5 et pH 8,0 (**Tableau 4A** - encadrés noirs) semblent indiquer une absence d’effet du pH sur les profils de réactivités obtenus. Afin de confirmer ces observations, des tests de Wilcoxon bilatéraux ont été réalisés sur R. L’hypothèse nulle formulée par le test de Wilcoxon bilatéral est  $H_0 : \theta = 0$ . Ainsi, si la p-value obtenue par ce test est  $< 0,05$ , l’hypothèse  $H_0$  peut être rejetée avec au moins 95% de confiance, ce qui implique que les deux jeux de données testés ont des

## AXE 1 – RESULTATS

médianes significativement différentes et donc des distributions différentes. Le **Tableau 4B** montre les p-values obtenues à partir de ces tests. Seule la condition HEPES pH 7,5 modifié au NMIA en présence de 10% de glycérol est capable de rejeter  $H_0$ , et serait donc statistiquement différente des autres conditions testées. Cependant, j'ai décidé de ne pas traiter cette condition comme un résultat aberrant. En effet, bien que les valeurs de réactivités de cette condition soient distinctes des autres, les coefficients de corrélation (**Tableau 4A**) montrent que le profil obtenu reste cohérent avec les autres conditions où les ARN sont modifiés au NMIA. Globalement, les résultats obtenus tendent à montrer que la présence de glycérol ainsi que la variation de pH de 7,5 à 8,0 sont des paramètres qui n'affectent pas les réactivités obtenues. Cependant, le réactif de modification utilisé (NMIA vs BzCN) a un impact considérable sur les réactivités obtenues, comme le montre les coefficients de corrélation relativement bas (**Tableau 4A**). Ces différences peuvent être expliquées par l'extrême réactivité du BzCN par rapport au NMIA. En effet, le BzCN a un temps de demi-vie de 0,25 s à 37°C (262), ce qui est 2 400 fois plus rapide que le NMIA. Ainsi, la moindre différence dans l'état de l'ARN ciblé peut être immédiatement détectée. A l'inverse, le temps de demi-vie du NMIA étant de 10 - 11 min à 37°C, les résultats de réactivités obtenus sont lissés et présentent moins de disparités entre eux. Un autre élément pouvant expliquer la variation des réactivités obtenues entre ces deux conditions est la présence d'un précipité lors de l'extraction au phénol-chloroforme réalisée à la suite de la modification au BzCN. Ainsi, il est possible qu'une partie de l'échantillon soit perdue dans la partie phénolique lors de l'extraction phénol-chloroforme. L'ensemble de ces observations m'a poussé à choisir un tampon HEPES, la sonde SHAPE NMIA pour modifier l'ARN et la présence de 10% de glycérol. L'étape suivante a donc consisté à tester ces conditions en présence de protéines pour s'assurer que le protocole optimisé était capable de mettre en évidence des différences induites par la fixation de protéines.

### II.2.3/ Validation des nouvelles conditions expérimentales sur les complexes ribonucléoprotéiques

J'ai évalué la robustesse de mon protocole en réalisant des expériences de formation de complexe suivi de hSHAPE sur les conditions « NoProt » et « 60 Gag », dans un tampon HEPES pH 7,5 ou 8,0 avec 10% de glycérol. La modification des ARN a été réalisée par du NMIA ou du BzCN, pour conserver un élément de comparaison avec le protocole publié

## AXE 1 – RESULTATS

précédemment par notre équipe. Après m’être assuré que les complexes se formaient correctement dans les conditions testées (**Figure 27B** et **27C**), j’ai réalisé les expériences hSHAPE et j’ai effectué l’analyse des comparaisons 2 à 2 de chacun de mes jeux de données (**Tableau 5A**) ainsi que les tests de Wilcoxon bilatéraux associés (**Tableau 5B**) de la même façon que précédemment.

<b>A</b>			0 Gag				60 Gag			
			NMIA		BzCN		NMIA		BzCN	
			HEPES pH 7,5	HEPES pH 8,0	HEPES pH 7,5	HEPES pH 8,0	HEPES pH 7,5	HEPES pH 8,0	HEPES pH 7,5	HEPES pH 8,0
			0,8 - 1 0,6 - 0,8 ≤ 0,6							
0 Gag	NMIA	HEPES pH 7,5	1	0,77	0,49	0,5	0,63	0,65	0,44	0,28
		HEPES pH 8,0		1	0,75	0,84	0,6	0,64	0,54	0,5
	BzCN	HEPES pH 7,5			1	0,95	0,66	0,72	0,78	0,73
		HEPES pH 8,0				1	0,56	0,67	0,69	0,67
60 Gag	NMIA	HEPES pH 7,5				1	0,93	0,86	0,78	
		HEPES pH 8,0					1	0,79	0,75	
	BzCN	HEPES pH 7,5						1	0,93	
		HEPES pH 8,0							1	

<b>B</b>			0 Gag				60 Gag			
			NMIA		BzCN		NMIA		BzCN	
			HEPES pH 7,5	HEPES pH 8,0	HEPES pH 7,5	HEPES pH 8,0	HEPES pH 7,5	HEPES pH 8,0	HEPES pH 7,5	HEPES pH 8,0
			p-value 0,05 - 0,1 ≤ 0,05							
0 Gag	NMIA	HEPES pH 7,5	1	0,003247	0,1101	0,1016	0,6835	0,05345	0,3227	0,02833
		HEPES pH 8,0		1	0,1447	0,1302	0,004764	0,4282	0,0292	0,2795
	BzCN	HEPES pH 7,5			1	0,9988	0,2389	0,5816	0,5277	0,5925
		HEPES pH 8,0				1	0,2142	0,5667	0,5166	0,5409
60 Gag	NMIA	HEPES pH 7,5				1	0,09117	0,5148	0,04847	
		HEPES pH 8,0					1	0,2611	0,9481	
	BzCN	HEPES pH 7,5						1	0,1931	
		HEPES pH 8,0							1	

**Tableau 5** : Comparaison 2 à 2 des données de réactivités pour les conditions avec ou sans Pr55<sup>Gag</sup>, repliées dans un tampon HEPES (pH 7,5 ou 8,0) et modifiées au NMIA ou au BzCN. (A) Les coefficients de corrélation entre chaque condition sont rapportés. Les cases encadrées en noires ou en rouges correspondent aux corrélations entre les pH 7,5 et 8,0, ou à celles entre 0 et 60 Pr55<sup>Gag</sup> par ARN, respectivement. La partie ombragée correspond aux comparaisons entre 0 et 60 Gag. (B) Les p-values obtenues après un test de Wilcoxon bilatéral sont rapportées. Les cases encadrées en rouge correspondent aux p-values significatives.

Premièrement, les résultats obtenus entre les conditions HEPES pH 7,5 et HEPES pH 8,0 montrent des corrélations relativement élevées (**Tableau 5A** - encadrés noirs), ce qui confirme les résultats obtenus avec l’ARN nu (**Tableau 4A** - encadrés noirs) et l’absence d’effet du pH

## AXE 1 – RESULTATS

sur les profils de réactivité. Ces observations suggèrent que les structures formées par l'ARNg 1-604 sont similaires dans un tampon HEPES, à pH 7,5 comme à pH 8,0.

Deuxièmement, l'ajout de protéines Pr55<sup>Gag</sup> aux ARNg 1-604 induit une modification de ces profils, comme démontré par les plus faibles corrélations entre les jeux de données « NoProt » et « 60 Gag » (**Tableau 5A** - encadrés rouges et zone ombragée). En effet, l'ajout de protéine induit une modification globale des profils de réactivité qui se traduit par une diminution de la corrélation entre les jeux de données. Dans le cas de la distribution de ces réactivités, les résultats sont plus divergents (**Tableau 5B** - encadrés rouges et zone ombragée). En effet, les tests de Wilcoxon bilatéraux indiquent une différence significative dans quasiment 25% des cas, ce qui ne permet pas de conclure sur l'effet de la fixation de la protéine bien qu'une tendance semble se dégager. Deux cas de figure sont donc envisageables. Soit la fixation de la protéine modifie la distribution des réactivités et les tests de Wilcoxon bilatéraux ne sont pas assez puissants pour détecter ces différences, soit la fixation n'induit pas de modification et les différences détectées sont artéfactuelles. D'après les résultats précédemment obtenus et publiés (118), l'hypothèse la plus probable est celle impliquant la faible puissance des tests statistiques réalisés.

Troisièmement, une différence persiste entre les résultats obtenus à partir des ARN modifiés au NMIA et ceux au BzCN (**Tableau 5A**), bien qu'elle soit majoritairement retrouvée dans les conditions en absence de protéines. Pour diminuer les risques de variation de réactivités obtenues d'un réplica à l'autre, ou même au sein d'un même groupe d'expérience, j'ai décidé d'utiliser le réactif NMIA et non le BzCN. De plus, j'ai choisi de tamponner mes solutions à pH 8,0 plutôt qu'à pH 7,5 car la réaction entre l'hydroxyle 2' de l'ARN et le réactif SHAPE est favorisée lorsque le pH augmente.

En conclusion, ces différentes expériences m'ont permis de valider un protocole optimisé de formation de complexes ribonucléoprotéiques suivi d'une approche de hSHAPE. De plus, ces conditions sont compatibles avec l'approche de MOHCA-seq (Axe 2), ce qui devrait faciliter le rapprochement des résultats obtenus par les deux techniques en éliminant la possibilité de l'effet d'une molécule tampon différente ou d'un pH différent. Ceci m'a permis de réaliser l'étude de la structure de l'ARNg 1-604 en présence des différentes protéines Pr55<sup>Gag</sup>, NCp15, NCp9 et NCp7. Les résultats de cette étude ont été publiés lors de ma dernière année de doctorat et constituent la prochaine partie.

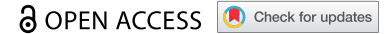


III/ Structural maturation of the HIV-1 RNA 5' untranslated region by Pr55<sup>Gag</sup> and its maturation products

Orian Gilmer, Elodie Mailler, Jean-Christophe Paillart, Assia Mouhand, Carine Tisé, Johnson Mak, Redmond P. Smyth, Roland Marquet & Valérie Vivet-Boudou

DOI: 10.1080/15476286.2021.2021677

RESEARCH PAPER



## Structural maturation of the HIV-1 RNA 5' untranslated region by Pr55<sup>Gag</sup> and its maturation products

Orian Gilmer<sup>a,†</sup>, Elodie Mailler<sup>a,†,\*</sup>, Jean-Christophe Paillart<sup>ib,a</sup>, Assia Mouhand<sup>b</sup>, Carine Tisé<sup>ib,b</sup>, Johnson Mak<sup>ib,c</sup>, Redmond P. Smyth<sup>a,†,\*\*</sup>, Roland Marquet<sup>a,†</sup>, and Valérie Vivet-Boudou<sup>ib,a,†</sup>

<sup>a</sup>Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, IBMC, Strasbourg, France; <sup>b</sup>Expression Génétique Microbienne, UMR 8261, CNRS, Université de Paris, Institut de Biologie Physico-chimique, Paris, France; <sup>c</sup>Institute for Glycomics, Griffith University, Gold Coast, Australia

### ABSTRACT

Maturation of the HIV-1 viral particles shortly after budding is required for infectivity. During this process, the Pr55<sup>Gag</sup> precursor undergoes a cascade of proteolytic cleavages, and whilst the structural rearrangements of the viral proteins are well understood, the concomitant maturation of the genomic RNA (gRNA) structure is unexplored, despite evidence that it is required for infectivity. To get insight into this process, we systematically analysed the interactions between Pr55<sup>Gag</sup> or its maturation products (NCp15, NCp9 and NCp7) and the 5' gRNA region and their structural consequences, *in vitro*. We show that Pr55<sup>Gag</sup> and its maturation products mostly bind at different RNA sites and with different contributions of their two zinc knuckle domains. Importantly, these proteins have different transient and permanent effects on the RNA structure, the late NCp9 and NCp7 inducing dramatic structural rearrangements. Altogether, our results reveal the distinct contributions of the different Pr55<sup>Gag</sup> maturation products on the gRNA structural maturation.

### ARTICLE HISTORY

Received 6 November 2021  
Revised 18 December 2021  
Accepted 19 December 2021



### KEYWORDS

HIV-1; genomic RNA; RNA structure; maturation; Gag; Pr55<sup>Gag</sup>; NCp15; NCp9; NCp7; RNA chaperone

## Introduction

Retroviruses, including human immunodeficiency virus type 1 (HIV-1), bud from infected cells as immature non-infectious particles [1]. Shortly after budding of HIV-1 particles, proteolytic cleavage of the Pr55<sup>Gag</sup> and Pr160<sup>GagPol</sup> precursors triggers morphological rearrangements of the immature Gag shell and lead to formation of the cone-shaped capsid characteristic of the lentivirus family [1–4]. At the same time, the genomic RNA (gRNA) rearranges from a 'loose' unstable dimer to a more stable and compact 'tight' dimer [5,6]. This process is broadly known as 'maturation' and is required for the formation of an infectious particle. The proteolytic cleavages taking place during maturation are well characterized [1,3,7,8]. Indeed, the tridimensional structures of the HIV-1 capsid (CA) in immature [9] and mature [10] virions, as well as in mutants blocked at different stages of maturation have been solved by cryo-electron microscopy [11], revealing the structural switch triggering CA maturation.

Proteolytic processing of Pr55<sup>Gag</sup> is driven by ordered, sequential cleavages at five positions by the viral protease enzyme [1,3,7,8] (Figure 1(a)). Correct proteolytic processing of Pr55<sup>Gag</sup> is necessary not only to form the mature matrix and capsid [1–3], but also for generating the mature gRNA dimer [5,6,12]. As the nucleocapsid (NC) domain exhibits potent RNA chaperone activity, interaction of Pr55<sup>Gag</sup> and its NC-containing maturation products with gRNA are expected to govern its structural rearrangements [13–15]. The primary cleavage between spacer peptide 1 (SP1) and the NC domain (Figure 1(a)) releases NCp15 inside the virion [6,12,16,17] and is associated with the initial condensation of the viral genome [18]. The secondary cleavage events, which release NCp9 (Figure 1(a)), have no further impact on gRNA compaction [19], but the late cleavage event that generates mature NCp7 (Figure 1(a)) is necessary to complete this process [20]. Intriguingly, mutant viruses blocked after the primary cleavage event contain the same level of reverse transcriptase activity as fully mature virions and are competent for virus-cell fusion, yet these viruses display a marked

**CONTACT** Roland Marquet  [r.marquet@ibmc-cnrs.unistra.fr](mailto:r.marquet@ibmc-cnrs.unistra.fr)  Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, IBMC, 2 Allée Konrad Roentgen, Strasbourg 67000, France

†These authors equally contributed to this work.

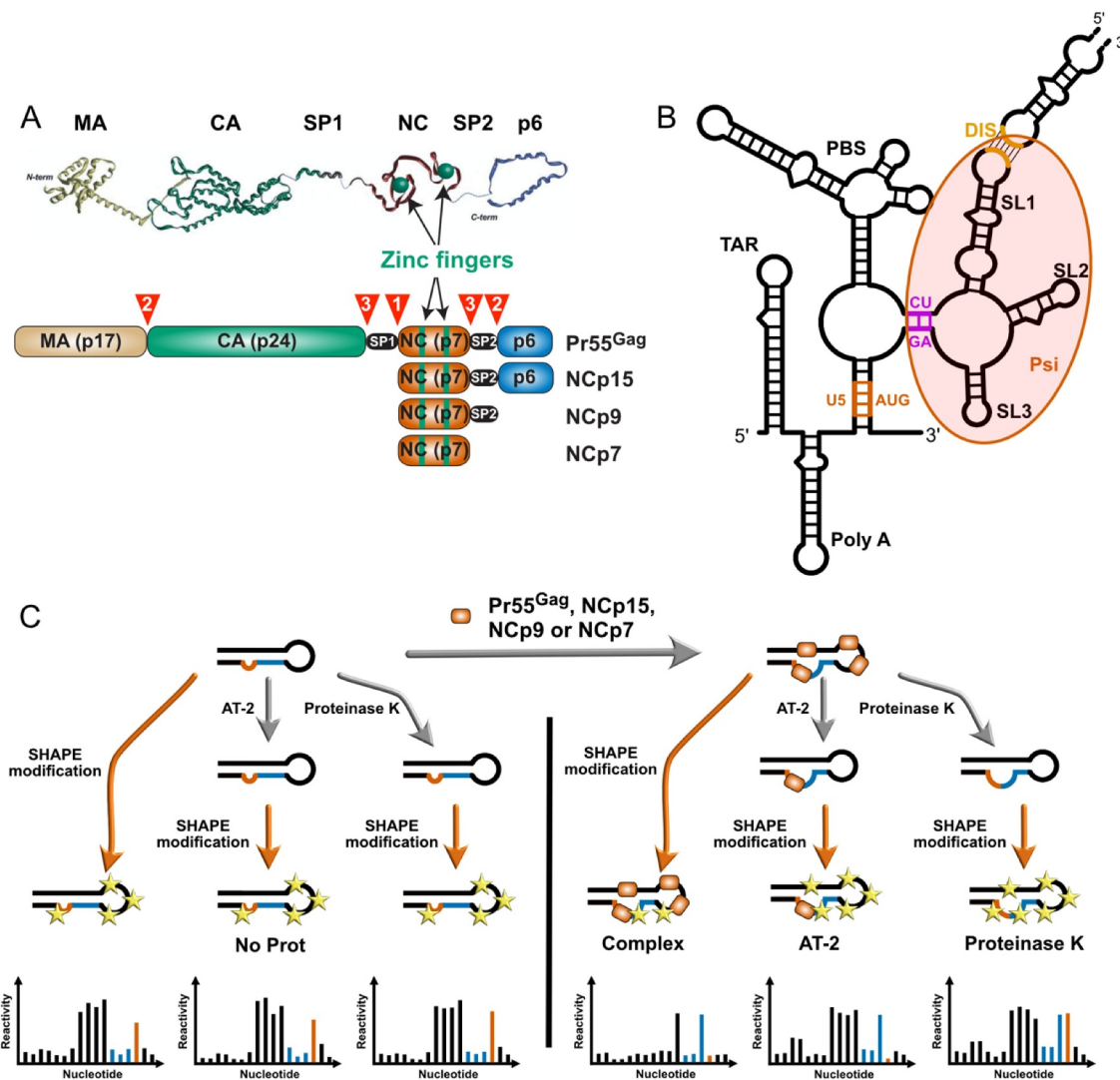
‡co-senior authors.

\*Present address: Cell Biology and Neurobiology Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, 20,892, USA

\*\*Present address: Genome Architecture and Evolution of RNA viruses, Helmholtz Institute for RNA-based Infection Research (HIRI), Joseph-Schneider-Str. 2, D15, 97,080 Würzburg, Germany

 Supplemental data for this article can be accessed [here](#)

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Players and experimental strategy. (a) The Pr55<sup>Gag</sup> precursor and its nucleocapsid-containing maturation products. Scheme of the Pr55<sup>Gag</sup>, NCp15, NCp9, and NCp7 proteins are drawn in the lower part to indicate the matrix (MA), capsid (CA), spacer peptide 1 (SP1), nucleocapsid (NC), spacer peptide 2 (SP2) and p6 domains. The sequential cleavages of the Pr55<sup>Gag</sup> precursor are indicated by numbered red arrowheads and the zinc fingers located in the NC domain are indicated in green. The 3D structures of the individual domains are shown on top of the figure and these structures are artificially linked together in a linear manner as no 3D structure of the full-length Pr55<sup>Gag</sup> is available. (b) Secondary structure of the 5' region of the HIV-1 genomic RNA (gRNA). One of the secondary structure models proposed in the literature is drawn to indicate the main elements present in this region, namely from 5' to 3': TAR, the *trans*-activating region of gRNA transcription; polyA, which contains the repressed 5' copy of the polyadenylation signal in its apical loop; U5:AUG (coloured in Orange), a proposed long-distance interaction between the U5 (unique in 5') region and the region surrounding the AUG initiation codon of the *gag* gene; PBS, the primer binding site domain to which tRNA<sup>Lys,3</sup> has to be annealed to initiate reverse transcription; CU:GA (coloured in purple) a proposed long distance interaction between CU- and GA-rich regions; SL1, stem-loop 1 which contains the gRNA dimerization initiation site and is involved in gRNA packaging; SL2, which contains the main 5' splice site; SL3, which is also involved in gRNA packaging. While most of these elements are present in the majority of secondary structure models proposed in the literature, the sequences forming the U5: AUG and CU:GA long distance interactions may be involved in alternative interactions. (c) Experimental strategy used in this study. gRNA 1–600 was refolded *in vitro* and incubated with Pr55<sup>Gag</sup>, NCp15, NCp9 or NCp7 protein (right). SHAPE was performed directly on the RNA:protein complex (**Complex** condition) or after treatment of the complex with AT-2 (**AT-2** condition), which is able to eject Zn<sup>2+</sup> ions from the two zinc fingers located in the NC domain. Alternatively, the protein was removed by proteinase K before performing SHAPE (**ProtK** condition). Controls without protein were performed for each of the three conditions (left). However, no significant differences were observed between these three controls, which were thus pooled together (**NoProt** condition).

defect in infectivity that is correlated to a failure to accumulate late reverse transcription products [18,19]. Furthermore, morphological changes in the virion require neither the NC domain nor gRNA [21] and they do not necessarily match with maturation of the gRNA [16,18,22]. Thus, morphological maturation of the virus is not sufficient for infectivity but must be accompanied by maturation of the gRNA structure, and it has been reported that the native structure of the HIV-1 gRNA is required for successful reverse transcription [23,24].

However, the relationship between virion protein maturation and structural maturation of the gRNA is poorly understood, despite the fact that these two phenomena are intricately linked, as gRNA increases the kinetics of Pr55<sup>Gag</sup> cleavage by the viral protease [8,25]. The structural rearrangements of the gRNA taking place during maturation remain largely unknown, mainly because it is extremely difficult to analyse the gRNA structure in virions blocked at different steps of the maturation process. Indeed, while recent

progresses in chemical probing has made it possible to analyse the structure of viral RNAs in infected cells or within viral particles or after extraction, such studies are still restricted to infectious viruses [26–29]. As an alternative to such an approach, here, we systematically compared binding of Pr55<sup>Gag</sup>, NCp15, NCp9, and NCp7 to the 5' untranslated region (5'-UTR) of HIV-1 gRNA. We used RNA chemical probing to compare the protein footprints and their transient effects of the RNA structure, the importance of the zinc fingers in the binding of these proteins to gRNA, and the permanent refolding they impose to the gRNA structure. Indeed, while several Gag [30–33] and NCp7 [26,34,35] footprinting studies have been reported in the literature, no systematic comparison of Pr55<sup>Gag</sup> and its consecutive maturation products, which would provide insight into the gRNA structural maturation process, is available.

We focused on the interaction of Pr55<sup>Gag</sup> and its NC-containing maturation products with the 5'-UTR of HIV-1 gRNA because this region is replete with sites that play key roles in HIV-1 replication and whose functions are directly linked to their structure (Figure 1(b)). Of special interest in the context of viral assembly and maturation are the primer binding site (PBS) domain and the packaging signal (Psi) region (Figure 1(b)). The PBS domain contains several sequences that are complementary to tRNA<sup>Lys3</sup>, which serves as a primer for HIV-1 reverse transcription, including the PBS itself, which is complementary to the 18 nucleotides (nts) at the 3' end of tRNA<sup>Lys3</sup>, and other regions complementary to the tRNA<sup>Lys3</sup> anticodon loop [36–38] and the TYC arm [39,40]. The Psi region corresponds to a ~ 110 nt long sequence that contains all the elements required for Pr55<sup>Gag</sup> binding [41] and is crucial for gRNA selection and packaging [42]. This region can adopt several different conformations, which have been proposed to regulate the switch from translation to packaging of the full-length unspliced gRNA [43–45]. It contains the conserved stem-loops 1 to 3 (SL1 to SL3) [46,47]: SL1 and SL3 play major roles in gRNA packaging [42,48], whereas SL2 regulates splicing of the full-length transcript and prevents cleavage and polyadenylation at the 5' copy of the polyadenylation site [42]. In addition, SL1 initiates dimerization of gRNA, which is mediated by a self-complementary sequence in the SL1 apical loop [49,50].

## Results

### Experimental strategy

In order to assess the capacity of Pr55<sup>Gag</sup> and its maturation products NCp15, NCp9 and NCp7 to bind and rearrange the 5'-end region of the HIV-1 genome, we performed detailed RNA structure and protein binding analysis using high-throughput selective 2'-hydroxyl acylation analysed by primer extension (hSHAPE). SHAPE experiments interrogate RNA structural dynamics using electrophilic anhydride that preferentially acylate the 2'-hydroxyl (2'-OH) ribose groups of single stranded RNA [51,52]. 2'-OH acetylation leads to a block to reverse transcription that can be quantitated using fluorescent oligonucleotides on a capillary electrophoresis device [26]. SHAPE experiments are routinely used to probe RNA

structures [51–53] and protein binding sites in viral genomes [26,30,54,55]. Here, we performed all hSHAPE experiments on *in vitro* transcribed RNA corresponding to the first 600 nts of HIV-1 gRNA under well-defined experimental conditions. In our analysis, we focused on the untranslated region (nucleotides 1–335), as it contains functional sites that play key roles in HIV-1 replication (Figure 1(b)) and is evolutionary more conserved than the HIV-1 gRNA coding regions [47].

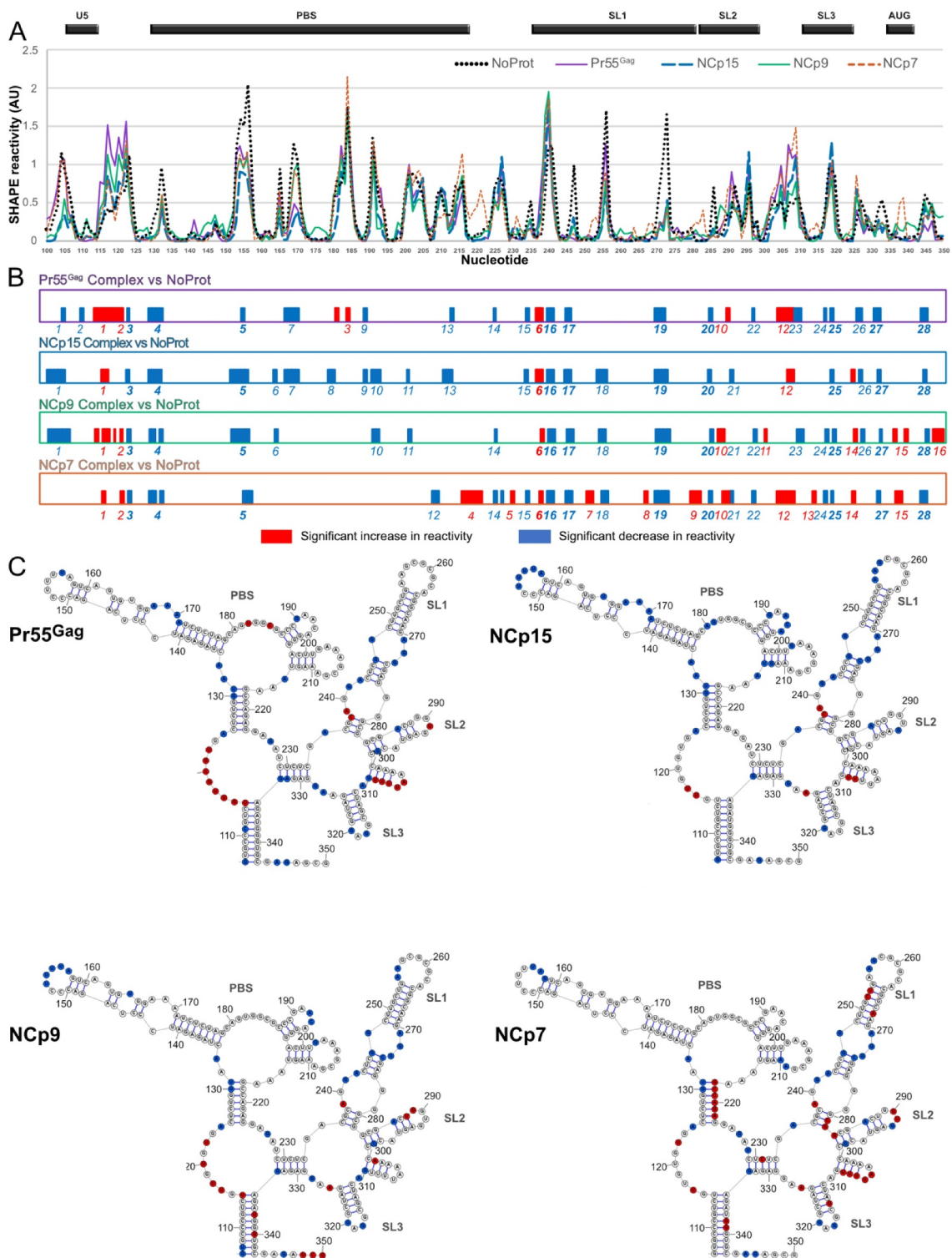
To test the effect of Pr55<sup>Gag</sup> and its maturation products on the structure of the 1–600 gRNA, we analysed four conditions (Figure 1(c)): 1) gRNA 1–600 in the absence of protein (**NoProt** condition), 2) gRNA 1–600 in complex with Pr55<sup>Gag</sup>, NCp15, NCp9, or NCp7 (**Complex** condition), 3) gRNA 1–600:protein complexes treated with aldrithiol-2 (AT-2), a compound that is able to eject the Zn<sup>2+</sup> ions from the two zinc fingers located in the NC domain [26] (**AT-2** condition), and 4) gRNA 1–600:protein complexes treated with proteinase K (**ProtK** condition). Since we wanted to mimic conditions prevailing in viral particles, and viral particles are known to contain between 1,000 and 5,000 of Pr55<sup>Gag</sup> or its maturation products [56], all complexes were formed at a ratio of 1 protein molecule per 10 nts of gRNA 1–600. Comparing the averaged SHAPE reactivity values of the **Complex** condition with SHAPE reactivity values of the **NoProt** condition allowed us to identify the protein binding sites on gRNA 1–600, as well as the effects of protein binding on the RNA structure. The reverse footprints of the two zinc-fingers located in the NC domain of full-length Pr55<sup>Gag</sup> and its maturation products (Figure 1(a)) were identified by comparing SHAPE reactivities of the **Complex** condition with the ones of the **AT-2** conditions. Finally, in order to visualize the permanent effect of the proteins on the RNA structure after they have been removed, *i.e.* their RNA chaperone activity, the SHAPE reactivities of the **ProtK** samples were compared to the ones of the **NoProt** samples.

Separate controls without protein were performed for each of the **Complex**, **AT-2**, and **ProtK** conditions (Figure 1(c), **left part**), but since they did not show any significant differences amongst conditions, all **NoProt** datasets were pooled together (**Supplementary Dataset 1**).

In order to minimize the errors on the SHAPE reactivity values, several data sets were obtained for each condition. Pairwise comparisons performed between all data sets obtained under the same condition usually showed very good correlation (**Supplementary Figure 1**), and if the correlation of a data set with the other ones was < 0.70 it was discarded. We averaged data from 3 to 6 experiments for **Complex**, **AT-2**, and **ProtK** conditions (**Supplementary Dataset 1**). For each protein, the averaged SHAPE values were then used to perform pairwise comparison between the different conditions.

### Pr55<sup>Gag</sup> and its maturation products differently bind and affect RNA structure

As a reference, we first analysed the structure of the 1–600 gRNA in the absence of protein (**NoProt** condition) (Figure 1(c), **left part**). NoProt SHAPE reactivities (**Supplementary dataset 1**) and the deduced secondary



**Figure 2.** Comparison of the SHAPE reactivity profiles of the 5' region of HIV-1 either alone or in complex with Pr55<sup>Gag</sup>, NCp15, NCp9, or NCp7. (a) The SHAPE reactivity profiles of nts 100–350 under *NoProt*, *Pr55<sup>Gag</sup> Complex*, *NCp15 Complex*, *NCp9 Complex*, and *NCp7 Complex* conditions are overlaid. (b) The significant differences between the *Complex* and *NoProt* conditions are drawn schematically and numbered. Reactivity decreases upon formation of the complexes are indicated by blue bars, while reactivity increases are indicated by red bars. Regions of decreased and increased reactivity upon complex formation are numbered in blue and red, respectively. All differences presented in this panel were statistically significant and considered to be biologically relevant (see Data analysis in the Method section). (c) These differences are plotted on the RNA secondary structure model obtained using the *NoProt* SHAPE reactivities as constraints, using the same colour code. Since the TAR and Poly A structures were not affected by addition of the proteins, they were omitted in the secondary structure models for clarity.

structure model (**Supplementary Figure 2**) correlated very well with previous studies [26,27], indicating that our 1–600 gRNA folds into the native conformation.

Pr55<sup>Gag</sup>, NCp15, NCp9, and NCp7 complexes with gRNA 1–600 all show SHAPE reactivity profiles that differ from the **NoProt** profile, as well as between themselves (**Figure 2(a)**). Schematic representation of the significant differences between the **Complex** and **NoProt** profiles (**Figure 2(b)**) indicates that protein binding mainly resulted in regions of decreased SHAPE reactivity, although regions of increased reactivity were observed with all proteins. While increased reactivity point to regions of gRNA 1–600 destabilized upon protein binding, decreased reactivity can either reflect protein footprints or RNA regions whose structure is stabilized by the protein.

Overall, vast majority of protected (blue) or destabilized (red) regions are common to several proteins, but only one-third (10/28) of protected regions and one amongst 16 destabilized regions are found for all four proteins (**Figure 2(a,b)**). This indicates that even if some similarities can be found in the binding of Pr55<sup>Gag</sup> and its maturation products, they each bind to gRNA and affect its structure in different manners. NCp7 distinguishes itself as having the most pronounced destabilizing effect on the RNA structure, while on the contrary NCp15 barely destabilized the RNA structure at all, and both Pr55<sup>Gag</sup> and NCp9 had intermediate behaviours (**Figure 2(b)**). The overall nt composition of the protected regions is similar for all proteins, but slightly more U-rich for NCp15 and NCp9 (**Supplementary Figure 3**). This is due to the fact that these two proteins protect longer stretches of nts than Pr55<sup>Gag</sup> and NCp7, and these stretches contain ~40% of uridines (**Table 1**).

In order to allow a better comparison between Pr55<sup>Gag</sup> and its maturation products, we plotted the differences

observed between the SHAPE profiles under the **Complex** and **NoProt** conditions on the RNA secondary structure model that best fits the SHAPE reactivity of the **NoProt** condition (**Figure 2(c)**). While protected regions are observed all along the studied region of HIV-1 gRNA, the vast majority of protected regions that were common to all four proteins are located between nts 235 to 333 (**Figure 2(b)**). This region contains all determinants required for optimal Pr55<sup>Gag</sup> binding [41] and is usually considered as containing the main HIV-1 packaging signals [42,43,45,57] and is therefore often referred to as the Psi region. Within this region, most common protections were observed in the internal loops of the extended SL1 (**Figure 2(c)**), which constitutes a Pr55<sup>Gag</sup> primary binding site [30,58]. An additional protection was observed in the SL1 apical loop (nts 255–257), which mediates HIV-1 gRNA dimerization [49], but only after the Gag precursor underwent the first proteolytic maturation (**Figure 2(b,c)**). The basal part of SL1 also contains the only region of gRNA 1–600 that is destabilized by the four proteins. (**Figure 2(b,c)**). Furthermore, single nt protections common to all proteins were also observed in SL2 and SL3, and in the GA-rich region (nts 328–333). Besides, protections common to all proteins also included 4 nts in the basal part of the PBS domain (nts 130–133) (**Figure 2(c)**).

Most regions destabilized by NCp7 were concentrated in the Psi region. NCp7 uniquely destabilizes the apical and basal stems of SL1 (**Figure 2(b,c)**), and this may promote the transition from a kissing complex to an extended duplex form of RNA dimer [50,59,60], which has been proposed to correspond to the stabilization of the gRNA dimer observed during the last steps of the viral particle maturation [5,6,16]. NCp7 also destabilized a stretch of four uridines (nts 305–308) that can form an unstable helix located between SL2 and SL3 (**Figure 2(c)**). Alternatively, these nucleotides might be rearranged into an extended form of SL3. Of note, one study recently proposed that unwinding of this unstable SL3 extension may be crucial for gRNA packaging [34]. Given that this model was based on experiments with mature NCp7 protein, it is interesting to point out that this uridine stretch is strongly destabilized both by NCp7 and Pr55<sup>Gag</sup>, but not by NCp15 or NCp9 (**Figure 2(b,c)**).

Outside of Psi, we observed that NCp7, but none of the other proteins included in this study, strongly destabilized the 3' strand of the basal stem of the PBS domain (nts 217–223) (**Figure 2(b,c)**). The 5' strand of this helix contains the primer activation signal (PAS) that is complementary to the TYC arm of the reverse transcription primer tRNA<sup>Lys,3</sup> and has been proposed to play a key role in the initiation of reverse transcription [39,40]. Additionally, NCp15 protects several stretches of nts in the PBS domain (**Figure 2(c)**). Of these, the A-rich internal loop (nts 168–171) is complementary to and interacts with the anticodon loop of tRNA<sup>Lys,3</sup> [36–38]. Altogether, these results suggest that the unique properties of NCp15 and NCp7 stabilize the reverse transcription initiation complex during maturation of the viral particles [61].

**Table 1.** Nucleotides protected by Pr55<sup>Gag</sup> and its maturation products.

PROTEIN	PROTECTED NUCLEOTIDES							Total	
	Stretch length (nts)								
Pr55 <sup>Gag</sup>	Nature	A	4	5	1	6	/	/	<b>35</b>
		C	1	/	/	/	/	/	
		G	6	7	2	1	/	/	
		U	1	/	/	1	/	/	
	<b>Total</b>	<b>12</b>	<b>12</b>	<b>3</b>	<b>8</b>	/	/		
NCp15	Nature	A	3	2	5	7	1	/	<b>48</b>
		C	/	/	1	1	/	/	
		G	7	4	2	3	3	/	
		U	1	/	1	1	6	/	
	<b>Total</b>	<b>11</b>	<b>6</b>	<b>9</b>	<b>12</b>	<b>10</b>	/		
NCp9	Nature	A	3	5	/	1	1	/	<b>39</b>
		C	/	1	/	1	/	/	
		G	8	5	/	2	/	3	
		U	1	1	/	/	4	3	
	<b>Total</b>	<b>12</b>	<b>12</b>	/	<b>4</b>	<b>5</b>	<b>6</b>		
NCp7	ature	A	4	5	1	1	/	/	<b>30</b>
		C	/	/	/	1	/	/	
		G	6	6	1	2	/	/	
		U	1	1	1	/	/	/	
	<b>Total</b>	<b>11</b>	<b>12</b>	<b>3</b>	<b>4</b>	/	/		

The total number of nts protected by each protein, as well as the number of nts involved in 1, 2, 3, 4, 5 or 6 protected nt stretches is indicated; the nt composition of these stretches is also reported. U-rich 5 and 6 nt stretches protected by NCp15 and NCp9 are highlighted in pink.

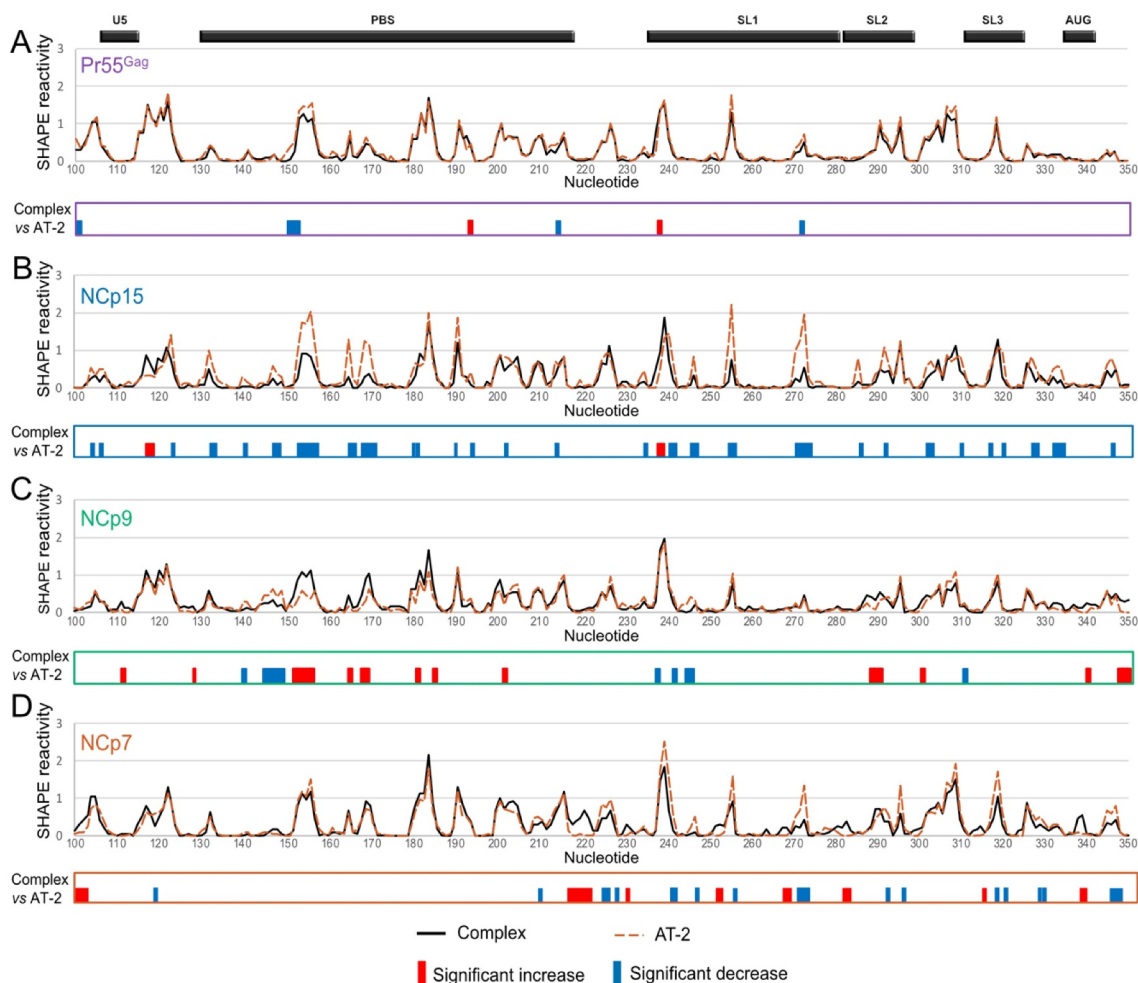
### The zinc fingers in the nucleocapsid domain differently interact with HIV-1 gRNA in Pr55Gag and its maturation products

Since Pr55<sup>Gag</sup> and its maturation products were found to bind gRNA 1–600 and to affect its structure in a protein-specific manner, we next asked whether the two conserved zinc fingers (Figure 1(a)) also play specific protein-dependent roles in RNA binding. To this aim, we treated the pre-formed gRNA 1–600:protein complexes with AT-2, which is able to eject the Zn<sup>2+</sup> ions from the CCHC-type zinc fingers [26], and compared the SHAPE reactivity values obtained in the **Complex** condition to those obtained in the **AT-2** condition (Figure 3).

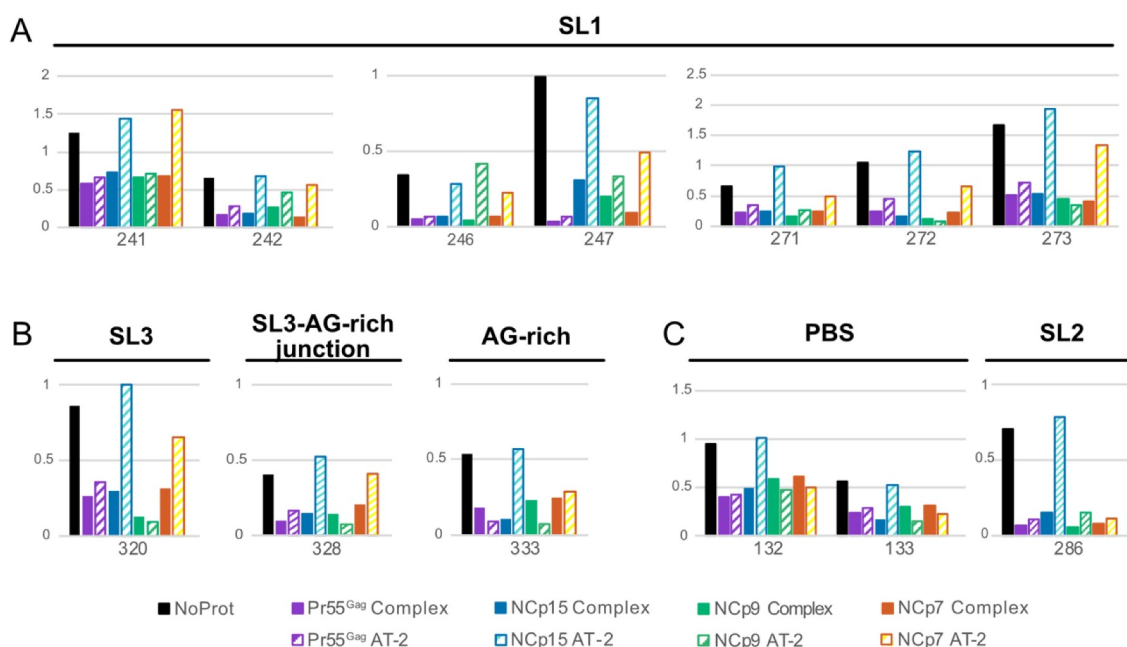
On the one hand, AT-2 treatment of the Pr55<sup>Gag</sup>:gRNA 1–600 complex had minimal effect on SHAPE reactivities, suggesting that once the complex is formed, the zinc fingers are not required to maintain the interaction with RNA (Figure 3(a)). Similarly, AT-2 had fairly limited effects on the NCp9:RNA complex, albeit more differences were observed between the **Complex** and **AT-2** conditions than in the case of the full-length Gag precursor, including a few reactivity decreases upon AT-2 treatment (Figure 3(c)). On the other hand, AT-2 treatment increased the SHAPE

reactivity of gRNA 1–600 in complex with NCp15 in numerous regions (Figure 3(b)). Indeed, for NCp15, the SHAPE reactivity differences between the **Complex** and **AT-2** conditions closely mimicked those observed between the **Complex** and **NoProt** conditions (compare Figures 2(b) and 3(b)), indicating that the zinc fingers play a crucial role in maintaining the NCp15:RNA complex. A similar observation can also be made for most NCp7 binding sites. With NCp7, not only most of the reactivity increases, but also most of the reactivity decreases were similar when comparing the **AT-2** and **Complex** conditions or the **NoProt** and **Complex** conditions (compare Figures 2(b) and 3(d)).

To obtain a clearer picture of the role of the zinc fingers, we took a closer look at the nts located in the Psi region whose reactivity was significantly affected by addition of all four proteins (Figure 4). For nts in SL1 (*i.e.* nts 241–242, 246–247, 271–273) addition of AT-2 to the Pr55<sup>Gag</sup>:RNA complex (**Pr55<sup>Gag</sup> AT-2** condition) had little or no effect on the SHAPE reactivity of the complex, while treatment of the NCp15:RNA complex with AT-2 (**NCp15 AT-2** condition) increased the SHAPE reactivity to levels similar to those observed in the absence of protein (**NoProt** condition)



**Figure 3.** Effect of AT-2 on the RNA:Pr55<sup>Gag</sup> (a), RNA:NCp15 (b), RNA:NCp9 (c) and RNA:NCp7 (d) complexes. In each panel, the **Complex** and **AT-2** SHAPE reactivity profiles are overlaid in the upper part, whereas the significant differences between the profiles are drawn in the lower part. All differences presented in the lower parts were statistically significant and considered to be biologically relevant (see Data analysis in the Method section). Reactivity increases upon AT-2 treatment are represented by blue bars, while reactivity decreases are indicated by red bars.



**Figure 4.** Effects of AT-2 on protected nucleotides located in SL1 (a), SL3, the AG-rich region and the junction between them (b), and PBS and SL2 regions (c). For each nt the SHAPE reactivity in the *NoProt* condition and in the *Complex* and *AT-2* conditions for each of the four proteins are compared.

(Figure 4(a)). Although the effects of AT-2 on the NCp9:RNA and NCp7:RNA complexes (*NCp9 AT-2* and *NCp7 AT-2* conditions, respectively) were less clear-cut, distinct trends could be observed. AT-2 usually had limited effect on the SHAPE reactivity of the NCp9:RNA complexes (albeit G246 is a noticeable exception), while AT-2 increased the reactivity of the NCp7:RNA complexes at most positions, without always reaching the same reactivity level as in the *NoProt* condition (Figure 4(a)). Similar AT-2 effects were observed at nts located in SL3 (nt 320), in the junction between SL3 and the CU:GA interaction (nt 328), and in the GA-rich sequence itself (nt 333) (Figure 4(b)). In the PBS domain (nts 132–133) and SL2 (nt 286), AT-2 retained its destabilizing effect on the NCp15:RNA complexes, but not on the NCp7:RNA complexes (Figure 4(c)), suggesting that NCp7 differently binds to the gRNA regions that are important for packaging and regions that are not. Besides, AT-2 had no effect on the Pr55<sup>Gag</sup>:RNA and NCp9:RNA complexes located in the PBS and SL2 domain (Figure 4(c)), as also observed in the other regions of gRNA 1–600 (Figure 4(a,b)).

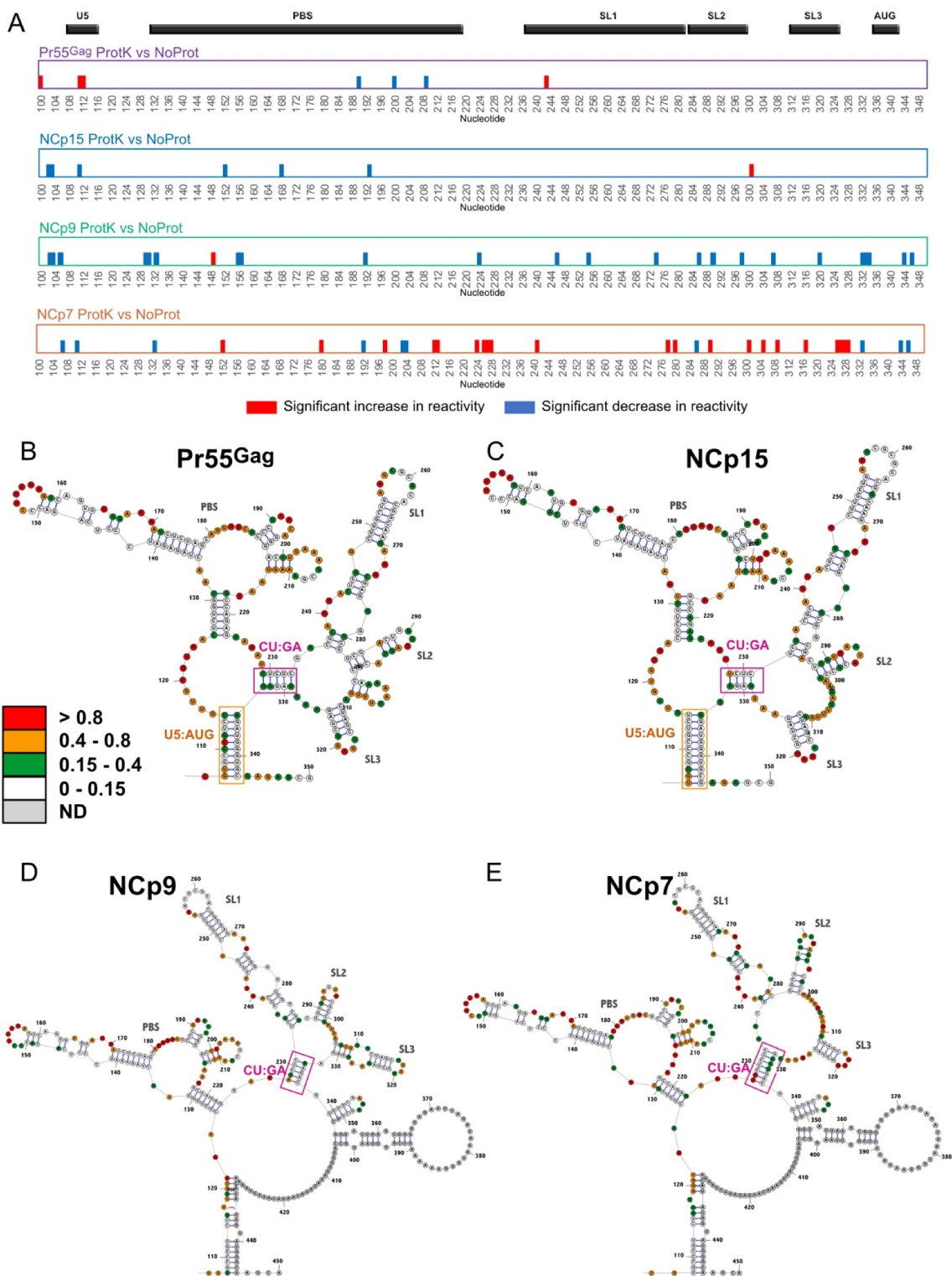
### *Pr55<sup>Gag</sup> and its maturation products progressively refold the gRNA structure*

Addition of Pr55<sup>Gag</sup> or its maturation products resulted in numerous protein-specific protections as well as destabilization of gRNA 1–600 (Figure 2), and we next asked whether these effects were transient (*i.e.* if they required the protein to be present in order to be observed) or if they were permanent (*i.e.* if they would persist after removing the protein). Permanent effects on the RNA structure are the hallmark of an RNA chaperone activity, which has been documented for Pr55<sup>Gag</sup> and its maturation products [13],

particularly NCp7 [14,15]. To this aim, we compared, for each protein, the SHAPE reactivity profiles in the *Complex*, *ProtK* and *NoProt* conditions (Supplementary Figure 4). Indeed, the *ProtK* SHAPE profiles were far more similar to the *NoProt* profile than to the *Complex* profiles, indicating that most of the effects of the proteins on the gRNA structure were transient. Nevertheless, while the differences between the *ProtK* and *NoProt* profiles were quite limited in the case of Pr55<sup>Gag</sup> and NCp15, they were more significant with NCp9 and NCp7 (Supplementary Figure 4 and Figure 5(a)), indicating that the Pr55<sup>Gag</sup> maturation products that are produced later in the viral maturation process have a more pronounced RNA chaperone activity. However, these two proteins have strikingly different effects on the gRNA structure: while the NCp9 *ProtK* profile mainly showed patches of nts whose reactivity decreased compared to the initial *NoProt* structure, the NCp7 *ProtK* profiles showed equivalent amounts of nt patches with increased and decreased reactivity compared to the *NoProt* Structure (Figure 5(a)).

We next took advantage of the *ProtK* data to analyse the effects of Pr55<sup>Gag</sup> and its maturation products on the gRNA secondary structure. In contrast to the *Complex* data sets, where SHAPE reactivities are modulated by protein binding, SHAPE reactivities in *ProtK* data sets can be used as constraints in RNA folding algorithms to obtain information on gRNA secondary structure. The first (*i.e.* the most stable) structure predicted when incorporating the *ProtK* SHAPE reactivities as constraints are shown for each protein in Figure 5(b–e). All structures presented an identical folding of the TAR and polyA domains, which were therefore omitted for clarity. Considering the limited differences between the *NoProt*





**Figure 5.** RNA chaperone activity of Pr55<sup>Gag</sup>, NCp15, NCp9, and NCp7 on the gRNA 1–600 structure. (a) The significant differences between the *ProtK* and *NoProt* SHAPE reactivity profiles are represented schematically for each protein. Reactivity increases in the *ProtK* conditions relative to the *NoProt* conditions are represented by red bars, while reactivity decreases are indicated by blue bars. All differences presented in this panel were statistically significant and considered to be biologically relevant (see Data analysis in the Method section). (b–e) Most stable secondary structure models of the 5' region of HIV-1 gRNA obtained using the Pr55<sup>Gag</sup> *ProtK* (b), NCp15 *ProtK* (c), NCp9 *ProtK* (d), and NCp7 *ProtK* (e) SHAPE values as constrains. The SHAPE reactivity values are reported on the structures and colour-coded as indicated in the insert. The U5:AUG (when existing) and the CU:GA interactions are indicated in Orange and purple, respectively. Since the TAR and Poly A domains were conserved in all structures, they were omitted for clarity.

and Pr55<sup>Gag</sup> *ProtK* SHAPE (Figure 5(a)), it is not very surprising that the secondary structure that best fits the experimental data obtained under these two conditions is identical (compare **Supplementary Figure 2** and **Figure 5 (b)**). Incubation with NCp15 had subtle effects on gRNA 1–600 once it was removed (compare **Figure 5(b,c)**). The metastable stem-loop located between SL2 and SL3 was destabilized, resulting in its complete unfolding. Besides, the basal part of SL2 was also unwound and the entire SL2 remodelled, and a concomitant extension of SL1 was observed. This in turn shifted the register of the interacting CU-rich and GA-rich sequences by two nucleotides, U230 interacting with A330 (**Figure 5(c)**) instead of A332 (**Figure 5(b)**). By contrast the RNA chaperone activity of NCp9 had a dramatic effect on the HIV-1 RNA structure. SL3 was stabilized, forming the extended hairpin recently proposed to play a key role in gRNA packaging [34], which in turn induced a further register shift of the CU:GA interaction, U230 now interacting with A334 (**Figure 5(d)**). Importantly, this remodelled CU:GA interaction is not compatible with the U5:AUG interaction that has been proposed in an number of HIV-1 RNA secondary structure models [31,41,62,63] and existed in the *NoProt*, Pr55<sup>Gag</sup> *ProtK* and NCp15 *ProtK* secondary structure models (**Supplementary Figure 2** and **Figure 5 (b,c)**), and hence a reorganization of the long-range interactions that maintain the global RNA structure was observed after incubation with and removal of NCp9 (**Figure 5(d)**). Of note, in this secondary structure model, region 227–337 of HIV-1 gRNA that contains the core Pr55<sup>Gag</sup> binding domain [41] and the main packaging signal [42] folds into an independent structural domain (**Figure 5(d)**), as proposed earlier [46]. In the gRNA 1–600 structure remodelled by NCp9, the U5 region surrounding nt 110 interacts with a region located downstream in the *gag* gene (nts 440–446) (**Figure 5(d)**), rather than with the *gag* AUG codon (**Figure 5(b,c)**). This new long-distance interaction is maintained in the structure resulting from the NCp7 RNA chaperone activity, while the CU:GA interaction and SL2 are stabilized at the expense of the basal part of the extended SL1 (**Figure 5(e)**). Similarly, the extended SL3 that was found in NCp9 *ProtK* is also destabilized (**Figure 5(e)**). Significantly, examination of the Los Alamos HIV sequence database (<https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html>) indicates that the sequences forming the CU:GA interaction are highly conserved amongst HIV-1 isolates. Likewise, the two lower stems of the U5:*gag* interaction are also highly conserved, indicating that the structural rearrangements we observed with the NL4.3 isolate are also possible in the other HIV-1 isolates, suggesting they are important for function.

Importantly, the structural changes observed with the different proteins do not reflect differences in the dimerization status of gRNA 1–600. Indeed, gRNA 1–600 was  $81 \pm 2\%$  dimeric in the absence of protein, and the dimer fraction further increased to 89–95% after incubation with the

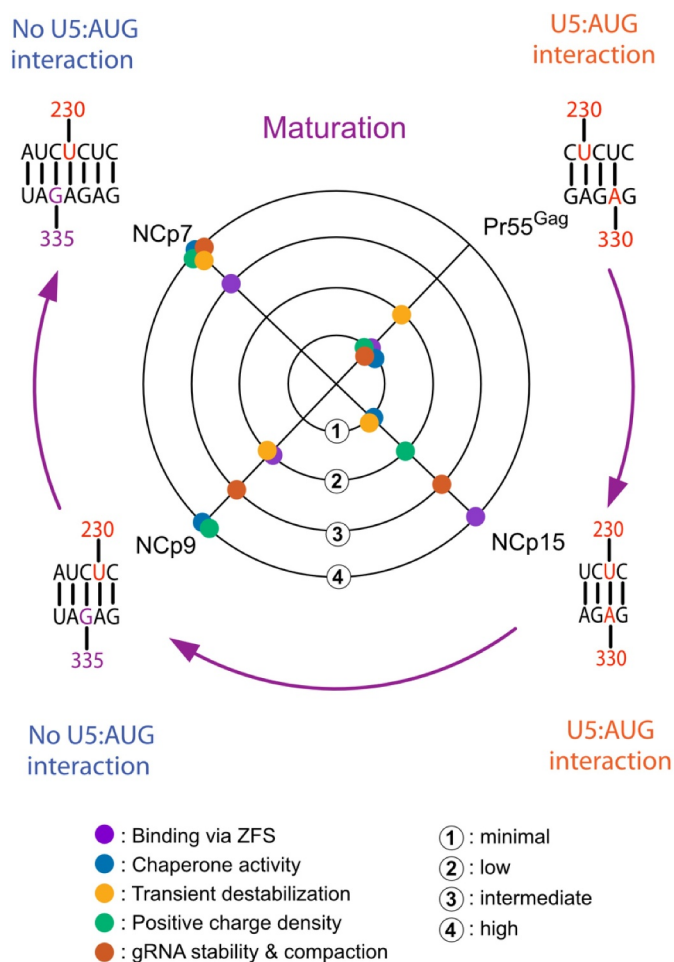
proteins, without any significant difference between proteins (**Supplementary Figure 5**).

## Discussion

The structural rearrangements of the HIV-1 gRNA during maturation of the viral particles remain unexplored, despite evidence that they are required for infectivity [18,19,23,24]. Indeed, despite recent progresses, chemical probing of viral genomes in infected cells, within viral particles, or after extraction is still restricted to infectious viruses [26–29], whereas immature and partially mature viruses are non-infectious. As an alternative approach, here we performed a systematic comparative study of Pr55<sup>Gag</sup> and its NC-containing maturation products regarding i) the protein binding sites and their transient structural effects, ii) the importance of the two NC zinc fingers in the stability of the complexes, and iii) the permanent effects of these proteins on the HIV-1 gRNA structure. Even though a few previous studies used structural approaches to identify Gag or/and NCp7 [26,32,34,35] binding sites *in vitro* [30–34,41] or directly in viral particles [26,35], no systematic comparison of the Pr55<sup>Gag</sup>, NCp15, NCp9, NCp7 binding sites has been reported. Furthermore, several of these studies were performed using fusion proteins [32,33] or GagΔp6 [31] as a surrogate for full-length Pr55<sup>Gag</sup>. Similarly, AT-2 has only been used to analyse the contribution of the zinc fingers of mature NCp7 [26], and, to the best of our knowledge, the chaperone activity of Pr55<sup>Gag</sup> and its maturation products on the 5' region of HIV-1 gRNA has never been described, even though others have analysed the *in vitro* RNA chaperone activity of these proteins using model systems [13,15,38,64].

For the part of this study that can be compared to published works, our results fit well with previous data. Of particular interest is the finding that our *in vitro* NCp7 binding analysis identifies the same binding sites as a similar analysis performed on mature viral particles, indicating that our *in vitro* analysis is relevant to the situation prevailing in viral particles [26]. Similarly, our study reveals that NCp7 has a much greater propensity to remodel the HIV-1 gRNA structure than Pr55<sup>Gag</sup>, in agreement with experiments conducted on small model systems [13,15,65].

Overall, our study reveals marked differences in the binding of Pr55<sup>Gag</sup>, NCp15, NCp9, and NCp7 to the 5' region of HIV-1 gRNA: some of the binding sites of these proteins are different and their transient effect on the RNA structure (**Figure 2**) as well as their chaperone activity differ, leading to different gRNA secondary structures (**Figure 5**). In addition, the role of the zinc fingers in maintaining the protein:RNA complexes varies from minimal to crucial between Pr55<sup>Gag</sup> and its NC-containing maturation products (**Figures 3 and 4**). Strikingly, neither the overall destabilizing effect nor the chaperone activity of Pr55<sup>Gag</sup> and its maturation products correlates with the role of the zinc fingers in the protein:gRNA complex stabilization (**Figure 6, central part**). Indeed, the zinc fingers have



**Figure 6.** Comparison of Pr55<sup>Gag</sup> and its maturation products with respect to the contribution of the zinc fingers to RNA binding, RNA chaperone activity, transient destabilization of RNA, overall charge density, and stabilization and compaction of the gRNA dimer (transient destabilization of RNA (central part) and evolution of the CU:AG interaction after exposure of the gRNA to Pr55<sup>Gag</sup> and its maturation products (outer part). Central part: Contribution of the zinc fingers to RNA binding, RNA chaperone activity, and transient destabilization of RNA were assessed semi-quantitatively from results of this study (Figs. 2, 3 & 4, and 5, respectively). Stabilization and compaction of the gRNA dimer were evaluated from published studies [18,19], and the positive charge density of each protein was calculated as the net positive charge at neutral pH divided by the number of amino acids in the protein. The positive charge density is 0.042, 0.107, 0.225 and 0.236 for Pr55<sup>Gag</sup>, NCp15, NCp9, and NCp7, respectively. Outer part: The CU:AG interaction resulting from exposure to NCp9 and NCp7 prevents formation of the U5:AUG interaction.

opposite contributions in maintaining the Pr55<sup>Gag</sup>:gRNA and NCp15:gRNA complexes once they are formed, but these two proteins both have minimal chaperone activity and minimal to low transient RNA destabilization ability (Figure 6, central part). Remarkably, the effect of AT-2 on Pr55<sup>Gag</sup> (or NCp15) is the same at all RNA binding sites (Figures 3 and 4), despite the fact that Pr55<sup>Gag</sup> binds different RNA binding sites with very different affinities [30,41,58]: therefore the opposite effect of AT-2 on Pr55<sup>Gag</sup> and NCp15 does likely not reflect differences in the affinities of these two proteins for RNA. The fact that AT-2 has no effect on the Pr55<sup>Gag</sup>:RNA complex may seem contradictory with the well-established observation that zinc fingers play a key role in the recognition of retroviral

gRNAs by their respective Gag precursors, leading to their specific packaging [66–68]. This contradiction is only apparent, and our data suggest that while the Pr55<sup>Gag</sup> zinc fingers are required for specific binding to gRNA, they are not required for maintaining the Pr55<sup>Gag</sup>:RNA complex once it is formed. An alternative but less likely interpretation could be that AT-2 is unable to extract the Zn<sup>2+</sup> ions from the Pr55<sup>Gag</sup> zinc fingers. Of note, while AT-2 has no significant effect on Pr55<sup>Gag</sup>:RNA complexes and limited effect on NCp9:RNA complexes, it completely destabilizes NCp15:RNA complexes and has also a strong effect on NCp7:RNA complexes. AT-2 thus differently affects the different protein:RNA complexes, and the magnitude of the effect does not correlate with the size of the protein, as would be expected if the protein domains flanking the Zn fingers would provide a steric protection against AT-2. Thus, if the ability of AT-2 to extract Zn<sup>2+</sup> ions differs amongst the different proteins, this must reflect different binding modes of the Zn fingers in the Pr55<sup>Gag</sup>:RNA and NCp9:RNA complexes on one side and in the NCp15:RNA and NCp7:RNA complexes on the other side.

The RNA chaperone activity of Pr55<sup>Gag</sup> and its maturation products correlates quite well with the positive charge density of these proteins (calculated as the net positive charge divided by the number of amino acids in the proteins) (Figure 6, central part). The primary role of the charge density in the RNA chaperone activity of Pr55<sup>Gag</sup> and its maturation products is in keeping with the observation that many proteins with disordered positively charged regions, including synthetic polypeptides, display RNA chaperone activity [69–71]. Likewise, the charge density seems to be the main factor governing compaction and stabilization of the gRNA dimer [19] (Figure 6, central part).

Pr55<sup>Gag</sup> and its maturation products mostly bind at identical sites in the Psi region of gRNA (Figure 2), likely reflecting the critical role of the Gag NC domain in the selection and packaging of retroviral gRNA [66–68]. The binding sites of Pr55<sup>Gag</sup>, NCp15, NCp9, and NCp7 in the Psi region exclusively consist in stretches of one to four nts that are highly enriched in purines, and particularly in G residues (73 of the 152 protected nucleotides are G residues (Figure 2 and Table 1)). This observation is in keeping with a recent study indicating that unpaired guanines in the 5'-UTR of HIV-1 gRNA act synergistically to mediate genome packaging [72] and with the central role of guanines for RNA chaperone function [55]. This however does not imply these sites binds Pr55<sup>Gag</sup>, NCp15, NCp9, and NCp7 with the same affinity or that any of these proteins binds all its binding sites with the same affinity [30,73]. The finding that Pr55<sup>Gag</sup> and its NC-containing maturation products bind at the same sites in the Psi region is however at odds with the widely accepted view that the mature NCp7 completely covers the viral genome in order to protect it [14]. Indeed, many regions remain accessible to the SHAPE probe in the NCp7:gRNA 1–600 complex, and some of these regions may also be accessible to RNases. In light of this observation, the recent finding that reverse transcription takes place in an

intact core that is imported into the nucleus and disassembles just before integration appears of particular importance [74,75]. In fact, in contrast with Pr55<sup>Gag</sup> and its first maturation products, NCp7 extensively destabilizes the Psi region of gRNA (Figure 2). Analysis of mutant HIV-1 blocked at various maturation stages suggests that this destabilization is required for successfully completing reverse transcription [18,19]. In contrast with the Psi region, Pr55<sup>Gag</sup>, NCp15, NCp9, and NCp7 bound differently in the PBS domain (Figure 2). In particular NCp15 and NCp7 show distinctive binding patterns in this domain that may drive the stabilization of the reverse transcription initiation complex [36–40,76] that takes progressively place during maturation of the viral particles [61].

The different RNA chaperone activity of Pr55<sup>Gag</sup> and its maturation products suggest that the gRNA structure is progressively remodelled inside maturing viral particles. A key region involved in these rearrangements seems to be the CU:GA interaction, with the two interacting strands being progressively shifted as Pr55<sup>Gag</sup> is replaced by its successive maturation products, eventually causing a dramatic change in the overall gRNA secondary structure (Figures 5 and 6, **outer part**). While SL1, SL2 and SL3 all exist in more or less extended forms (Figure 5), extension of these SLs do not seem to regulate the register shift of the CU:GA interaction since the same forms of SL1, SL2 and SL3 can coexist with different CU:GA interaction registers (compare Figure 5(b,e)). Importantly, the remodelled CU:GA interaction existing after gRNA exposure to NCp9 and NCp7 is not compatible with the U5:AUG interaction that has been proposed in an number of HIV-1 RNA secondary structure models [31,41,62,63] and exists in the *NoProt*, *Pr55<sup>Gag</sup> ProtK* and *NCp15 ProtK* secondary structure models (Compare **Supplementary Figure 2** and **Figure 5(b-d)**). Of note, in the secondary structure models resulting from the NCp9 and NCp7 RNA chaperone activity, region 227–337 of HIV-1 gRNA that contains the core Pr55<sup>Gag</sup> binding domain [41] and the main packaging signal [42] folds into an independent structural domain (Figure 5(d)), as proposed in an earlier study [46]. It thus cannot be excluded that, *in vivo*, this gRNA conformer might already exist in producer cells, maybe as a minor one, and be selectively packaged into budding viral particles. In-cell RNA probing methodologies able to detect the coexistence of several RNA conformers, which are currently being developed, should be able to address directly this question. *In vivo* probing could also allow to determine the effect of tRNA<sup>Lys,3</sup> annealing on the structural rearrangements of gRNA during the maturation process [61]. It would also take into account the fact that there are more Pr55<sup>Gag</sup> copies in immature particles than NCp7 copies in mature ones [56], whereas we kept the protein concentration constant in all our *in vitro* assays in order to allow a strict comparison between the experimental conditions. Finally, while the gRNA secondary structure models after exposure to NCp9 and NCp7 are very similar (Figures 5(d,e) and 6, **outer part**), the ability of NCp7 to transiently unwind

RNA secondary structures (Figures 2 and 6, **central part**), which is much more pronounced than that of NCp9, is likely crucial for efficient reverse transcription, explaining why infectivity is only acquired after complete maturation of the viral particles.

## Methods

### Protein expression and purification

Expression, purification and characterization of NL4.3 Pr55<sup>Gag</sup> with an appended C-terminal His<sub>6</sub>-tag was performed as described by McKinstry et al [77]. Recombinant wild-type NCp7, NCp9 and NCp15, respectively 55, 71 and 123 amino acids in length, were expressed and purified as described previously [78–80].

### RNA synthesis

gRNA 1–600 corresponding to the first 600 nts of the wild type NL4.3 HIV-1 gRNA was synthesized by *in vitro* transcription from plasmid pDR4607 [81], after linearization with the *PvuII* restriction enzyme. Transcription was performed using a MEGAscript T7 Transcription kit (Thermo Fisher Scientific) following the manufacturer's instructions and purified by exclusion chromatography on a TSKgel G4000SW column as previously described [81]. RNA integrity and purity were confirmed by denaturing polyacrylamide gel electrophoresis.

### Formation of the RNA-protein complexes

In order to form RNA-protein complexes, RNA was folded in a buffer favouring dimerization. Briefly, four pmoles of gRNA 1–600 were denatured in water at 90°C for 2 min then placed on ice for 2 min. The RNA was then incubated at 37°C for 30 min in a 20 µl final volume of 1X Folding Buffer (30 mM HEPES pH 8.0; 300 mM KCl; 5 mM MgCl<sub>2</sub>) supplemented with 40 U RNasin® (Promega) and 1 µg total yeast tRNA. In parallel, 240 pmoles of protein (Pr55<sup>Gag</sup>, NCp15, NCp9 or NCp7) were incubated on ice for 15 min in the 1X Folding Buffer supplemented with 0.01% Triton X-100; 0.01 M DTT and 4 µg BSA in a final volume of 100 µl.

The ribonucleoprotein (RNP) complexes were formed by mixing 20 µl of RNA mixture with 100 µl of protein solution or 100 µl 1X Folding buffer as a control without protein and incubated at 37°C for 30 min then at 0°C for 15 min. The RNP samples were divided in two equal fractions that were treated with a SHAPE reagent (+) or with an equal volume of DMSO as a negative control (-) (see below).

### RNA modification with NMIA

#### Structural effects of protein binding to gRNA 1–600

A 60 µl solution of refolded gRNA 1–600 (*NoProt* condition) or of protein-RNA complex (*Complex* condition) was treated with 12 µl of 10 mM NMIA (Sigma-Aldrich) in anhydrous DMSO (Sigma-Aldrich) (+) or with 12 µl anhydrous DMSO (control, -). After 50 min at room temperature, 128 µl Milli-Q water were

added and material was precipitated with 3 volumes of ethanol, 1/10 volume of 3 M sodium acetate pH 5.0, 1  $\mu$ l of glycoBlue (ThermoFisher) for 30 min in a dry ice/ethanol bath and collected by centrifugation at 20,800 g for 30 min at 4°C. The pellets were washed twice with 70% ethanol, dried and resuspended in 15  $\mu$ l water and 4  $\mu$ l 5X Folding Buffer. One  $\mu$ l of proteinase K (Roche) was added and the digestion was performed at 37°C for 30 min before adding 79  $\mu$ l water. The modified RNA was purified by phenol-chloroform extraction (1:1, pH 7.5) and ethanol precipitated as described above. RNA pellets were resuspended in 6  $\mu$ l Milli-Q water.

### Effects of AT-2 on the gRNA 1-600-protein complexes

The RNP complexes (60  $\mu$ l) were treated with 2  $\mu$ l of 30 mM 2,2'-dithiodipyridine (Aldrihiol, AT-2) (Sigma-Aldrich) in DMSO for 1 h at 37°C. Then, RNA was modified by NMIA using the same protocol as above (AT-2 condition).

### RNA chaperone activity of Pr55<sup>Gag</sup> and its maturation products

After formation of the RNP complexes as described above, including incubations for 30 min at 37°C and for 15 min at 0°C, the proteins were eliminated by proteinase K treatment prior to RNA modification. The 60  $\mu$ l mixture was supplemented with 1  $\mu$ l proteinase K and incubated at 37°C for 30 min. Then, RNA was modified with 12  $\mu$ l of 10 mM NMIA in anhydrous DMSO or 12  $\mu$ l DMSO (negative control) and incubated for 50 min at room temperature (ProtK condition). Milli-Q water was added up to 200  $\mu$ l and the RNA was phenol-chloroform extracted and ethanol precipitated as described above.

### cDNA synthesis and analysis by capillary electrophoresis

RNA samples (6  $\mu$ l) treated with NMIA (+) or DMSO (-) were mixed with 2  $\mu$ l of a 1 mM AS primer 1 (5'- AGC TCC CTG CTT GCC CAT ACT A-3': complementary to nts 436–457 of gRNA 1–600) or AS primer 2 (5'- CTT CTG ATC CTG TCT GAA GG-3': complementary to nts 536–555 of gRNA 1–600) labelled with Vic (Life Technologies SAS, France). The mixture was heated at 90°C for 2 min and placed on ice for 2 min. After the addition of 2  $\mu$ l AMV RT Buffer 5x (125 mM Tris-HCl pH 8.3; 250 mM KCl; 10 mM DTT; 25 mM MgCl<sub>2</sub>) and 10 min incubation at room temperature, reverse transcription was performed by adding 2  $\mu$ l AMV RT Buffer 5x, 6  $\mu$ l dNTPs 2.5 mM (Invitrogen), 2 U of AMV RT (Life Sciences) and water to 20  $\mu$ l. Elongation was ensured by incubation 20 min at 42°C followed by 30 min at 50°C. The enzyme was inactivated at 60°C for 10 min. Simultaneously, a sequencing reaction was performed with 2 pmoles of unmodified RNA and 2  $\mu$ l of a 2 mM AS primer 1 or AS primer 2 labelled with Ned (Life Technologies SAS, France). Reverse transcription was performed as for the SHAPE (+) and (-) elongation reactions except for the reaction mix added which was composed of 6  $\mu$ l G10 (0.25 mM dGTP, 1 mM dATP, 1 mM dCTP, 1 mM TTP), 2  $\mu$ l ddGTP at 100  $\mu$ M, 2  $\mu$ l AMV RT Buffer 5x and 2 U of AMV RT (Life Sciences).

The reaction volumes were adjusted to 100  $\mu$ l and cDNAs were phenol-chloroform extracted (Roti-phenol). For each

experiment, the modified (+) and unmodified (-) samples were pooled each with a ddG sequencing reaction before ethanol precipitation. The cDNAs were resuspended in 10  $\mu$ l HiDi Formamide (Applied Biosystem), denatured at 90°C for 5 min, then placed on ice for 5 min and finally centrifuged for 5 min at 6,000 g. The primer extension products were loaded on an ABI 3130XL Genetic Analyser (Applied Biosystem) and the electropherograms were analysed with the QuShape software [82].

### Data analysis

For pairwise comparisons between the different conditions, we first used deltaSHAPE [83] to identify statistically significant SHAPE differences between two conditions [84]<sup>\*\*\*</sup>. DeltaSHAPE performs a modified Z-factor test that identifies nts in the RNA of interest whose SHAPE differences under the two conditions differ by > 1.96 standard deviations of the SHAPE errors under these two conditions (Z-factor > 0), ensuring that the 95% confidence intervals of each measurements do not overlap [83]. However, not all statistically significant differences may be biologically relevant as neither an absolute nor a relative threshold can be used alone to identify the biologically relevant differences between conditions. Using an absolute difference threshold between two SHAPE datasets retrieves false positives (*i.e.* irrelevant differences) when SHAPE reactivities are high in both conditions and may miss biologically relevant differences (false negative) when the SHAPE reactivities are low in both conditions. On the other side, using a relative difference threshold can generate false positive when SHAPE reactivities are low in both conditions, and the difference between reactivity values is not significant. We therefore considered differences between conditions biologically relevant only if the absolute SHAPE reactivities differed by  $\geq$  0.20 and if the relative difference differed by  $\geq$  40%.

In order to compare the RNA chaperone effect of Pr55<sup>Gag</sup> and its maturation products on the gRNA 1–600 structure 'NoProt' or 'ProtK' SHAPE reactivity values were used as constraints to fold the RNA secondary structure with the RNAstructure software version 6.0. No other constraints than the SHAPE reactivities were applied to the fold. Based on the RNAstructure data, the structure of naked RNA and deproteinated RNAs were drawn using the Structure Editor graphical tool, a module of the RNAstructure software.

### Acknowledgments

We thank Erwan Quignon for his help with bioinformatics tools and Serena Bernacchi for fruitful discussions.

### Author contributions

VVB, RM and RPS conceived the experiments; AM, CT and JM provided reagents; OG, EM and VVB performed experiments; OG, EM, JCP, RPS, RM, and VVB analyzed the data; RM and VVB wrote the paper with contributions from the other authors.

### Disclosure statement

The authors declare no competing interests.

## Data availability statement

All data are available in the main article or in the supplementary materials.

## Funding

This work was supported by grants from the Agence Nationale de la Recherche sur le SIDA et les hépatites virales (ANRS) and from Sidaction to RM and by doctoral fellowships from the French Ministry of Higher Education and Research to EM and OG and from the Fondation de la Recherche Médicale (F.R.M) to EM.

## ORCID

Jean-Christophe Paillart  <http://orcid.org/0000-0003-1647-8917>

Carine Tisé  <http://orcid.org/0000-0001-5534-4650>

Johnson Mak  <http://orcid.org/0000-0002-5229-5707>

Valérie Vivet-Boudou  <http://orcid.org/0000-0001-8702-1047>

## References

- Swanstrom R, Wills JW. Synthesis, assembly, and processing of viral proteins [Internet]. In: Coffin JM, Hughes SH, and Varmus HE, editors. *Retroviruses*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1997. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK19456/>
- Sundquist WI, Kräusslich H-G. HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med*. 2012;2:a006924.
- Freed EO. HIV-1 assembly, release and maturation. *Nat Rev Microbiol*. 2015;13:484–496.
- Kleinpeter AB, Freed EO. HIV-1 maturation: lessons learned from inhibitors. *Viruses*. 2020;121:940. <https://doi.org/10.3390/v12090940>
- Fu W, Gorelick RJ, Rein A. Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J Virol*. 1994;68:5013–5018.
- Jalalirad M, Laughrea M. Formation of immature and mature genomic RNA dimers in wild-type and protease-inactive HIV-1: differential roles of the Gag polyprotein, nucleocapsid proteins NCp15, NCp9, NCp7, and the dimerization initiation site. *Virology*. 2010;407:225–236.
- Lee S-K, Potempa M, Kolli M, et al. 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*. 2012;287:13279–13290.
- Potempa M, Nalivaika E, Ragland D, et al. A direct interaction with RNA dramatically enhances the catalytic activity of the HIV-1 protease in vitro. *J Mol Biol*. 2015;427:2360–2378.
- Schur FKM, Hagen WJH, Rumlová M, et al. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature*. 2015;517(7535):505–8. <https://doi.org/10.1038/nature13838>
- Mattei S, Glass B, Hagen WJH, et al. The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science*. 2016;354(6318):1434–1437.
- Mattei S, Tan A, Glass B, et al. High-resolution structures of HIV-1 Gag cleavage mutants determine structural switch for virus maturation. *Proc Natl Acad Sci USA*. 2018;115:E9401–10.
- Shehu-Xhilaga M, Kraeusslich HG, Pettit S, et al. Proteolytic processing of the p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation. *J Virol*. 2001;75:9156–9164.
- Wu T, Datta SAK, Mitra M, et al. Fundamental differences between the nucleic acid chaperone activities of HIV-1 nucleocapsid protein and Gag or Gag-derived proteins: biological implications. *Virology*. 2010;405:556–567.
- Darlix J-L, de Rocquigny H, Mély Y. The multiple roles of the nucleocapsid in retroviral RNA conversion into proviral DNA by reverse transcriptase. *Biochem Soc Trans*. 2016;44:1427–1440.
- Levin JG, Guo J, Rouzina I, et al. Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog Nucleic Acid Res Mol Biol*. 2005;80:217–286.
- Feng YX, Copeland TD, Henderson LE, et al. HIV-1 nucleocapsid protein induces “maturation” of dimeric retroviral RNA in vitro. *Proc Natl Acad Sci USA*. 1996;93:7577–7581.
- Müller B, Anders M, Akiyama H, et al. HIV-1 Gag processing intermediates trans-dominantly interfere with HIV-1 infectivity. *J Biol Chem*. 2009;284:29692–29703.
- Wieggers K, Rutter G, Kottler H, et al. Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J Virol*. 1998;72:2846–2854.
- Ohishi M, Nakano T, Sakuragi S, et al. The relationship between HIV-1 genome RNA dimerization, virion maturation and infectivity. *Nucleic Acids Res*. 2011;39:3404–3417.
- Song R, Kafaie J, Yang L, et al. HIV-1 viral RNA is selected in the form of monomers that dimerize in a three-step protease-dependent process; the DIS of stem-loop 1 initiates viral RNA dimerization. *J Mol Biol*. 2007;371:1084–1098.
- Mattei S, Flemming A, Anders-Össwein M, et al. RNA and nucleocapsid are dispensable for mature HIV-1 capsid assembly. *J Virol*. 2015;89:9739–9747.
- Kräusslich HG, Fäcke M, Heuser AM, et al. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *J Virol*. 1995;69:3407–3419.
- Berkhout B, Das AT, van Wamel JL. The native structure of the human immunodeficiency virus type 1 RNA genome is required for the first strand transfer of reverse transcription. *Virology*. 1998;249:211–218.
- Paillart JC, Berthoux L, Ottmann M, et al. A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J Virol*. 1996;70:8348–8354.
- Lyonnais S, Sadiq SK, Lorca-Oró C, et al. The HIV-1 ribonucleoprotein dynamically regulates its condensate behavior and drives acceleration of protease activity through membrane-less granular phase-separation [Internet]. 2019. [cited 2021 Sept 23] Available from: <https://www.biorxiv.org/content/10.1101/528638v2>
- Wilkinson KA, Gorelick RJ, Vasa SM, et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol*. 2008;6:e96.
- Watts JM, Dang KK, Gorelick RJ, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*. 2009;460:711–716.
- Pirakitikulr N, Kohlway A, Lindenbach BD, et al. The coding region of the HCV genome contains a network of regulatory RNA structures. *Mol Cell*. 2016;62:111–120.
- Manfredonia I, Nithin C, Ponce-Salvatierra A, et al. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res*. 2020;48:12436–12452.
- Ew AE-W, Smyth RP, Mailler E, et al. Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat Commun*. 2014;5:4304.
- Kenyon JC, Prestwood LJ, Lever AML. A novel combined RNA-protein interaction analysis distinguishes HIV-1 Gag protein binding sites from structural change in the viral RNA leader. *Sci Rep*. 2015;5:14369.
- Damgaard CK, Dyhr-Mikkelsen H, Kjems J. Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions. *Nucleic Acids Res*. 1998;26:3667–3676.

- [33] Zeffman A, Hassard S, Varani G, et al. The major HIV-1 packaging signal is an extended bulged stem loop whose structure is altered on interaction with the Gag polyprotein. *J Mol Biol.* 2000;297:877–893.
- [34] Ding P, Kharytonchik S, Waller A, et al. Identification of the initial nucleocapsid recognition element in the HIV-1 RNA packaging signal. *Proc Natl Acad Sci USA.* 2020;117:17737–17746.
- [35] Paillart J-C, Dettenhofer M, Yu X-F, et al. First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J Biol Chem.* 2004;279:48397–48403.
- [36] Isel C, Ehresmann C, Keith G, et al. Initiation of reverse transcription of HIV-1: secondary structure of the HIV-1 RNA/tRNA (3Lys) (template/primer). *J Mol Biol.* 1995;247:236–250.
- [37] Isel C, Westhof E, Massire C, et al. Structural basis for the specificity of the initiation of HIV-1 reverse transcription. *EMBO J.* 1999;18:1038–1048.
- [38] Sleiman D, Goldschmidt V, Barraud P, et al. Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. *Virus Res.* 2012;169:324–339.
- [39] Beerens N, Groot F, Berkhout B. Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. *J Biol Chem.* 2001;276:31247–31256.
- [40] Beerens N, Berkhout B. Switching the in vitro tRNA usage of HIV-1 by simultaneous adaptation of the PBS and PAS. *RNA.* 2002;8:357–369.
- [41] Smyth RP, Despons L, Huili G, et al. Mutational interference mapping experiment (MIME) for studying RNA structure and function. *Nat Methods.* 2015;12:866–872.
- [42] Smyth RP, Smith MR, Jousset A-C, et al. In cell mutational interference mapping experiment (in cell MIME) identifies the 5' polyadenylation signal as a dual regulator of HIV-1 genomic RNA production and packaging. *Nucleic Acids Res.* 2018;46:e57.
- [43] Lu K, Heng X, Summers MF. Structural determinants and mechanism of HIV-1 genome packaging. *J Mol Biol.* 2011;410:609–633.
- [44] Keane SC, Summers MF. NMR studies of the structure and function of the HIV-1 5'-leader. *Viruses.* 2016;8(12):338. <https://doi.org/10.3390/v8120338>
- [45] Mailler E, Bernacchi S, Marquet R, et al. The Life-Cycle of the HIV-1 Gag-RNA complex. *Viruses.* 2016;8(9):248. <https://doi.org/10.3390/v8090248>
- [46] Harrison GP, Lever AM. The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J Virol.* 1992;66:4144–4153.
- [47] Berkhout B. Structure and function of the human immunodeficiency virus leader RNA. *Prog Nucleic Acid Res Mol Biol.* 1996;54:1–34.
- [48] Harrison GP, Miele G, Hunter E, et al. Functional analysis of the core human immunodeficiency virus type 1 packaging signal in a permissive cell line. *J Virol.* 1998;72:5886–5896.
- [49] Skripkin E, Paillart JC, Marquet R, et al. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc Natl Acad Sci USA.* 1994;91:4945–4949.
- [50] Paillart JC, Skripkin E, Ehresmann B, et al. A loop-loop “kissing” complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci USA.* 1996;93:5572–5577.
- [51] Merino EJ, Wilkinson KA, Coughlan JL, et al. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc.* 2005;127:4223–4231.
- [52] Badorrek CS, Weeks KM. RNA flexibility in the dimerization domain of a gamma retrovirus. *Nat Chem Biol.* 2005;1:104–111.
- [53] Aktar SJ, Jabeen A, Ali LM, et al. SHAPE analysis of the 5' end of the Mason-Pfizer monkey virus (MPMV) genomic RNA reveals structural elements required for genome dimerization. *RNA.* 2013;19:1648–1658.
- [54] Gherghe C, Lombo T, Leonard CW, et al. Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proc Natl Acad Sci USA.* 2010;107:19248–19253.
- [55] Grohman JK, Gorelick RJ, Lickwar CR, et al. A guanosine-centric mechanism for RNA chaperone function. *Science.* 2013;340:190–195.
- [56] Briggs JAG, Simon MN, Gross I, et al. The stoichiometry of Gag protein in HIV-1. *Nat Struct Mol Biol.* 2004;11:672–675.
- [57] Rein A. Retroviral RNA packaging: a review. *Arch Virol Suppl.* 1994;9:513–522.
- [58] Bernacchi S, Abd El-Wahab EW, Dubois N, et al. HIV-1 Pr55Gag binds genomic and spliced RNAs with different affinity and stoichiometry. *RNA Biol.* 2017;14:90–103.
- [59] Muriaux D, Girard PM, Bonnet-Mathonière B, et al. Dimerization of HIV-1 RNA at low ionic strength. An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide. *J Biol Chem.* 1995;270:8209–8216.
- [60] Laughrea M, Jetté L. Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248–271 are dispensable for dimer formation. *Biochemistry.* 1996;35:1589–1598.
- [61] Seif E, Niu M, Kleiman L. In vitro SHAPE analysis of tRNA(Lys3) annealing to HIV-1 genomic RNA in wild type and protease-deficient virus. *Retrovirology.* 2015;12:40.
- [62] Abbink TEM, Berkhout B. A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J Biol Chem.* 2003;278:11601–11611.
- [63] Keane SC, Heng X, Lu K, et al. RNA structure. Structure of the HIV-1 RNA packaging signal. *Science.* 2015;348:917–921.
- [64] Stewart-Maynard KM, Cruceanu M, Wang F, et al. Retroviral nucleocapsid proteins display nonequivalent levels of nucleic acid chaperone activity. *J Virol.* 2008;82:10129–10142.
- [65] Cruceanu M, Urbaneja MA, Hixson CV, et al. Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res.* 2006;34:593–605.
- [66] Berkowitz RD, Ohagen A, Höglund S, et al. Retroviral nucleocapsid domains mediate the specific recognition of genomic viral RNAs by chimeric Gag polyproteins during RNA packaging in vivo. *J Virol.* 1995;69:6445–6456.
- [67] Zhang Y, Barklis E. Nucleocapsid protein effects on the specificity of retrovirus RNA encapsidation. *J Virol.* 1995;69:5716–5722.
- [68] Dupraz P, Spahr PF. Specificity of Rous sarcoma virus nucleocapsid protein in genomic RNA packaging. *J Virol.* 1992;66:4662–4670.
- [69] Holmstrom ED, Liu Z, Nettek D, et al. Disordered RNA chaperones can enhance nucleic acid folding via local charge screening. *Nat Commun.* 2019;10:2453.
- [70] Basu S, Bahadur RP. A structural perspective of RNA recognition by intrinsically disordered proteins. *Cell Mol Life Sci.* 2016;73:4075–4084.
- [71] Chu T-H, Liou A-T, Su P-Y, et al. Nucleic acid chaperone activity associated with the arginine-rich domain of human hepatitis B virus core protein. *J Virol.* 2014;88:2530–2543.
- [72] Nikolaitchik OA, Somoulay X, Rawson JMO, et al. Unpaired guanines in the 5' untranslated region of HIV-1 RNA act synergistically to mediate genome packaging. *J Virol.* 2020;94:e00439–20.
- [73] Wang W, Naiyer N, Mitra M, et al. Distinct nucleic acid interaction properties of HIV-1 nucleocapsid protein precursor NCp15 explain reduced viral infectivity. *Nucleic Acids Res.* 2014;42:7145–7159.
- [74] Zila V, Margiotta E, Turoňová B, et al. Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell.* 2021;184:1032–1046.e18.

- [75] Li C, Burdick RC, Nagashima K, et al. HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. *Proc Natl Acad Sci U S A*. 2021;118.e2019467118.
- [76] Larsen KP, Mathiharan YK, Kappel K, et al. Architecture of an HIV-1 reverse transcriptase initiation complex. *Nature*. 2018;557:118–122.
- [77] McKinstry WJ, Hijnen M, Tanwar HS, et al. Expression and purification of soluble recombinant full length HIV-1 Pr55 (Gag) protein in *Escherichia coli*. *Protein Expr Purif*. 2014;100:10–18.
- [78] Lee BM, De Guzman RN, Turner BG, et al. Dynamical behavior of the HIV-1 nucleocapsid protein. *J Mol Biol*. 1998;279:633–649.
- [79] De Guzman RN, Wu ZR, Stalling CC, et al. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science*. 1998;279:384–388.
- [80] Mouhand A, Belfetmi A, Catala M, et al. Modulation of the HIV nucleocapsid dynamics finely tunes its RNA-binding properties during virion genesis. *Nucleic Acids Res*. 2018;46:9699–9710.
- [81] Sinck L, Richer D, Howard J, et al. In vitro dimerization of human immunodeficiency virus type 1 (HIV-1) spliced RNAs. *RNA*. 2007;13:2141–2150.
- [82] Karabiber F, McGinnis JL, Favorov OV, et al. QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA*. 2013;19:63–73.
- [83] Smola MJ, Calabrese JM, Weeks KM. Detection of RNA-Protein interactions in living cells with SHAPE. *Biochemistry*. 2015;54:6867–6875.
- [84] Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*. 2010;11:129.



## A NoProt

NoProt	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	#25	#26	#27	#28	#29	#30	#31	#32	#33	#34	Mean
Exp. #1	1	0.915	0.927	0.933	0.917	0.932	0.916	0.855	0.907	0.908	0.939	0.828	0.875	0.897	0.933	0.850	0.855	0.793	0.913	0.905	0.815	0.864	0.888	0.886	0.869	0.850	0.854	0.878	0.878	0.848	0.857	0.905	0.835	0.885	0.945
Exp. #2		1	0.936	0.912	0.894	0.961	0.949	0.837	0.923	0.937	0.916	0.861	0.884	0.908	0.899	0.875	0.876	0.830	0.910	0.911	0.844	0.843	0.897	0.873	0.889	0.861	0.874	0.808	0.882	0.848	0.854	0.900	0.880	0.875	0.950
Exp. #3			1	0.927	0.918	0.965	0.950	0.837	0.872	0.903	0.927	0.852	0.911	0.884	0.925	0.884	0.889	0.860	0.923	0.919	0.830	0.857	0.888	0.883	0.899	0.843	0.900	0.844	0.900	0.829	0.861	0.924	0.872	0.871	0.954
Exp. #4				1	0.941	0.918	0.946	0.884	0.901	0.911	0.918	0.850	0.852	0.899	0.926	0.869	0.863	0.829	0.918	0.906	0.818	0.883	0.904	0.895	0.867	0.853	0.833	0.829	0.848	0.853	0.832	0.884	0.867	0.889	0.945
Exp. #5					1	0.883	0.934	0.890	0.880	0.918	0.919	0.868	0.849	0.892	0.908	0.844	0.832	0.798	0.892	0.887	0.837	0.901	0.908	0.903	0.872	0.853	0.846	0.837	0.836	0.854	0.850	0.877	0.885	0.872	0.940
Exp. #6						1	0.938	0.819	0.898	0.913	0.933	0.852	0.908	0.901	0.916	0.867	0.889	0.834	0.911	0.905	0.824	0.838	0.882	0.870	0.876	0.849	0.893	0.810	0.905	0.847	0.866	0.918	0.852	0.861	0.948
Exp. #7							1	0.870	0.903	0.944	0.918	0.857	0.874	0.911	0.920	0.886	0.901	0.860	0.934	0.933	0.841	0.881	0.913	0.893	0.907	0.847	0.882	0.832	0.875	0.864	0.864	0.912	0.905	0.891	0.961
Exp. #8								1	0.847	0.874	0.845	0.815	0.797	0.899	0.850	0.837	0.784	0.759	0.855	0.827	0.790	0.880	0.884	0.906	0.820	0.850	0.780	0.768	0.805	0.878	0.829	0.828	0.821	0.867	0.900
Exp. #9									1	0.890	0.912	0.820	0.819	0.879	0.908	0.809	0.805	0.755	0.869	0.877	0.805	0.847	0.887	0.862	0.844	0.833	0.814	0.831	0.831	0.843	0.840	0.869	0.858	0.864	0.920
Exp. #10										1	0.898	0.870	0.868	0.947	0.920	0.853	0.871	0.800	0.900	0.904	0.858	0.897	0.927	0.904	0.911	0.905	0.878	0.804	0.882	0.887	0.881	0.911	0.901	0.881	0.957
Exp. #11											1	0.824	0.849	0.899	0.938	0.849	0.833	0.778	0.901	0.897	0.810	0.869	0.891	0.878	0.853	0.838	0.844	0.861	0.866	0.859	0.856	0.891	0.853	0.884	0.939
Exp. #12												1	0.880	0.873	0.835	0.834	0.834	0.796	0.847	0.843	0.875	0.842	0.873	0.894	0.872	0.882	0.882	0.787	0.855	0.855	0.859	0.861	0.841	0.810	0.912
Exp. #13													1	0.891	0.855	0.849	0.874	0.836	0.875	0.879	0.870	0.839	0.869	0.901	0.894	0.884	0.952	0.839	0.936	0.845	0.874	0.945	0.836	0.835	0.934
Exp. #14														1	0.890	0.851	0.849	0.795	0.892	0.887	0.860	0.890	0.918	0.921	0.878	0.909	0.881	0.803	0.899	0.916	0.877	0.916	0.870	0.890	0.951
Exp. #15															1	0.811	0.820	0.764	0.902	0.910	0.807	0.872	0.908	0.881	0.875	0.860	0.850	0.870	0.884	0.839	0.840	0.909	0.877	0.877	0.941
Exp. #16																1	0.925	0.933	0.895	0.879	0.849	0.840	0.867	0.877	0.878	0.825	0.858	0.783	0.848	0.861	0.865	0.867	0.848	0.873	0.920
Exp. #17																	1	0.937	0.903	0.884	0.841	0.829	0.862	0.865	0.908	0.828	0.893	0.790	0.881	0.849	0.878	0.898	0.850	0.842	0.924
Exp. #18																		1	0.883	0.868	0.811	0.799	0.841	0.825	0.861	0.783	0.850	0.749	0.843	0.809	0.827	0.862	0.847	0.838	0.886
Exp. #19																			1	0.975	0.851	0.876	0.930	0.890	0.889	0.849	0.887	0.870	0.912	0.873	0.864	0.925	0.900	0.930	0.959
Exp. #20																				1	0.851	0.887	0.934	0.885	0.909	0.861	0.890	0.889	0.899	0.859	0.849	0.931	0.917	0.948	0.958
Exp. #21																					1	0.848	0.861	0.876	0.888	0.871	0.898	0.796	0.859	0.843	0.858	0.875	0.838	0.836	0.906
Exp. #22																						1	0.943	0.925	0.887	0.896	0.841	0.829	0.853	0.903	0.900	0.880	0.889	0.912	0.935
Exp. #23																							1	0.939	0.902	0.909	0.875	0.840	0.907	0.923	0.894	0.927	0.943	0.915	0.963
Exp. #24																								1	0.903	0.903	0.883	0.842	0.885	0.921	0.926	0.917	0.867	0.878	0.954
Exp. #25																									1	0.895	0.919	0.859	0.888	0.853	0.884	0.923	0.890	0.885	0.947
Exp. #26																										1	0.891	0.807	0.901	0.899	0.885	0.895	0.857	0.856	0.929
Exp. #27																											1	0.849	0.943	0.867	0.889	0.953	0.850	0.846	0.938
Exp. #28																												1	0.837	0.798	0.828	0.869	0.812	0.856	0.930
Exp. #29																													1	0.884	0.880	0.959	0.867	0.863	0.943
Exp. #30																														1	0.920	0.896	0.848	0.877	0.929
Exp. #31																															1	0.902	0.842	0.845	0.930
Exp. #32																																1	0.892	0.899	0.966
Exp. #33																																	1	0.879	0.930
Exp. #34																																		1	0.937
Mean																																			1

## B Pr55<sup>Gag</sup>

Complex	#1	#2	#3	#4	#5	Mean
Exp. #1	1	0.953	0.818	0.928	0.904	0.965
Exp. #2		1	0.766	0.900	0.871	0.941
Exp. #3			1	0.884	0.918	0.920
Exp. #4				1	0.956	0.976
Exp. #5					1	0.973
Mean						1

AT-2	#1	#2	Mean
#1	1	0.937	0.983
#2		1	0.985
Mean			1

ProtK	#1	#2	#3	#4	#5	Mean
Exp. #1	1	0.944	0.908	0.914	0.819	0.967
Exp. #2		1	0.927	0.911	0.869	0.979
Exp. #3			1	0.860	0.830	0.952
Exp. #4				1	0.809	0.950
Exp. #5					1	0.904
Mean						1

## C NCp15

Complex	#1	#2	#3	#4	#5	#6	Mean
Exp. #1	1	0.882	0.757	0.833	0.859	0.845	0.906
Exp. #2		1	0.751	0.860	0.879	0.857	0.916
Exp. #3			1	0.897	0.900	0.887	0.921
Exp. #4				1	0.979	0.978	0.976
Exp. #5					1	0.987	0.986
Exp. #6						1	0.977
Mean							1

AT-2	#1	#2	#3	Mean
Exp. #1	1	0.952	0.892	0.977
Exp. #2		1	0.873	0.968
Exp. #3			1	0.959
Mean				1

ProtK	#1	#2	#3	#4	#5	Mean
Exp. #1	1	0.867	0.750	0.801	0.790	0.900
Exp. #2		1	0.822	0.821	0.827	0.928
Exp. #3			1	0.925	0.920	0.940
Exp. #4				1	0.968	0.961
Exp. #5					1	0.959
Mean						1

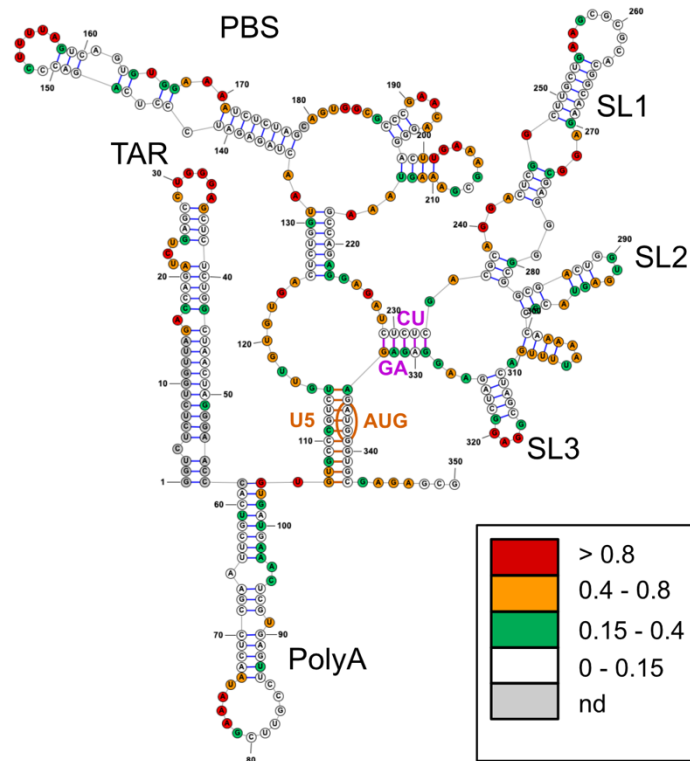
## D NCp9

Complex	#1	#2	#3	Mean
Exp. #1	1	0.714	0.732	0.909
Exp. #2		1	0.710	0.875
Exp. #3			1	0.917
Mean				1

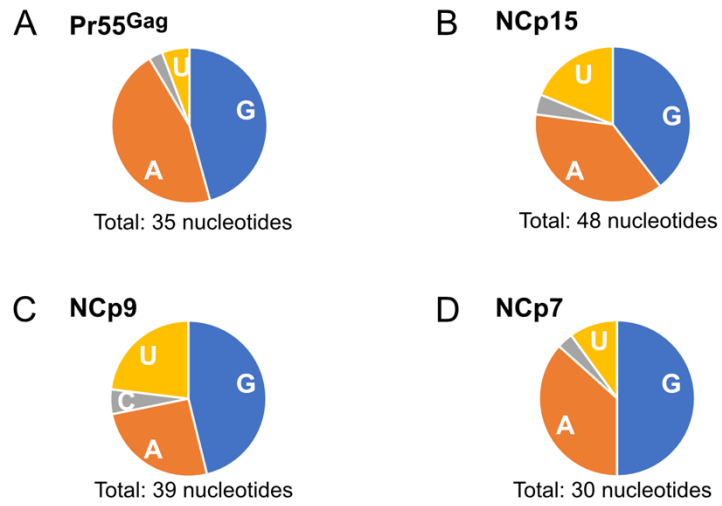
AT2	#1	#2	#3	Mean
Exp. #1	1	0.808	0.836	0.935
Exp. #2		1	0.877	0.943
Exp. #3			1	0.957
Mean				1

ProtK
-------

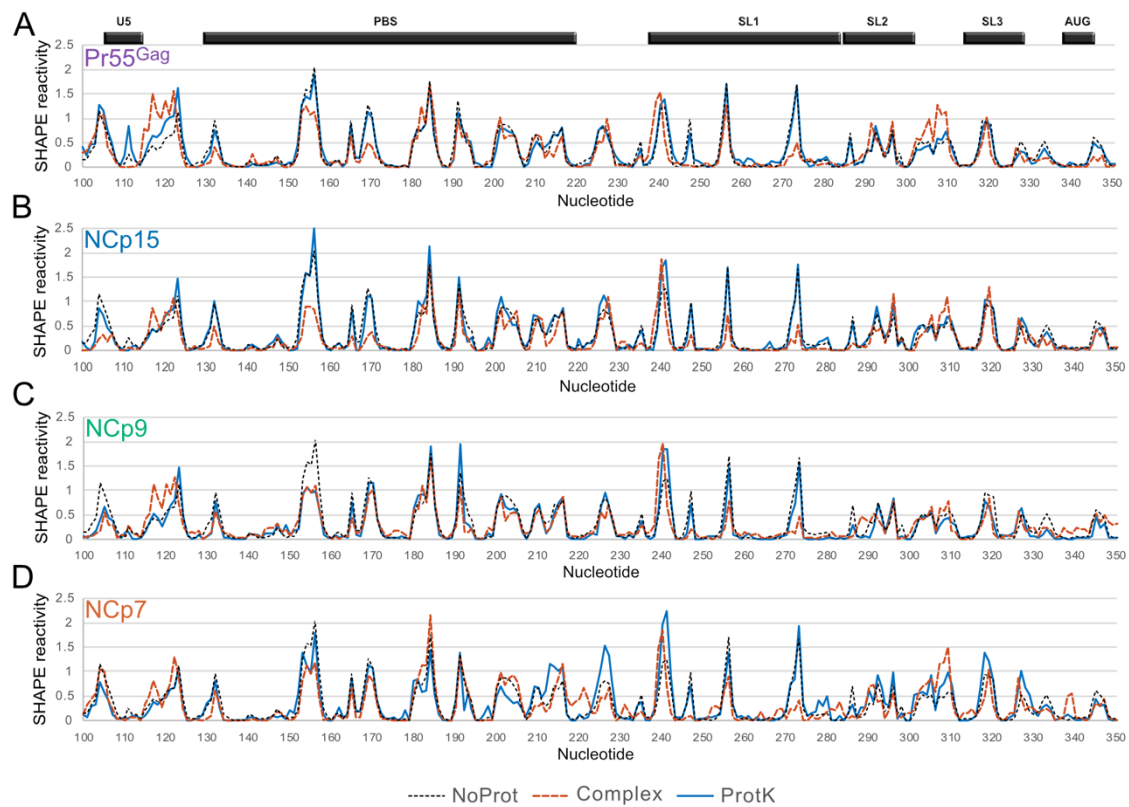
are presented. In each table, the columns and lines correspond to individual replicates, except for the last columns and lines, which correspond to the mean value of the replicates.



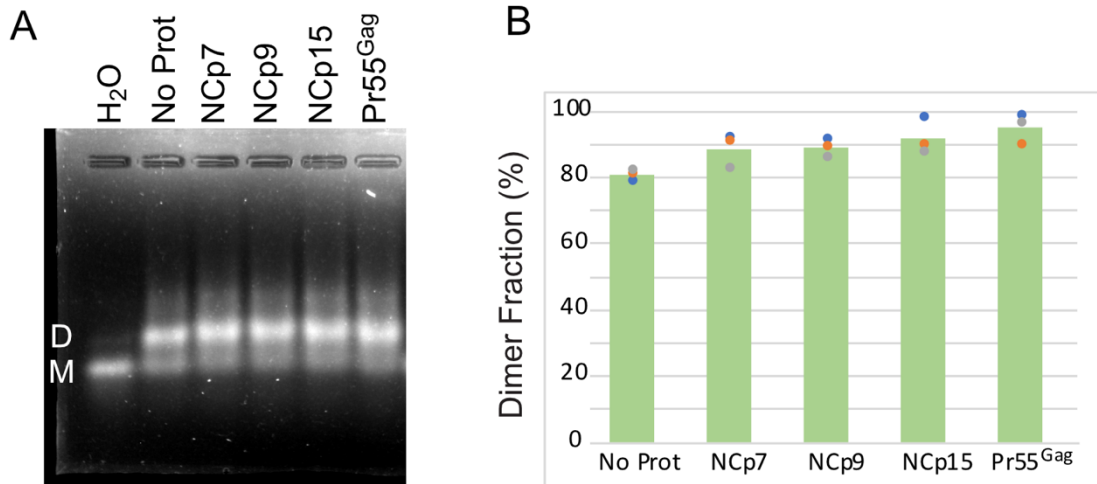
**Supplementary Figure 2: Secondary structure model of the 5' region of the “naked” HIV-1 gRNA.** This structure was obtained using the mean SHAPE reactivity values of the *NoProt* condition (mean values of 34 data sets) as constraints for the RNAstructure software (version 6.0). The SHAPE reactivity values are drawn on the structure and color-coded as indicated in the insert. The U5:AUG and CU:GA long distance interactions are indicated in orange and purple, respectively.



**Supplementary Figure 3: Nucleotide composition of the regions protected by Pr55<sup>Gag</sup> (A), NCp15 (B), NCp9 (C), and NCp7 (D).**



**Supplementary Figure 4: SHAPE reactivity profiles under the *NoProt*, *Complex*, and *ProtK* conditions for Pr55<sup>Gag</sup> (A), NCp15 (B), NCp9 (C), and NCp7 (D).**



**Supplementary Figure 5: Fraction of RNA dimer after incubation without or with protein.**

gRNA 1-600 was incubated in the absence of protein or in the presence of Pr55<sup>Gag</sup>, NCp15, NCp9 or NCp7 as described for the *ProtK* condition (see material and methods). After removal of the proteins by proteinase K treatment, the samples were analyzed on 1% agarose gels run at 4 °C in Tris-Borate -Magnesium buffer. Experiments were run in triplicates. One of the gels stained with ethidium bromide is shown in (A). M and D stand for dimer and monomer, respectively. RNA dissolved in water was used as a control for monomeric RNA. Quantification of the Dimer Fraction is shown in (B). Fraction Dimer =  $100 \cdot [D / (M + D)]$ .

## Supplementary Dataset 1

See accompanying excel file

**Legend to Supplementary Dataset 1. Sheet 1:** SHAPE reactivities obtained under the “*NoProt*” condition. **Sheet 2:** SHAPE reactivities obtained in the presence of Pr55<sup>Gag</sup> under the *Complex ProtK*, and *AT-2* conditions. **Sheet 3:** SHAPE reactivities obtained in the presence of NCp15 under the *Complex ProtK*, and *AT-2* conditions. **Sheet 3:** SHAPE reactivities obtained in the presence of NCp9 under the *Complex ProtK*, and *AT-2* conditions. **Sheet 4:** SHAPE reactivities obtained in the presence of NCp7 under the *Complex ProtK*, and *AT-2* conditions. On each sheet, the first column corresponds to the nucleotide position, and for each condition the following columns correspond to different replicates, except the last column, which corresponds to the mean of all replicates obtained under the specified condition. **Sheet 5** summarizes the mean SHAPE values obtained without protein (*NoProt*) and with Pr55<sup>Gag</sup>, NCp15.

## IV/ Analyses complémentaires

Les résultats présentés dans notre article mettent en évidence l'effet propre de la fixation de chacune des protéines contenant le domaine NC à l'ARNg 1-604. Ainsi, les protéines Pr55<sup>Gag</sup> et NCp15 semblent posséder la plus forte capacité de fixation à l'ARN. Cette fixation ne s'accompagne pas de réarrangements structuraux conséquents, comme déjà rapporté par Carlson *et al.*, (2016) (235). A l'inverse, les protéines NCp9 et NCp7, bien que plus discrètes au niveau de leurs sites de fixation, ont des effets déstabilisateurs de la structure initiale bien plus marqués (72). De plus, les expériences en présence d'aldrithiol-2 (AT-2) ont permis de montrer l'implication des domaines en doigts de zincs dans le maintien de la fixation. Ainsi, il a été conclu que les doigts de zinc des différentes protéines avaient une importance croissante au cours de la maturation. Cependant, par manque de place, certaines analyses complémentaires n'ont pas été intégrées dans le manuscrit. Dans cette partie, je vais donc compléter l'analyse des données obtenues par des tests statistiques supplémentaires et je vais présenter une étude comparative plus poussée avec la littérature.

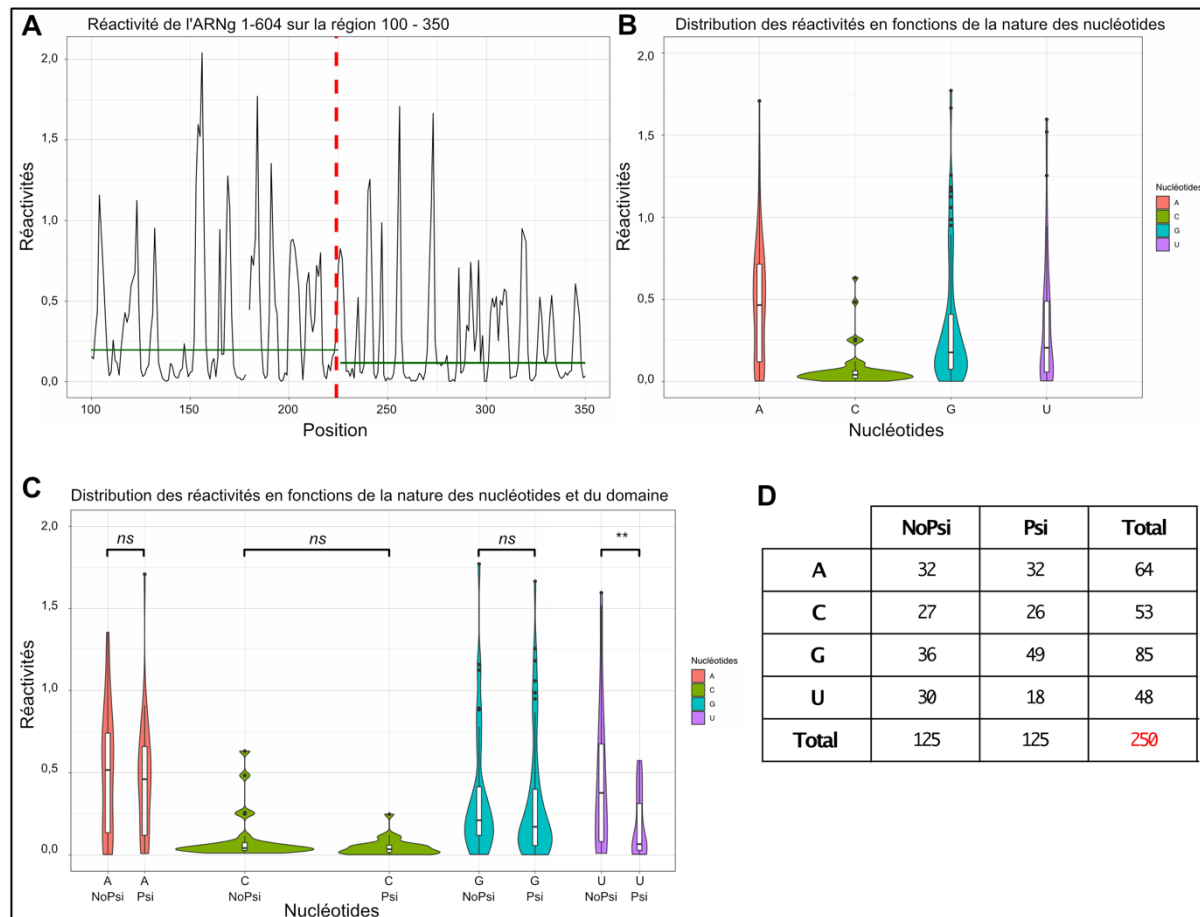
### IV.1/Analyse globale des réactivités associées à l'ARNg 1-604

L'analyse globale des réactivités SHAPE obtenues pour l'ARNg 1-604 en absence de protéine (condition « NoProt ») ainsi qu'une analyse plus précise du domaine Psi précédemment identifié au laboratoire entre les nucléotides 227 et 337 (57) ont été réalisées sur R (**Figure 28**). Pour ce faire, j'ai séparé les réactivités SHAPE en deux jeux de données distincts, mais de tailles similaires, SHAPE\_NoPsi, entre les nucléotides 100 et 225, et SHAPE\_Psi, entre les nucléotides 225 et 350 (**Figure 28A**, séparation rouge). Les distributions des réactivités SHAPE de ces deux domaines ont ensuite été analysées ensemble (**Figure 28B**) ou séparément (**Figure 28C**). Globalement, ces réactivités semblent légèrement plus faibles dans le domaine Psi que dans le domaine NoPsi (**Figure 28A**), ce qui est confirmé par un test de Wilcoxon unilatéral ( $p$ -value = 0,02114). Une analyse plus fine de la réactivité en fonction de la nature du nucléotide montre que cette différence entre les deux régions est principalement due à la diminution marquée ( $p$ -value = 0,009302) des réactivités associées aux uridines dans le domaine Psi (**Figure 28C**), mais également au fait que les uridines sont



## AXE 1 – RESULTATS

relativement moins présents dans cette région Psi (**Figure 28D**). Cependant, il n’y a pas de différences significatives de réactivités associées aux autres nucléotides entre ces deux domaines. Au niveau de la structure de l’ARNg 1-604 en absence de protéines, ces observations se traduisent par un nombre légèrement plus élevé de nucléotides simples brins dans le domaine NoPsi (62 sb vs 47 sb, **Figure 2**, papier).



**Figure 28** : Distributions des réactivités SHAPE de l’ARNg 1-604 en absence de protéine. (A) Les réactivités sont représentées en fonction de leur position dans l’ARN, la ligne rouge représente la séparation entre les domaines NoPsi (gauche) et Psi (droite). Les lignes vertes indiquent la médiane pour chaque domaine. (B et C) Les distributions des réactivités en fonction de la nature des nucléotides sont représentées pour (B) l’ARN complet ou (C) séparé en domaines NoPsi et Psi. (D) Le tableau récapitule le nombre et la nature des nucléotides par domaine.

### IV.2/Analyse globale des réactivités associées à l’ARNg 1-604 en présence de protéines

L’analyse est poursuivie en comparant les profils globaux de l’ARNg 1-604, entre la condition « NoProt » et chacune des conditions en complexe avec les protéines. Pour ce faire,

## AXE 1 – RESULTATS

j'ai soumis l'ensemble de ces conditions à des tests de Wilcoxon unilatéraux, afin de déterminer si des différences significatives de réactivités sont présentes et quelles sont leur portée. Les résultats de ces tests sont présentés dans le **Tableau 6**. Premièrement, la protéine NCp15 semble avoir les effets les plus importants parmi les quatre protéines testées. En effet, une diminution significative de la réactivité sur trois des quatre nucléotides (A, C et G) est observée par rapport à la condition NoProt. Deuxièmement, la fixation des quatre protéines semble principalement affecter les guanines avec une diminution significative dans le cas des protéines Pr55<sup>Gag</sup>, NCp15 et NCp7. Finalement, la fixation de la protéine NCp9 semble induire une augmentation des réactivités associées à des C. Les nucléotides U ne montrent pas de différences significatives en complexe avec les protéines. Les résultats de ces tests statistiques sont en accord avec les observations rapportées dans notre article en partie III. En effet, l'impact de chacune des protéines testées ici est notablement distinct des autres. De plus, une spécificité pour les guanines semble apparaître avec un effet similaire observé pour 3 des 4 protéines. Cette spécificité peut être expliquée par le rôle central que jouent les guanines dans la reconnaissance de l'ARNg par Pr55<sup>Gag</sup> lors de la sélection et de la formation de la particule immature (228), ainsi que dans l'activité chaperon d'ARN (263).

	Tests de Wilcoxon entre conditions NoProt et Protéine								
	Pr55Gag		NCp15		NCp9		NCp7		
	Bilatéral	Unilatéral	Bilatéral	Unilatéral	Bilatéral	Unilatéral	Bilatéral	Unilatéral	
<b>A</b>	ns	ns	0,03415	0,01707	ns	ns	ns	ns	
<b>C</b>	ns	ns	0,04216	0,0218	0,000105	5,26E-05	ns	ns	
<b>G</b>	0,02172	0,01086	0,000183	9,17E-05	ns	ns	ns	0,04831	
<b>U</b>	ns	ns	ns	ns	ns	ns	ns	ns	

Différent
Plus grand
Plus petit

**Tableau 6 :** Tests de Wilcoxon entre la condition « NoProt » et les conditions en présence de protéines pour l'ARNg 1-604 en fonction de la nature des nucléotides. Les p-values significatives sont représentées dans des cases colorées.

### IV.3/Analyse spécifique des réactivités associées à l'ARNg 1-604 en présence de protéines et en fonction des domaines

Afin d'affiner l'analyse des réactivités et de leur distribution en fonction de la nature nucléotidique, j'ai également réalisé des tests de Wilcoxon unilatéraux en comparant les

## AXE 1 – RESULTATS

profils de réactivités des domaines NoPsi et Psi lorsque les complexes étaient étudiés (**Tableaux 7A et 7B**). Dans un premier temps, il est intéressant de noter la différence de répartition des changements significatifs de réactivités entre les deux domaines. En effet, des répartitions différentes sont observées pour les protéines à l'exception de NCp9, qui semble augmenter les réactivités associées aux cytosines de manière globale et non spécifique. Pour les protéines Pr55<sup>Gag</sup> et NCp15, la diminution de la réactivité des guanines qui avait été identifiée pour l'ARN entier est retrouvée, avec une spécificité pour le domaine Psi. En effet, seules les guanines sont affectées par ces deux protéines dans le domaine Psi, alors que dans le domaine NoPsi les nucléotides A, C et G sont affectés.

A	Tests de Wilcoxon entre conditions NoProt et Protéine - Domaine NoPsi							
	Pr55Gag		NCp15		NCp9		NCp7	
	Bilatéral	Unilatéral	Bilatéral	Unilatéral	Bilatéral	Unilatéral	Bilatéral	Unilatéral
A	ns	0,04345	ns	0,02679	ns	ns	ns	ns
C	ns	ns	0,03612	0,01806	0,00418	0,00209	0,001884	0,00942
G	ns	ns	0,0126	0,006298	ns	ns	ns	ns
U	ns	ns	ns	ns	ns	ns	ns	ns

B	Tests de Wilcoxon entre conditions NoProt et Protéine - Domaine Psi							
	Pr55Gag		NCp15		NCp9		NCp7	
	Bilatéral	Unilatéral	Bilatéral	Unilatéral	Bilatéral	Unilatéral	Bilatéral	Unilatéral
A	ns	ns	ns	ns	ns	ns	ns	ns
C	ns	ns	ns	ns	0,008175	0,004088	ns	ns
G	0,03867	0,01934	0,004791	0,002395	ns	ns	ns	ns
U	ns	ns	ns	ns	ns	ns	ns	ns

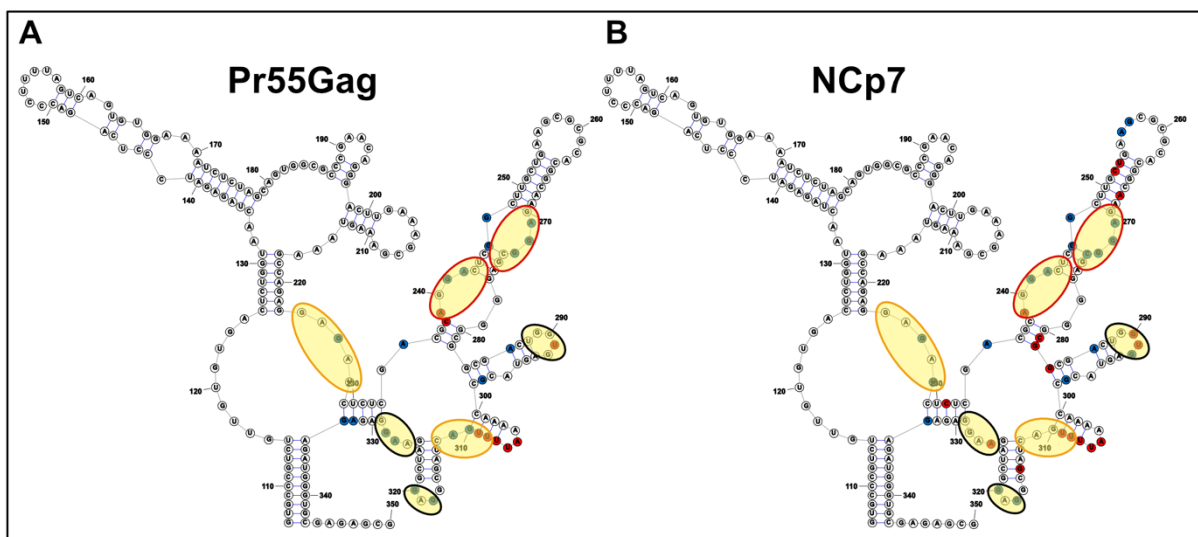
Différent
Plus grand
Plus petit

**Tableau 7 :** Tests de Wilcoxon entre la condition « NoProt » et les conditions en présence de protéines pour l'ARNg 1-604 en fonction de la nature des nucléotides et des domaines (A) NoPsi et (B) Psi. Les p-values significatives sont représentées dans des cases colorées.

En revanche, la diminution des réactivités des guanines associées à NCp7 sur l'ARN entier (**Tableau 6**), n'est plus significative lorsque l'ARN est séparé en domaines (**Tableau 7**). Ces résultats suggèrent que l'impact de NCp7 n'est pas localisé sur une région particulière, mais possède un effet global, non spécifique sur l'ARN. Finalement, nous pouvons remarquer grâce à ces tests que les uridines ne sont impactées par aucune des protéines.

IV.4/Comparaison avec la littérature

Les résultats obtenus dans notre étude semblent en accord avec l'idée générale de la fixation des protéines contenant le domaine NC rapportés dans la littérature. En effet, les protéines Pr55<sup>Gag</sup> semblent impacter plus spécifiquement la région Psi que les autres protéines. Cependant, pour vérifier la véracité de cette affirmation, j'ai décidé d'approfondir mon analyse en comparant les résultats obtenus lors de ma thèse avec différentes études rapportées dans la littérature. Pour ce faire, les sites de diminution significative de réactivités identifiées pour les protéines Pr55<sup>Gag</sup> et NCp7 ont été comparés aux sites de fixation proposés dans la littérature (**Figures 29 et 30**). Une étude parue en 2020 s'est intéressée aux guanines présentes dans le domaine Psi et à leur rôle dans l'encapsidation (119). Les résultats indiquent que la mutation de chacune de ces guanines a un impact, plus ou moins fort, sur l'encapsidation, sans impacter de manière trop importante la structure du domaine Psi.



**Figure 29** : Comparaison des données obtenues en hSHAPE aux sites de fixation proposés par Nikolaitchik *et al.*, (2020). Les diminutions (bleu) et les augmentations (rouge) de réactivités en présence des protéines (A) Pr55<sup>Gag</sup> et (B) NCp7 sont représentées. Les zones jaunes rapportent les sites de fixation proposés par Nikolaitchik *et al.*, (2020). Les contours rouges, oranges et noirs des zones jaunes indiquent les sites de fixation primaires, secondaires et tertiaires, respectivement.

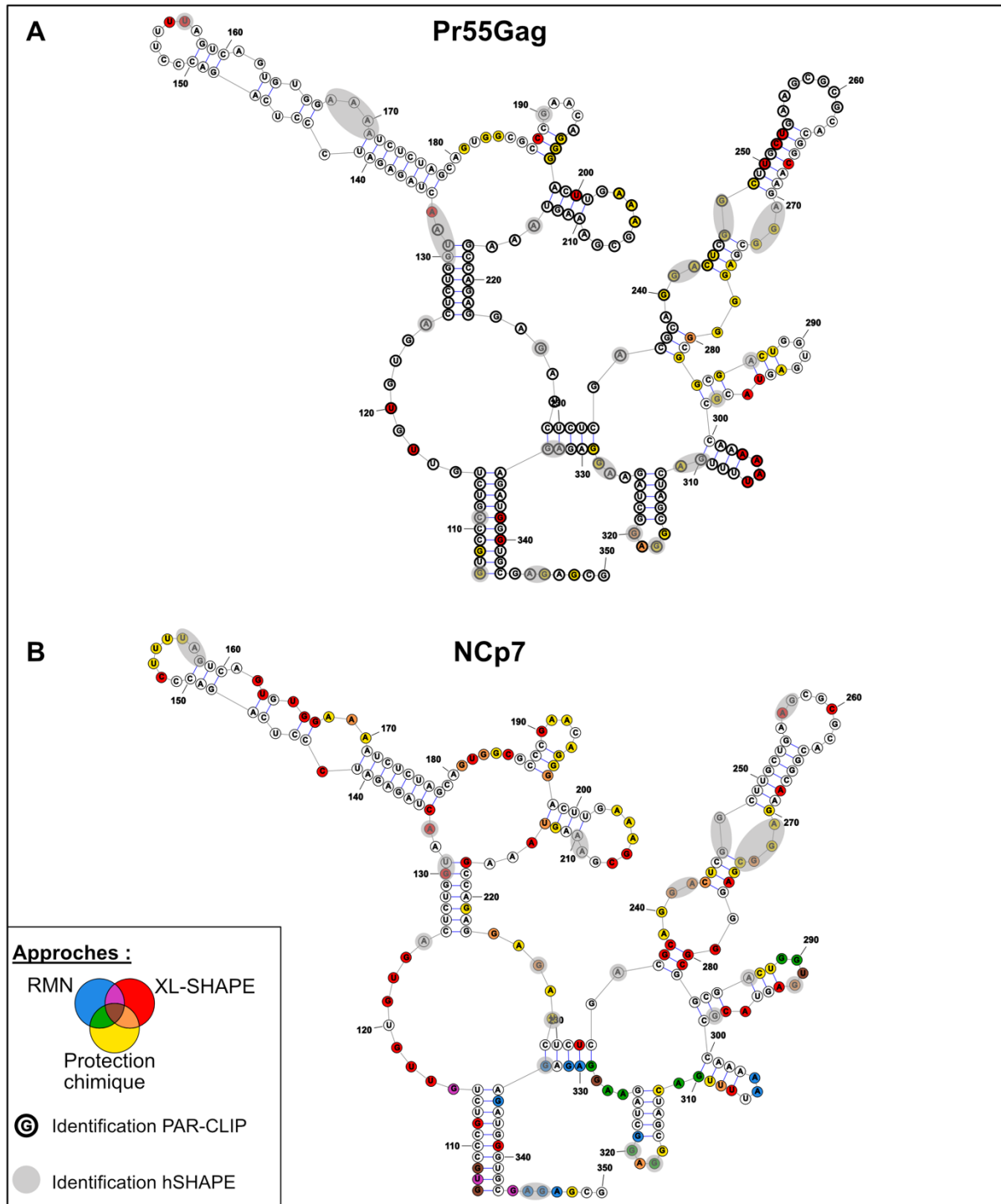
Après m'être assuré *in silico* que les mutations n'affectent pas la structure bidimensionnelle du domaine Psi, j'ai donc comparé les sites de diminution de réactivités induits par la présence des protéines Pr55<sup>Gag</sup> et NCp7 que j'ai identifiés dans mon étude (nucléotides bleus) aux sites de fixation proposés dans l'étude de Nikolaitchik *et al.*, (2020)

(**Figure 29A** et **29B** - zones jaunes). Les résultats indiquent une excellente corrélation puisque les guanines de 6 des 7 sites de fixation de Pr55<sup>Gag</sup> sont retrouvés dans notre étude ainsi que 5 des 7 sites de fixation de NCp7.

Je me suis également intéressé à d'autres données de la littérature obtenues à partir d'expériences de CLIP ou de cartographie chimiques sur les complexes entre l'ARNg du VIH-1 et les protéines Pr55<sup>Gag</sup> (118,201,228) ou son mutant Pr55<sup>GagΔp6</sup> (247,264), ainsi que NCp7 (**Figure 30**). Pour NCp7, des approches de RMN (120,212,225–227), XL-SHAPE (244,264) et de protection chimique ou enzymatique (201,220) ont été réalisées.

Globalement, les analyses des sites de fixation ou de protection par Pr55<sup>Gag</sup> / Pr55<sup>GagΔp6</sup> montrent que la région Psi est favorisée. En effet, 18 nucléotides sont identifiés dans la région NoPsi contre 43 dans la région Psi, soit 70% (**Figure 30A**). Ces résultats sont cohérents avec ceux obtenus dans notre étude où 57% des positions protégées sont retrouvées dans le domaine Psi pour Pr55<sup>Gag</sup>. Cependant, les sites identifiés dans les différentes études ne se superposent pas toujours entre eux. Dans le cas des sites de protection en présence de Pr55<sup>Gag</sup> (**Figure 30A** - jaune), les résultats des études de Damgaard *et al.*, (1998) et de Abd El-Wahab *et al.*, (2014) se recouvrent relativement bien malgré l'utilisation d'une protéine Pr55<sup>Gag</sup> taguée GST par Damgaard *et al.*, (1998) (118,201). De même, les résultats obtenus dans notre étude corrélaient relativement bien avec ces études précédentes (**Figure 30A** - zones grisées). A l'inverse, les résultats de ces études corrélaient extrêmement mal avec ceux des études de D'Souza et de Kenyon qui utilisent une approche de cross-linking de la protéine mutante Pr55<sup>GagΔp6</sup> (247,264) (**Figure 30A** - orange). En effet, bien que la majorité des régions identifiées soit identique, les positions nucléotidiques varient énormément. De plus, il est intéressant de noter que les travaux de D'Souza *et al.*, (2021) n'identifient qu'un unique site de fixation de Pr55<sup>GagΔp6</sup> (A133) sur la région étudiée. Cette observation est d'autant plus étrange que le protocole, les protéines et l'ARN utilisés sont très semblables à ceux de la publication de Kenyon *et al.*, (2015). Dans le cas des expériences de CLIP réalisées par Kutluay *et al.*, (2014), la majorité de la région 5' terminale est retrouvée (228). En effet, seules la région PBS et une partie de SL1 et de SL2 manquent. En conclusion, parmi l'ensemble des nucléotides identifiés dans la littérature, nos travaux permettent de retrouver la majorité des nucléotides identifiés par les approches de cartographie chimique utilisant une protéine Pr55<sup>Gag</sup> non délétée du domaine p6.

D'un autre côté, l'analyse des effets de la fixation de la protéine NCp7 montre que la favorisation du domaine Psi est perdue, avec 54% des nucléotides identifiés se situant dans ce domaine (**Figure 30B**). Ainsi, la spécificité pour cette région de la protéine Pr55<sup>Gag</sup> n'apparaît plus autant pour NCp7, ce qui est en adéquation avec les résultats statistiques décrit plus haut qui semblaient pointer vers un effet global de la protéine sur la structure (**Tableaux 4 et 5**). Les autres résultats obtenus en RMN ont permis d'identifier des sites de fixation de la NCp7 sur SL2, le SL2-SL3 linker, SL3, l'interaction CU:GA et la tige-boucle métastable SL4. Parmi ces sites, seuls ceux présents sur SL3, l'interaction CU:GA et la tige-boucle métastable SL4 ont été retrouvés. Cependant, ces approches de RMN ont été contraintes à des analyses sur de petites régions de l'ARN, ce qui peut expliquer certaines des disparités observées. A l'inverse, dans l'étude de Ding *et al.*, (2020), plus de deux douzaines de sites de fixation de NCp7 sont identifiés sur un ARN de 344 nucléotides par RMN (120). Ces résultats sont cohérents avec les expériences de XL-SHAPE qui ont permis de mettre en évidence 61 nucléotides affectés par la fixation de NCp7 (244,264). Dans les expériences de Liu *et al.*, (2021), une diminution de réactivité est observée pour les positions « G241 ; A242 ; G246 ; A271 ; G272 ; G273 ; G279 ; G280 ; G289 ; G290 ; G292 ; G310 ; G312 », ce qui est cohérent avec les résultats obtenus par nos expériences de SHAPE en présence de NCp7 (zones grises – **Figure 30B**). Cependant, l'ARN utilisé dans l'étude était incapable de dimériser (GAGA en boucle apicale de SL1), ce qui peut expliquer certaines différences. Finalement, les expériences de protections chimiques ou enzymatiques montrent des effets de la protéine NCp7 sur les régions U5, PBS, Psi de façon générale (201,220).



**Figure 30** : Comparaison des données obtenues en hSHAPE à des données issues de la littérature. Les résultats issus d'expériences de cartographie chimique, de RMN ou de XL-SHAPE en présence de (A) Pr55<sup>Gag/GagΔp6</sup> ou de (B) NCp7 sont représentés. Pour Pr55<sup>Gag</sup>, des expériences de CLIP ont également été réalisées (228). Les zones grisées correspondent aux diminutions de réactivités identifiées dans nos expériences (72).

## Discussion :

L'objectif de ce premier axe de ma thèse a été d'étudier la structure de la région 5' terminale de l'ARNg du VIH-1 et les réarrangements structuraux induits par la fixation des protéines Pr55<sup>Gag</sup>, NCp15, NCp9 et NCp7. Ces protéines jouent un rôle essentiel lors des étapes tardives du cycle, avec Pr55<sup>Gag</sup> intervenant dans la sélection, l'encapsidation ou encore la maturation en coordination avec les protéines NCp15, NCp9 et NCp7. En effet, la maturation génomique se produit concomitamment à l'étape de maturation protéolytique de la particule virale (64). Lors de cette maturation, l'ARNg du VIH-1, encapsidé sous forme de dimère instable, va subir des réarrangements structuraux aboutissant à une structure plus stable, plus compacte et plus à même d'être rétro-transcrite (66,67,250). Afin de décomposer cette maturation de l'ARNg, nous avons utilisé une approche de cartographie chimique en solution pour étudier systématiquement les complexes formés par l'ARNg 1-604 et chacune des différentes protéines.

La première partie de mon travail s'est focalisée sur la formation des complexes dans des conditions permettant d'observer les interactions ARN-protéine de façon répétable. Après m'être assuré de la pureté de l'ARNg 1-604, j'ai pu déterminer des conditions optimales à la formation du complexe ARNg-Pr55<sup>Gag</sup>. En effet, les premières expériences dans un tampon au cacodylate de sodium (pH 7,5) n'ont pas donné les résultats escomptés (**Figure 25**). L'ARNg semblait pris dans un agrégat protéique, ce qui a été vérifié plus tard en DLS (**Figure 26**). Ainsi, la présence de ces agrégats pourrait expliquer le caractère erratique des données obtenues. En effet, il est possible que ces agrégats aient entraîné une partie des ARN dans la phase phénolique, ne laissant que les ARNs n'ayant pas formés de complexes lors de l'analyse SHAPE. De plus, dans l'hypothèse où les ARN ne seraient pas entraînés dans la phase phénolique, les interactions ARN-protéine dans un complexe et dans un agrégat sont certainement bien différentes. En effet, la formation d'agrégats implique généralement un repliement différentiel de la protéine et une oligomérisation de cette dernière (265–267). De telles hypothèses seraient cohérentes avec les données montrant une forte corrélation entre les conditions « NoProt », « 10 Gag » et « 60 Gag », mais également avec l'absence de similarité retrouvée avec les anciennes données du laboratoire (118).



Le problème d'agrégation des protéines du lot 1 (**Figure 26**) ne pouvait pas être résolu par un changement de lot. En effet, les quantités nécessaires de protéines pour l'ensemble des expériences n'étaient pas atteinte par les autres lots. Ainsi, le passage au Tris-HCl ou à l'HEPES, supplémenté de 10% de glycérol, a permis de restaurer les interactions ARN-protéines préalablement identifiées dans la littérature, ainsi qu'un profil de migration cohérent avec la concentration protéique (**Figure 27**). Dans l'ensemble, les problèmes expérimentaux détectés lors des tests initiaux ont été résolus et ont permis une homogénéisation des conditions entre le hSHAPE utilisé ici, et le MOHCA-seq qui sera présenté dans l'Axe 2.

Après avoir optimisé le protocole expérimental, je l'ai mis en œuvre dans l'étude systématique des complexes ribonucléoprotéiques. J'ai ainsi pu évaluer les sites de fixation des différentes protéines et l'effet de ces fixations, l'importance des deux domaines en doigt de zinc présents dans la NC et les effets chaperon de l'ARN de chacune des protéines (Partie III). Les analyses statistiques que j'ai pu réaliser (**Tableaux 6 et 7**) viennent également renforcer les conclusions tirées. Cependant, il est important de préciser que les tests statistiques réalisés ne s'intéressent pas à déterminer la significativité des différences une à une, mais à la tendance globale pour chaque nucléotide. Ainsi, une absence de différence significative globale pour un nucléotide n'impliquera pas forcément l'absence de différences significatives pour un nucléotide particulier. En effet, un des points avancés dans notre article est l'effet distinct de chaque protéine par rapport à la condition « Noprot ». Cette observation est également faite sur le **Tableau 6** où chaque protéine présente un profil différent des autres, à l'exception de Pr55<sup>Gag</sup> et NCp7 qui ne semblent diminuer significativement que les réactivités associées aux guanines. Cependant, cette similarité « macroscopique » est effacée lorsque les réactivités sont observées plus en détails (**Figure 2B et 2C** – Article – Partie III). De plus, nos résultats mettent en évidence que la capacité de protection des nucléotides par NCp15 est la plus importante des quatre protéines. Cette importante capacité de protection transparait également dans les **Tableaux 6 et 7** où NCp15 présente le plus de diminutions significatives globales de réactivités. A l'inverse, la protéine NCp9 est la seule qui induit une augmentation de réactivités significative sur un nucléotide. Ce contraste apparent entre les résultats pour NCp15 et NCp9 sont cohérents avec l'hypothèse d'une réorganisation de la structure de l'ARNg lors de la maturation, où la perte du domaine p6 (passage de NCp15 à NCp9) permet à la NC d'augmenter ses propriétés chaperon d'ARN (64,237,240). Finalement,

une caractéristique du domaine NC semble également apparaître dans nos résultats. En effet, 3 des 4 protéines induisent une diminution significative des réactivités des guanines (**Tableau 4**), ce qui corrèle bien avec l'idée d'une reconnaissance spécifique de ces protéines au niveau des guanines simples brins (53,195,268). Cependant, ce modèle suggère que la reconnaissance des guanines simples brins soit localisée dans le domaine Psi.

En s'intéressant alors plus spécifiquement aux différences entre les domaines Psi et NoPsi, il apparaît que les réactivités du domaine Psi sont significativement moins élevées que celles du domaine NoPsi (**Figure 28A**). Cependant, cette diminution significative est due principalement à la différence entre les uridines de ces deux domaines (**Figure 28C**). En effet, la répartition des réactivités associées aux guanines est quasiment identique entre les domaines. Ainsi, la diminution significative des réactivités associées aux guanines du domaine Psi et induite par les protéines Pr55<sup>Gag</sup> semble confirmer l'hypothèse de reconnaissance des guanines par les domaines en doigts de zinc du domaine NC (**Tableau 5B**). Dans le cas de NCp15, 7 diminutions significatives de la réactivité associées au domaine Psi sont observées par rapport à Pr55<sup>Gag</sup>. Cependant, dans le domaine NoPsi, NCp15 possède 14 sites de diminutions significatives supplémentaires. Ces données permettent d'expliquer la présence des diminutions significatives des réactivités A, C et G dans le domaine NoPsi (**Tableau 5A**). Finalement, en s'intéressant aux effets de NCp7 domaine à domaine, la diminution observée sur les guanines disparaît dans notre test statistique. Ce changement peut être expliqué par la diminution de la taille de l'échantillon qui impacterait le calcul de la p-value.

Pour continuer sur l'hypothèse de la reconnaissance spécifique des guanines simples brins, nos résultats mettent en évidence que Pr55<sup>Gag</sup>, NCp15, NCp9 et NCp7 reconnaissent 73 guanines parmi les 152 nucléotides protégés, montrant à nouveau l'importance de ces dernières. Ainsi, un lien peut être fait avec l'étude de Nikolaitchik *et al.*, (2020), qui pointe vers un effet synergique des guanines pour la sélection et l'encapsidation. Cette étude utilise des mutants (substitutions de G en A) de l'ARNg sur 7 sites (**Figure 29** – zones jaunes) et évalue l'effet de ces mutations sur l'encapsidation, et donc, indirectement sur la reconnaissance spécifique de l'ARNg par le domaine NC de Pr55<sup>Gag</sup>. En comparant plus en détail les résultats de protection obtenus pour les protéines Pr55<sup>Gag</sup> (**Figure 29A**) et NCp7 (**Figure 29B**), il est intéressant de noter que 100% des sites primaires de fixation (contour rouge) sont retrouvés. Pour les sites secondaires de fixation (contour orange), Pr55<sup>Gag</sup> reconnaît à nouveau 100% de ceux identifiés, à l'inverse de NCp7 qui ne reconnaît que le site en amont du domaine Psi. Pour

finir, deux des trois sites de reconnaissance appelés tertiaires (contour noir) sont retrouvés avec Pr55<sup>Gag</sup> et le dernier est reconnu par NCp7. En étendant l'analyse de ces sites aux protéines NCp15 et NCp9, 5 d'entre eux sont protégés par NCp15, contre 6 pour NCp9. Dans l'ensemble, les différents sites cités par Nikolaitchik *et al.*, (2020) sont tous retrouvés dans notre étude et reconnus par au moins une des quatre protéines. Cependant, aucune protéine à elle seule ne reconnaît tous les sites. Cette observation va dans le sens d'une spécificité de fixation des différentes protéines par le domaine NC au niveau des guanines simple brin du domaine Psi. De même, elle soulève la question du mécanisme de reconnaissance spécifique des guanines simple brin de Psi par le domaine NC de Pr55<sup>Gag</sup>. Comment la protéine peut-elle reconnaître spécifiquement les guanines de ce domaine ? Une hypothèse serait la présence d'un motif de reconnaissance spécifique. Cependant, aucun motif spécifique ne peut être déduits de l'ensemble des sites de fixation. En retirant le site de fixation secondaire présent sur le SL2-SL3 linker, un motif GGHG semble se dégager (H = A, C ou U), avec la nature du nucléotide central n'étant que peu importante. En effet, ce dernier peut être tourné vers l'extérieur sans empêcher la fixation de la protéine sur la boucle apicale de SL2 ou SL3(225–227). De plus, la réactivité de ces nucléotides (U291 et A319) n'est pas diminuée par la fixation de la protéine, ce qui suggèrerait que sa nature n'est pas essentielle à l'interaction. Cette hypothèse serait également soutenue par l'existence de modèles structuraux présentant un tige-boucle SL3 allongée où les nucléotides GA du site SL2-SL3 linker feraient face au site de fixation en GGAG dans une boucle interne (52,182,222,268). Ce type de structure serait relativement proche de celle adoptée par l'ARNg 1-604 après la fixation de la protéine NCp9 (**Figure 5** – Article - partie III). Une autre hypothèse permettant d'expliquer cette préférence du domaine NC de la protéine Pr55<sup>Gag</sup> pour la région Psi serait la structure tridimensionnelle de l'ARNg. En effet, une exposition particulière au solvant des guanines libres dans ce domaine structural pourrait également fournir une explication sur la spécificité de sélection. Cette suggestion est également soutenue par des données expérimentales ayant montré l'importance du dinucléotide G226/A227 dans le repliement tridimensionnel du domaine Psi (269) et plus généralement, des guanines non appariées (214). Cependant, il n'existe que peu d'études de la structure tridimensionnelle de la région 5' terminale de l'ARNg du VIH-1 dans la littérature (122,214,247,269–274) et ces dernières sont rarement modélisées à partir de données expérimentales tridimensionnelles (122,214,269,271,274).

## AXE 1 – DISCUSSION

Pour finir, la comparaison de nos résultats à ceux rapportés dans la littérature a mis en évidence le nombre important de sites de fixation potentiels des protéines sur la région 5' terminale de l'ARNg du VIH-1, mais également l'importance des approches utilisées lors de ces études. En effet, parmi les 250 nucléotides étudiés, les protéines Pr55<sup>Gag</sup> et NCp7 affectent respectivement 61 et 113 d'entre eux, dont 15 et 18 sont reconnus dans notre étude (72). Ainsi, ces nucléotides ne sont pas communs à l'ensemble des approches. Cette divergence est due aux différences méthodologiques et aux limitations de ces dernières. En effet, les approches de cartographie chimiques ne sont pas capables de reconnaître les potentiels sites de fixation des protéines présents sur des régions peu flexibles, comme les régions double brins. De plus, l'utilisation d'ARN mutés (244) ou de protéines tronquées (247,264) influencent également les repliements étudiés et donc les interactions identifiées. Ces variables expliquent, en partie, la diversité de sites de fixation reconnus. Dans l'ensemble, cette figure illustre à la fois l'intérêt porté aux problématiques structurales de la région 5' terminale et la nécessité d'atteindre un consensus pour pouvoir avancer dans la compréhension des mécanismes moléculaires régis par ces structures.

## Axe 2

Étude des interactions à longue distance  
de la région 5' de l'ARNg du VIH-1

Mise au point d'une cartographie  
chimique multidimensionnelle à haut  
débit « MOHCA-seq »



Les molécules d'ARN sont capables de remplir une myriade de fonctions cellulaires (275,276). Dans chacun de ces processus, ces molécules s'organisent selon des structures complexes qui régissent leurs fonctions (277,278). De façon similaire aux protéines, ces fonctions reposent sur le repliement adopté par les ARN et les interactions possibles avec le milieu extérieur. Ainsi, la compréhension de ces fonctions nécessite l'identification et la caractérisation des régions structurales qui les portent (279). Cependant, ces structures ne sont pas statiques. En effet, leurs dynamiques sont extrêmement complexes et permettent souvent l'existence de multiples conformations. Parmi ces conformations, certaines peuvent être formées avec une forte probabilité et compter pour une portion importante de la population d'ARN étudiée tandis que d'autres sont bien plus rares. Ainsi, l'étude d'une population composée de l'ensemble de ces conformations rend l'analyse structurale des ARN délicate. Malgré ces difficultés, les nombreuses investigations pour tenter de résoudre les structures de différents ARN ont permis d'obtenir une idée générale de la formation des dites structures (280). En effet, ces structures sont repliées hiérarchiquement avec un premier niveau de structure usuellement appelé structure secondaire et qui est constitué par la formation d'hélices double brin grâce à des interactions Watson-Crick (WC)  $\leftrightarrow$  Watson-Crick par  $\sim 50\%$  des nucléotides. L'assemblage de ces hélices entre elles par des interactions non-WC organise ensuite la structure tertiaire. Des interactions WC-WC à longue distance peuvent également être impliquées dans le repliement tridimensionnel des ARN.

Les approches actuelles permettant de déterminer la structure des ARN peuvent être séparées en deux catégories : biophysiques et biochimiques. Parmi les approches biophysiques, la cristallographie, la diffraction des rayons X aux petits angles (SAXS), la RMN et la cryomicroscopie électronique (cryo-EM) permettent d'atteindre des modèles structuraux frôlant une résolution atomique. Cependant, ces approches font également face à des restrictions inhérentes imposées par les ARN. En effet, la multiplicité de structures adoptées, les larges régions non structurées, les états de compaction distincts et la capacité à interagir avec différents partenaires rendent la résolution de ces structures plus complexe que celle des protéines (281). Ces limitations expliquent pourquoi la Protein Data Bank (PDB) ne contient que 6 288 structures d'ARN sur les 206 847 structures recensées au mois de septembre 2022 (soit  $\sim 3\%$ ). A l'inverse des techniques permettant l'obtention de modèles à haute résolution, les approches biochimiques telles que les méthodes de cartographies peuvent être appliquées à la majorité des ARN, dans des conditions inaccessiblelement

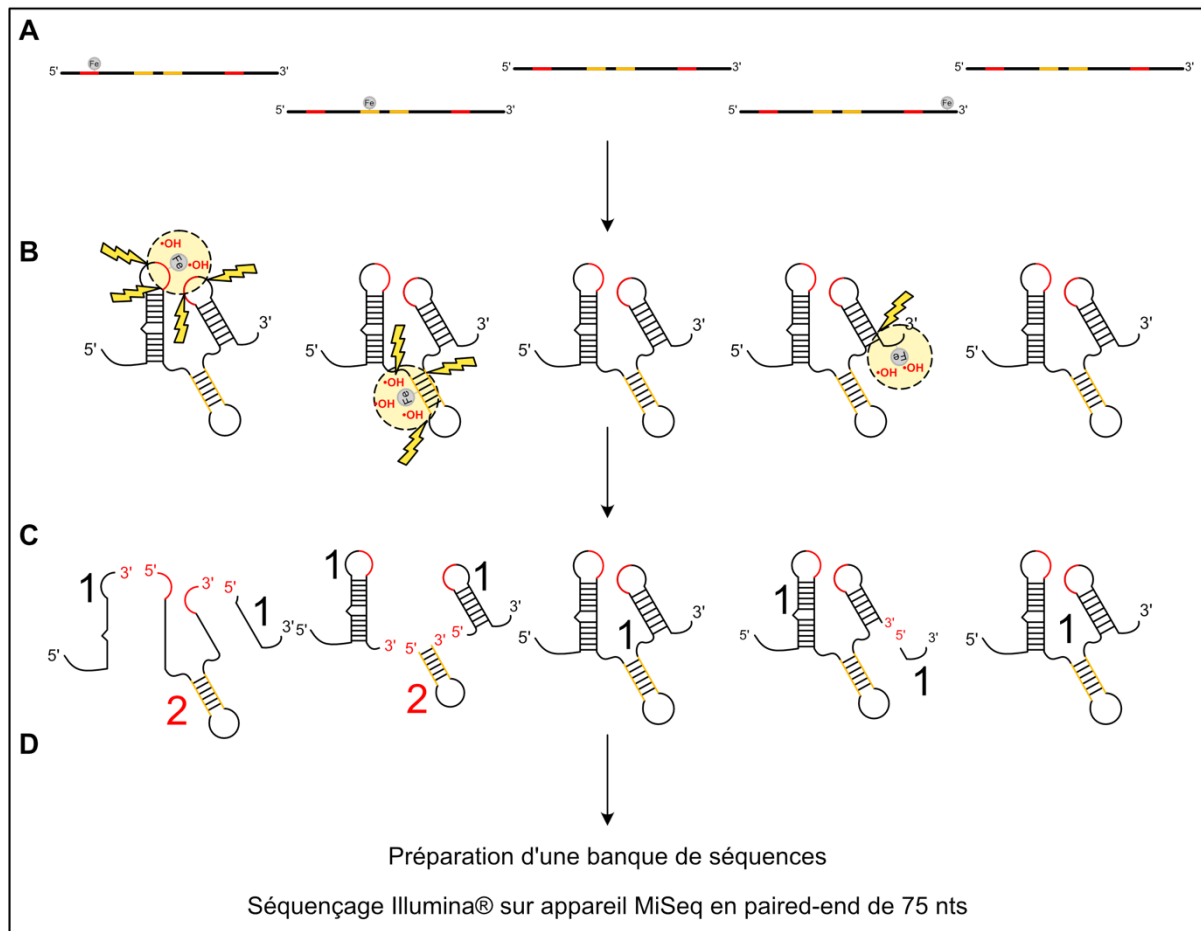
par les approches biophysiques (282). Ces méthodes sont basées sur l'accessibilité au solvant (283–286), l'accessibilité des faces WC et Hoogsteen (287), ou la flexibilité (261,288) des nucléotides ciblés par les attaques chimiques. Ainsi, elles ne sont pas soumises aux inconvénients de tailles, d'environnements ou de structures qui freinent les approches biophysiques (289). Dans les dernières années, plusieurs de ces approches ont été développées afin de mieux appréhender les repliements impliqués dans la structure tertiaire des ARN (282). Cependant, ces techniques nécessitent de détecter les clivages ou modifications chimiques réalisées, incorporant de possibles biais dans l'analyse qui peuvent se répercuter sur les structures modélisées. Ainsi, ces approches biochimiques sont moins précises et plus sujettes aux biais liés aux méthodes de détection et de prédiction des structures (276).

Dans le cas de l'ARNg du VIH-1, malgré de nombreuses études biophysiques et biochimiques de sa structure 5' terminale, peu de repliements impliqués dans sa structure tridimensionnelle ont été formellement identifiées. Ce manque de certitude concernant cette structure provient de l'hétérogénéité structurale dont fait preuve cette région, mais également de la multitude de modèles présents dans la littérature. En effet, plusieurs modèles rapportent différentes interactions à longue distance pour un ou plusieurs des partenaires impliqués (272,290). Ces variations dans les structures secondaires proposées peuvent avoir des conséquences immenses dans le repliement tridimensionnel, puisque les contraintes imposées par les doubles hélices d'ARN modulent les interactions possibles. Ainsi, ces interactions, distinctes et parfois contradictoires, empêchent l'apparition d'un consensus sur la structure 5' terminale. Jusqu'à présent quelques études se sont intéressées aux interactions à longue distance et à celles non canoniques mises en place dans l'ARNg du VIH-1. Ces travaux détiennent certaines des clefs permettant d'expliquer les différences et les incohérences retrouvées lors de la comparaison des modèles présents dans la littérature. En effet, l'interaction triple A-U-A dans la boucle SL2 (A276, U295 et A296) ou l'interaction entre le U131 et C179 dans le domaine PBS (207), sont rarement représentées dans les modèles structuraux. Cependant, la mise en évidence de leur existence permettrait d'expliquer certains résultats expérimentaux. De même, une interaction Hoogsteen:Watson-Crick existe entre les guanines anti-G240 et syn-G279 (133) et il a été montré que la nature des nucléotides C243 et G277 était importante à l'infectivité virale (291). Ainsi, ces différentes informations



pourraient permettre d'améliorer la compréhension des modèles présents dans la littérature, et à terme, peut-être de les clarifier.

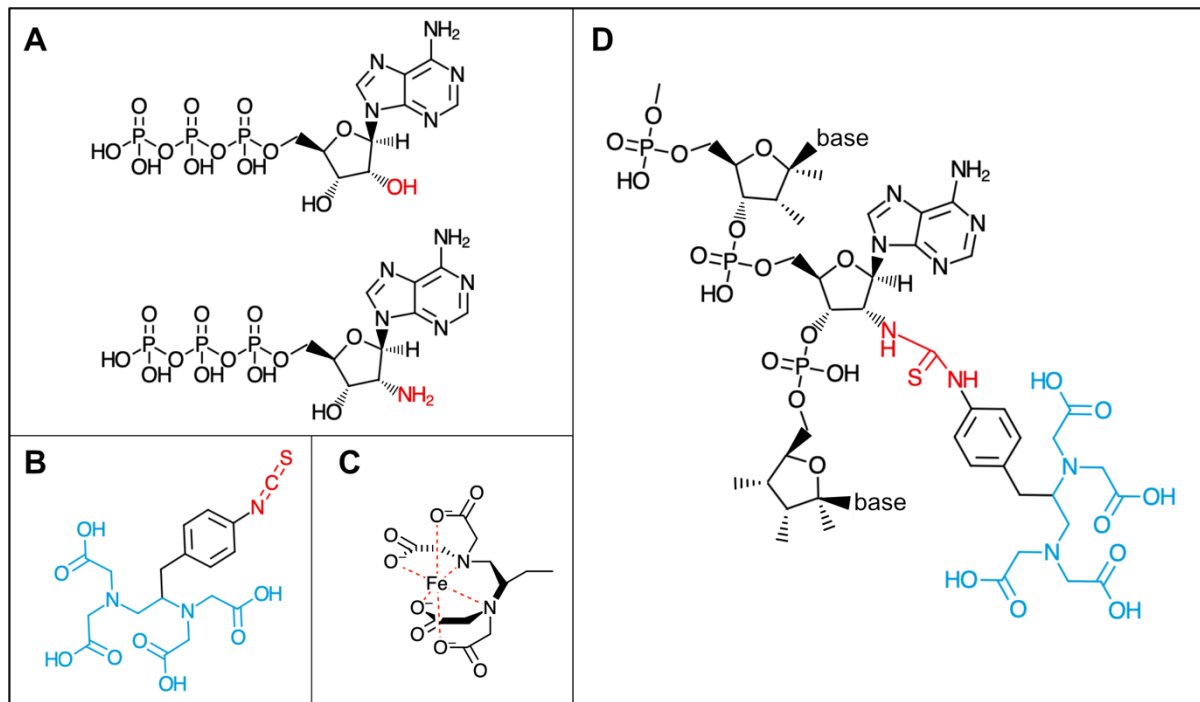
Dans le but de tenter de clarifier les interactions à longue distance présentes dans la structure de l'ARNg du VIH-1, j'ai adapté et optimisé une approche de cartographie chimique multidimensionnelle développée par l'équipe de Rhiju Das (260). Ce choix est justifié par différents points ayant évolué ces dernières années. Premièrement, les approches biophysiques n'étaient pas capables de résoudre la structure de la région 5' terminale de l'ARNg. En effet, des approches comme la cristallographie, le SAXS ou la RMN sont limitées par rapport à la taille et l'hétérogénéité structurale de la molécule étudiée. Un ARN long de 604 nucléotides reste donc bien en dehors de leurs capacités d'analyse. Par exemple, dans le cas du VIH-1, le plus grand ARN étudié à ce jour par RMN correspond aux 356 premiers nucléotides de la séquence de l'ARNg (211). Dans le cas de la cryo-EM, cette dernière n'était pas capable de fournir une structure suffisamment résolutive seule pour un ARN nu (133). En effet, selon la PDB, seules 187 structures d'ARN provenaient d'expériences de cryo-EM jusqu'en 2014. Parmi ces structures, 21 atteignaient une résolution entre 3,0 et 4,5 Angströms (Å). Il a fallu attendre 2020 pour que des structures d'ARN seuls soient enfin obtenues à des résolutions inférieures à 10 Å (292). Ainsi, le projet s'est axé autour d'une approche biochimique pour la détermination des interactions à longue distance. Cependant, de nombreuses méthodes existent également dans la littérature (282). L'ARNg du VIH-1 étant capable de dimériser, les approches s'appuyant sur des cross-linking comportent un risque non nul de faux positifs issus d'interactions en *trans*. Dans le but de diminuer ce risque, l'approche de MOHCA-seq ou *Multiplexed •OH Cleavage Analysed with paired-end sequencing* a donc été sélectionnée (**Figure 31**).



**Figure 31** : Représentation schématique de l’approche du MOHCA-seq. (A) Après la transcription *in vitro* des ARN d’intérêts en présence d’ATP 2’NH<sub>2</sub>, un groupement ITCB-EDTA est fixé et sert à chélater un ion Fe<sup>3+</sup>. (B) Les ARN d’intérêt sont repliés et le clivage aux radicaux hydroxyles est déclenché. La sphère d’action des radicaux est représentée sous forme de cercle jaune. (C) Les fragments d’ARN produits par le clivage sont récupérés. Ceux présentant (fragments 1) une ou plusieurs extrémités parentales n’apportent pas d’informations structurales importantes. Les fragments (2) comportant deux extrémités non parentales (5’ et 3’) contiennent des informations sur les régions proches dans l’espace lors du clivage. (D) Ces informations sont analysées par un séquençage Illumina® sur un appareil MiSeq en mode paired-end 75 nts, après que les fragments aient été préparés sous forme de banque de séquences.

Cette approche permet de mettre en évidence la proximité spatiale de nucléotides distants dans la séquence (260). Pour ce faire, l’ARN d’intérêt est transcrit *in vitro* en présence d’un nucléoside triphosphate modifié portant un groupement amine sur le carbone C2’ (**Figure 32A**). Le nucléotide modifié est incorporé aléatoirement dans la séquence à hauteur d’une modification maximum par transcrit. Un groupement isothiocyanobenzyl-EDTA (**Figure 32B et 32C**) est ensuite fixé au groupement 2’ amine précédemment incorporé ce qui permet la chélation d’un ion Fe<sup>3+</sup> à l’endroit où le nucléotide modifié a été introduit. L’ARN est

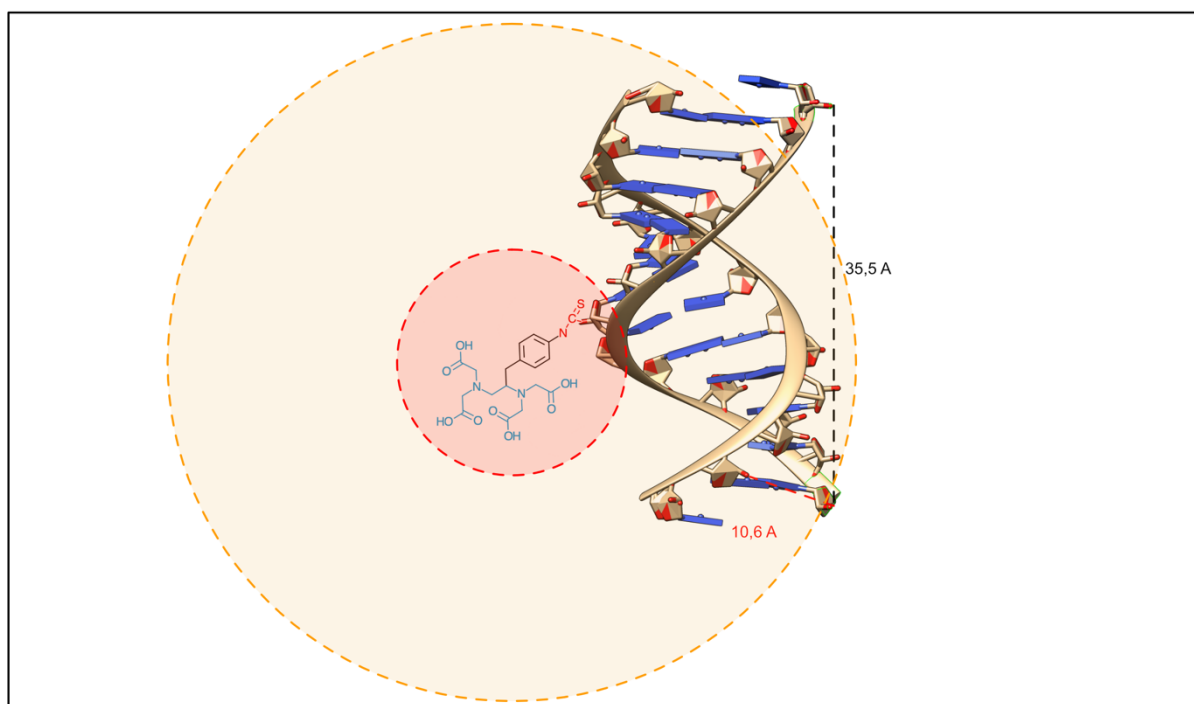
replié dans la condition d'intérêt et supplémenté d'ascorbate de sodium qui permet l'induction d'une réaction de type Fenton. Lors de cette réaction, les ions  $Fe^{3+}$  sont réduits en  $Fe^{2+}$  par l'ascorbate de sodium. Les ions  $Fe^{2+}$  produits peuvent alors réagir avec l' $O_2$  pour être oxydé en  $Fe^{3+}$ . Des radicaux hydroxyles ( $\bullet OH$ ) sont issus de cette réaction et peuvent induire des clivages dans l'ARN à proximité (**Figure 33**).



**Figure 32** : Représentation de réactifs principaux de l'approche de MOHCA-seq. (A) La différence entre l'ATP 2'OH et l'ATP 2'NH<sub>2</sub> est représentée en rouge. (B) L'isothiocyanobenzyl-EDTA (ITCB-EDTA) est représenté avec le groupement EDTA, capable de chélater un ion  $Fe^{3+}$ , en cyan et le groupe réagissant avec les amines en rouge. (C) La structure du groupement EDTA est représentée après la chélation d'un ion  $Fe^{3+}$ . (D) La liaison covalente entre un AMP 2'NH<sub>2</sub> incorporé dans un brin d'ARN et l'ITCB-EDTA est représentée en rouge. Les molécules ont été dessinées à l'aide du logiciel ChemDoodle.

Ces radicaux hydroxyles ( $\bullet OH$ ), espèce radicalaire extrêmement réactive dont la durée de vie est très courte (1/10<sup>ème</sup> de  $\mu$ sec), sont produits et réagissent de façon non sélective avec les composés à la limite de leur portée de diffusion allant de 10 à 30 Å (**Figure 33**) (293,294). Lorsque les radicaux hydroxyles réagissent au niveau du ribose, ils induisent le clivage de l'ARN. Après réparation des extrémités, les fragments sont séparés en fonction de leur taille et préparés pour un séquençage couplé des deux extrémités *paired-end sequencing*. L'analyse des résultats est réalisée en partie à l'aide du logiciel MAPSeeker

(<https://ribokit.github.io/MAPseeker/>) (260,295) et la modélisation tridimensionnelle de l'ARN est réalisée par le logiciel FARFAR2 (<https://rosie.rosettacommons.org/farf2>) (296,297). Dans le cas du MOHCA-seq, les fragments d'intérêts produits après le clivage sont ceux ne comportant aucune des extrémités parentales de l'ARN étudié. En effet, si le fragment possède des extrémités 5' et 3' différentes de celles de l'ARN de départ non clivé, il est probable qu'une des deux extrémités soit la source de production des radicaux hydroxyles et que la seconde soit une région proche de la source et clivée par les radicaux produits (**Figure 31C – fragment 2**).



**Figure 33** : Représentation de la sphère d'effet des radicaux hydroxyles produits en réaction au fer. La taille de la sphère est comprise entre 10 et 30 Å (cercle rouge = 10 Å et cercle jaune = 30 Å). Elle est comparée à une molécule d'ARN double brin (Source : PBD 413D) et aux distances des nucléotides dans l'hélice (10,6 Å entre 2 groupements 2'OH et 35,5 Å entre les deux extrémités)

# Matériel & Méthodes :

## I/ Préparation des ARN modifiés pour le MOHCA-seq

### I.1/ Transcription *in vitro* avec 2'NH<sub>2</sub> ATP

La transcription *in vitro* est réalisée selon le même protocole que présenté précédemment à l'exception de la présence d'ATP modifié en 2' avec un groupement amine (**Figure 32A**). Le ratio ATP 2'NH<sub>2</sub>:ATP 2'OH a varié au cours de la mise au point de l'approche allant de 1:2 ; 2:5 ; 1:3 ; 1:10 à 1:20. Finalement, le ratio qui a été choisi est de 1 ATP-2'-NH<sub>2</sub> pour 20 ATP-2'-OH.

### I.2/ Purification de l'ARN transcrit *in vitro*

Après la précipitation à l'éthanol, les ARN 1-604 transcrits *in vitro* sont repris dans de l'eau MQ. Ils sont ensuite purifiés sur un gel de polyacrylamide dénaturant à 6% d'acrylamide et 8 M d'urée. Les ARN sont dénaturés à 90°C pendant 2 min avant d'être déposés sur le gel. Après 4 – 6 h de migration à 300 V et température ambiante, les ARN sont visualisés et identifiés sur un trans-illuminateur afin d'être excisés. Les morceaux de gel contenant l'ARN d'intérêt sont placés dans une membrane de dialyse (Spectra/Por® 3 RC) préalablement conditionnée 15 min dans de l'eau MQ. Les morceaux de membranes sont recouverts de 1,5 mL d'eau MQ, la membrane est refermée et immergée dans une cuve d'électrophorèse horizontale contenant du TBE 1x. Les ARN sont électro-élués hors des morceaux de gel par migration à 95 V pendant 1 h à température ambiante. Après 1 h, les électrodes sont inversées pendant 2-3 min à 95 V. La solution située dans la membrane de dialyse est prélevée et les ARN sont précipités à l'éthanol. Les culots d'ARN purifiés sont repris dans 50 µL de H<sub>2</sub>O MQ puis quantifiés au QuBit™ 4 Fluorometer (Invitrogen™). La pureté des échantillons est également contrôlée sur un gel de polyacrylamide dénaturant à 8% d'acrylamide et 8 M d'urée ainsi que par électrophorèse capillaire sur une puce RNA 6000 Nano (Agilent) à l'aide d'un 2100 Bioanalyzer (Agilent).

### I.3/ Modification des ARN 2'NH<sub>2</sub> par l'ITCB-EDTA

Après s'être assuré de la pureté des ARN ayant incorporé les nucléotides 2'NH<sub>2</sub>, 15 µL d'ARN à 1 µg/µL sont placés dans un tube Eppendorf® et mélangé à 15 µL de tampon phosphate de potassium à 0,8 M (pH 8,5). La solution est supplémentée de 5,7 µL d'ITBC-EDTA (Dojindo) à 200 mM solubilisé dans un tampon phosphate de potassium à 0,4 M (pH 8,5). Le mélange est incubé 16 h à 37°C sous agitation (300 rpm) dans un thermomixer compact (Eppendorf®).

### I.4/ Chélation du fer sur l'ITBC-EDTA

Après 16 h d'incubation à 37°C, les ARN modifiés à l'ITCB-EDTA sont supplémentés de 12 µL de FeCl<sub>3</sub> à 200 mM. L'ajout de fer à la solution induit la formation d'un précipité jaune laiteux. La solution est incubée 15 min à 37°C sous agitation (300 rpm) puis le fer en excès est chélaté par l'ajout de 10 µL d'EDTA à 0,5 M et une incubation de 15 min à température ambiante, sous agitation à 300 rpm. Les ARN sont purifiés successivement sur une colonne BioSpin® 6 (Bio-Rad) puis sur une colonne RNA Clean and Concentrator Zymo®, selon le protocole des fournisseurs. Le précipité jaune laiteux est éliminé lors du passage sur la colonne BioSpin®.

## II/ MOHCA-seq : fragmentation des ARN

Lors de la conceptualisation de l'expérience de MOHCA-seq, nous avons choisi 5 conditions d'intérêts pour l'étude de la structure de la région 5' terminale de l'ARNg. Ainsi, l'ARN est replié dans [1] H<sub>2</sub>O MQ, [2] un tampon favorisant la monomérisation de l'ARN (30 mM HEPES pH 8,0 ; 40 mM KCl ; 0,1 mM MgCl<sub>2</sub>) ou [3] un tampon favorisant la dimérisation de l'ARN (30 mM HEPES pH 8,0 ; 300 mM KCl ; 5 mM MgCl<sub>2</sub>). De plus, nous avons décidé d'étudier l'effet de deux oligonucléotides antisens sur l'ARN replié dans un tampon de dimérisation (**Tableau 8**). Ces oligonucléotides ciblent soit [4] la région U5 (nucléotides 102 à 116), soit [5] la région AUG (nucléotides 334 à 348).

### II.1/ Clivages aux radicaux hydroxyles

Dans un premier temps, 4 pmoles (0,8 µg) d'ARN sont repris dans un volume final de 8 µL, en présence ou non de 2 µL d'oligonucléotide antisens à 100 µM. L'ARN est incubé 1 min à 70°C pour être dénaturé avant d'être refroidi sur glace. L'ARN est ensuite replié par incubation 30 min à 37°C dans 10 µL final de solution contenant le tampon adapté à la condition étudiée. La réaction de Fenton est déclenchée par l'ajout de 1 µL d'ascorbate de sodium à 100 mM et la réaction est poursuivie à température ambiante pendant 20 min. Afin d'arrêter la réaction, le mélange est supplémenté de 1 µL de thiourée à 100 mM et incubé 2 min à température ambiante. Les fragments d'ARN obtenus sont ensuite purifiés sur une colonne RNA Clean and Concentrator Zymo® selon les instructions du fournisseur.

### II.2/ Déphosphorylation et rephosphorylation des fragments d'ARN

Les fragments d'ARN obtenus par l'étape de clivage au radicaux hydroxyles présentent des extrémités 5' et 3' non conformes aux réactions de ligations. Une étape de réparation est donc nécessaire pour poursuivre la création de la librairie de séquençage. Ainsi, les fragments d'ARN sont déphosphorylés et rephosphorylés en une étape en supplémentant l'ARN de 10 U de T4 PNK (ThermoFisher Scientific) et de tampon T4 PNK 1x (50 mM Na-MES pH 6,0 ; 10 mM MgCl<sub>2</sub> ; 5 mM DTT ; 10 µM ATP). Le mélange est incubé 30 min à 37°C et la réaction enzymatique est arrêtée par ajout de H<sub>2</sub>O MQ qsp 50 µL. L'ARN est purifié par une extraction phénol-chloroforme (vol/vol) suivie d'une précipitation à l'éthanol.

### II.3/ Séparation des fragments en fonction de leur taille

Les ARN sont repris dans 300 µL d'eau MQ et séparés en fonction de leur taille à l'aide de billes magnétiques SPRISelect (Beckman Coulter). Brièvement, 2 volumes de billes SPRISelect sont mélangés aux ARN par aspiration refoulement puis incubés à température ambiante pendant 5 min. Le tube Eppendorf® contenant les ARN et les billes est placé sur un portoir magnétique afin de prélever le surnageant qui contient les fragments d'ARN plus petits que 200 nucléotides. Les billes sont lavées à l'aide de 600 µL d'éthanol 85% pendant 30 sec à température ambiante. L'éthanol 85% est éliminé et les billes sont séchées à l'air pendant 15 min. Le tube Eppendorf® est ensuite retiré du portoir magnétique et 300 µL de H<sub>2</sub>O MQ

sont ajoutés pour resuspendre les billes. Le mélange est laissé 2 min à température ambiante avant d'être replacé sur le portoir magnétique. Après 3 min, l'éluât contenant les grands fragments d'ARN est prélevé. Les petits et les grands fragments d'ARN sont précipités à l'éthanol.

### III/ Préparation de la librairie de séquençage à haut-débit

#### III.1/ Ligation d'un adaptateur en 3' des fragments d'ARN :

Après la précipitation, les ARN sont repris dans 8 µL de H<sub>2</sub>O MQ et supplémentés de 100 pmoles d'adaptateur ARN (**Tableau 8**). L'ARN est dénaturé 2 min à 90°C puis refroidi sur glace. La réaction de ligation de l'adaptateur en 3' est réalisée par l'ajout de 2 µL de tampon T4 RNA Ligase 1 (NEB®), 10 U de T4 RNA Ligase 1 (NEB®) et 8 µL de PEG 8000 à 50%. Le mélange est incubé 16 h à 25°C dans un thermomixer compact (Eppendorf®) sous agitation à 300 rpm. La réaction est arrêtée par ajout de 80 µL de H<sub>2</sub>O MQ. Les ARN ligués sont purifiés par une extraction phénol-chloroforme (vol/vol, pH 7.5) suivie d'une précipitation à l'éthanol.

#### III.2/ Transcription inverse des fragments ARN ligués :

Les ARN ligués sont repris dans 8 µL de H<sub>2</sub>O MQ et supplémentés de 4 µL de dNTPs à 2,5 mM et de 1 µL de primer de transcription inverse complémentaire à l'adaptateur ligué à 10 µM. Le mélange est incubé à 65°C pendant 5 min puis refroidi sur glace. La réaction de transcription inverse est catalysée par l'ajout de 4 µL de tampon SuperScript IV 5x (Invitrogen), 1 µL de DTT à 100 mM, 1 µL de RNasine à 40 U/µL et 1 µL de SuperScript IV Reverse Transcriptase (Invitrogen) à 200 U/µL. Les ARN sont rétro-transcrit par une incubation à 55°C pendant 25 min suivie d'une inactivation thermique de l'enzyme par un passage à 80°C pendant 10 min. Les primers de transcription inverse sont retirés par ajout de 1 µL d'Exo I (NEB®) à 20 U/µL et incubation 45 min à 37°C. La matrice ARN est dégradée par ajout de 10 µL de NaOH à 500 mM suivie d'une incubation à 90°C pendant 20 min. La réaction est interrompue par l'ajout de 10 µL de Tris-HCl pH 7.2 (à 25°C) à 1 M. Les ADNc synthétisés sont purifiés par une extraction phénol-chloroforme et une précipitation à l'éthanol.



### III.3/ Ligation d'un adaptateur en 3' des ADNc :

Les ADNc sont repris dans 4 µL de H<sub>2</sub>O MQ et 1 µL d'adaptateur 3' (**Table 8**) à 10 µM est ajouté. Le mélange est dénaturé à 90°C pendant 2 min puis refroidi sur glace. La réaction de ligation en 3' des ADNc est débutée par l'ajout de 3 µL de tampon T4 RNA Ligase 1 10x (NEB®), 1,5 µL d'ATP à 20 µM, 1,5 µM de DTT à 20 µM, 4,5 µL de T4 RNA Ligase 1 à 20 U/µL et 15 µL de PEG 8000 à 50%. Le mélange est incubé à 25°C sous agitation à 300 rpm sur un thermomixer compact (Eppendorf®), pendant 70 h. La réaction est arrêtée par ajout de H<sub>2</sub>O MQ qsp 100 µL et les ADNc ligués sont purifiés par une extraction phénol-chloroforme (vol:vol , pH 7.5) suivie d'une précipitation à l'éthanol. Les ADNc sont repris dans 8 µL de H<sub>2</sub>O MQ.

Oligonucléotide	Séquence	Région d'hybridation	Modification
@U5	5'- dCAG ACG GGC ACA CAC -3'	102 - 116	/
@AUG	5'- dCTC TCG CAC CCA TCT -3'	334 - 348	/
Adaptateur 3' ARN	5'- dTGG AAT TCT CGG GTG CCA AGG - 3'	/	5' rApp 3' ddC
Amorce RT	5'- dGCC TTG GCA CCC GAG AAT TCC A -3'	Adaptateur 3' ARN	
Adaptateur 3' ADN	5'- dGAT CGT CGG ACT GTA GAA CTC TGA AC -3'	/	5' Phosphate 3' Phosphate

**Tableau 8 :** Séquences des oligonucléotides utilisés lors du MOHCA-seq

### III.4/ Amplification de la librairie d'ADNc par PCR

Les séquences des adaptateurs Illumina P5 et P7 nécessaires au séquençage sont ajoutés lors de l'étape de PCR. Dans un premier temps, un mélange réactionnel contenant 10 µL de tampon Phusion® HF 5x, 26,5 µL de H<sub>2</sub>O MQ, 1 µL de dNTP à 10 mM, 2 µL d'oligo RP1 TrueSeq Small RNA (TruSeq) et 0,5 µL de Phusion® DNA Polymerase (ThermoFisher) à 2 U/µL est préparé. Les 40 µL de mélange sont ajoutés à 8 µL d'échantillons purifiés et supplémentés de 2 µL d'oligo RPI\_X TruSeq Small RNA, en fonction de la condition étudiée (**Tableau 9**). Une amplification PCR de 12 cycles est réalisée sur un appareil T100™ Thermal Cycler selon le programme suivant : une dénaturation initiale de 30 s à 98°C, suivi des 12 cycles de 10 s à 98°C, 30 s à 60°C et 45 s à 72°C. Une élongation finale de 10 min à 72°C est réalisée et la température est stabilisée à 10°C. Les échantillons sont contrôlés par une analyse sur une puce ADN au Bioanalyzer (Agilent).

N° Tube	Échantillon	Amorce (TruSeq)	Séquence index
1	H <sub>2</sub> O MQ - Court	RPI-1	ATCACG
2	Tp Mono - Court	RPI-3	TTAGGC
3	Tp Dimère – Court	RPI-5	ACAGTG
4	Tp Dimère @U5 - Court	RPI-8	ACTTGA
5	Tp Dimère - @AUG - Court	RPI-10	TAGCTT
6	H <sub>2</sub> O MQ – Long	RPI-2	CGATGT
7	Tp Mono – Long	RPI-4	TGACCA
8	Tp Dimère – Long	RPI-7	CAGATC
9	Tp Dimère @U5 – Long	RPI-9	GATCAG
10	Tp Dimère @AUG – Long	RPI-11	GGCTAC
11	Contrôle H <sub>2</sub> O MQ	RPI-11	GGCTAC

**Tableau 9** : Récapitulatif des conditions étudiées en MOHCA-seq et des index utilisés pour chaque échantillon.

### III.5/ Purification de la librairie d'ADNc

Après la PCR, il est nécessaire de purifier les amplicons obtenus sur gel afin d'éliminer les amorces et les oligomères formés lors des cycles d'amplification. Pour ce faire, 8 µmoles d'un mélange approximativement équimolaire des 5 produits PCR issus des fragments courts

## AXE 2 – MATERIEL & METHODES

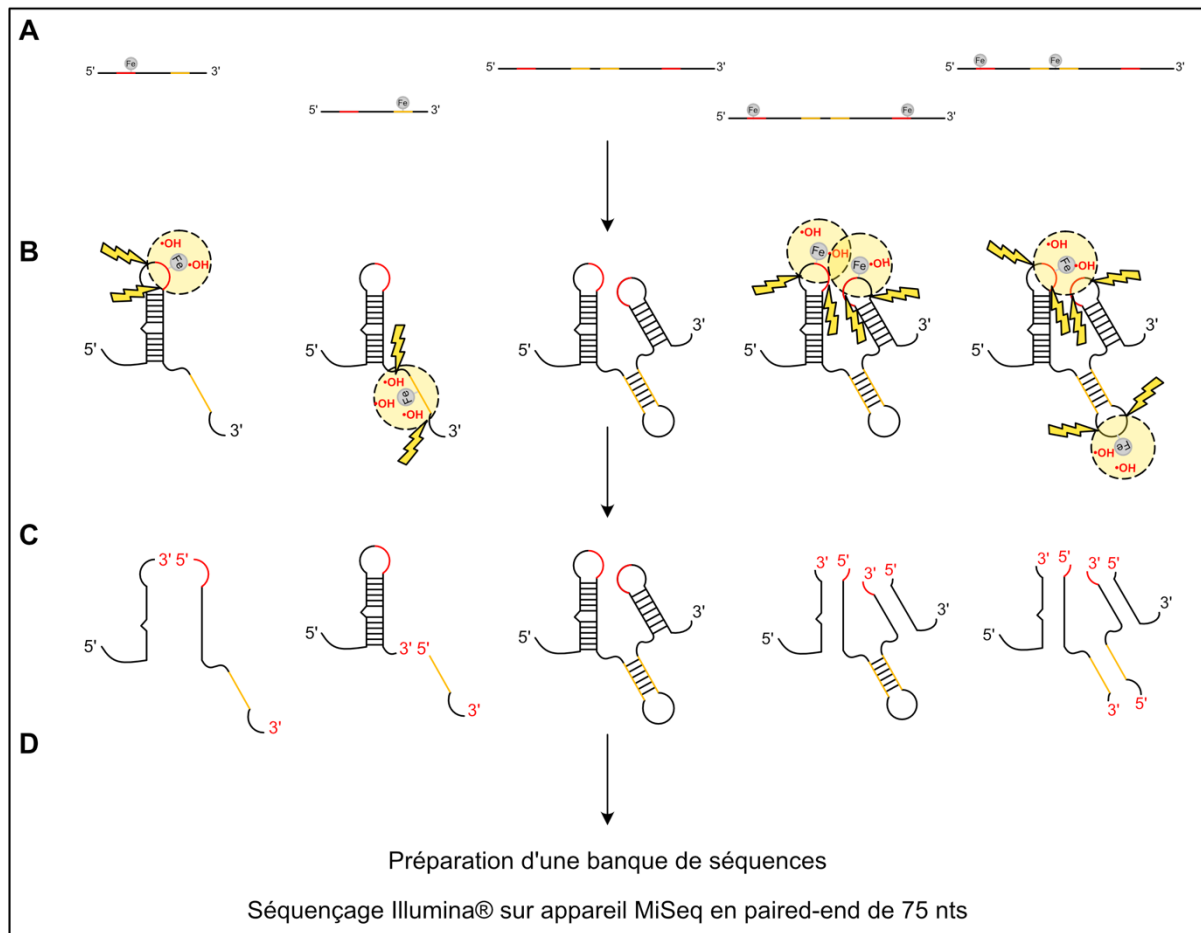
sont rassemblés dans un tube Eppendorf® d'un côté et 5 µmoles d'un mélange approximativement équimolaire des 5 produits PCR issus des fragments longs sont rassemblés dans un second tube. Les amplicons sont concentrés à 30 µl par évaporation au SpeedVac®. La purification est réalisée par électrophorèse sur un gel pré-coulé TBE-Urea NOVEX™ (Invitrogen) à 6% d'acrylamide, 1 h30 à 145 V, à température ambiante dans un tampon TBE 1x. Les fragments d'intérêts sont identifiés au trans-illuminateur puis découpés. Les morceaux de gels sont broyés, supplémentés de 300 µL de H<sub>2</sub>O MQ et élués sur la nuit à température ambiante, sous agitation à 700 rpm. La solution est filtrée sur des filtres Spin-X® d'acétate de cellulose à 0,45 µm (Costar®) et les amplicons sont précipités à l'éthanol. Les échantillons sont repris dans 12 µL de H<sub>2</sub>O MQ et ils sont quantifiés au QuBit et analysés au Bioanalyser (Agilent). Après la quantification, les échantillons sont séquencés sur un appareil MiSeq en mode paired-end 75 nts.



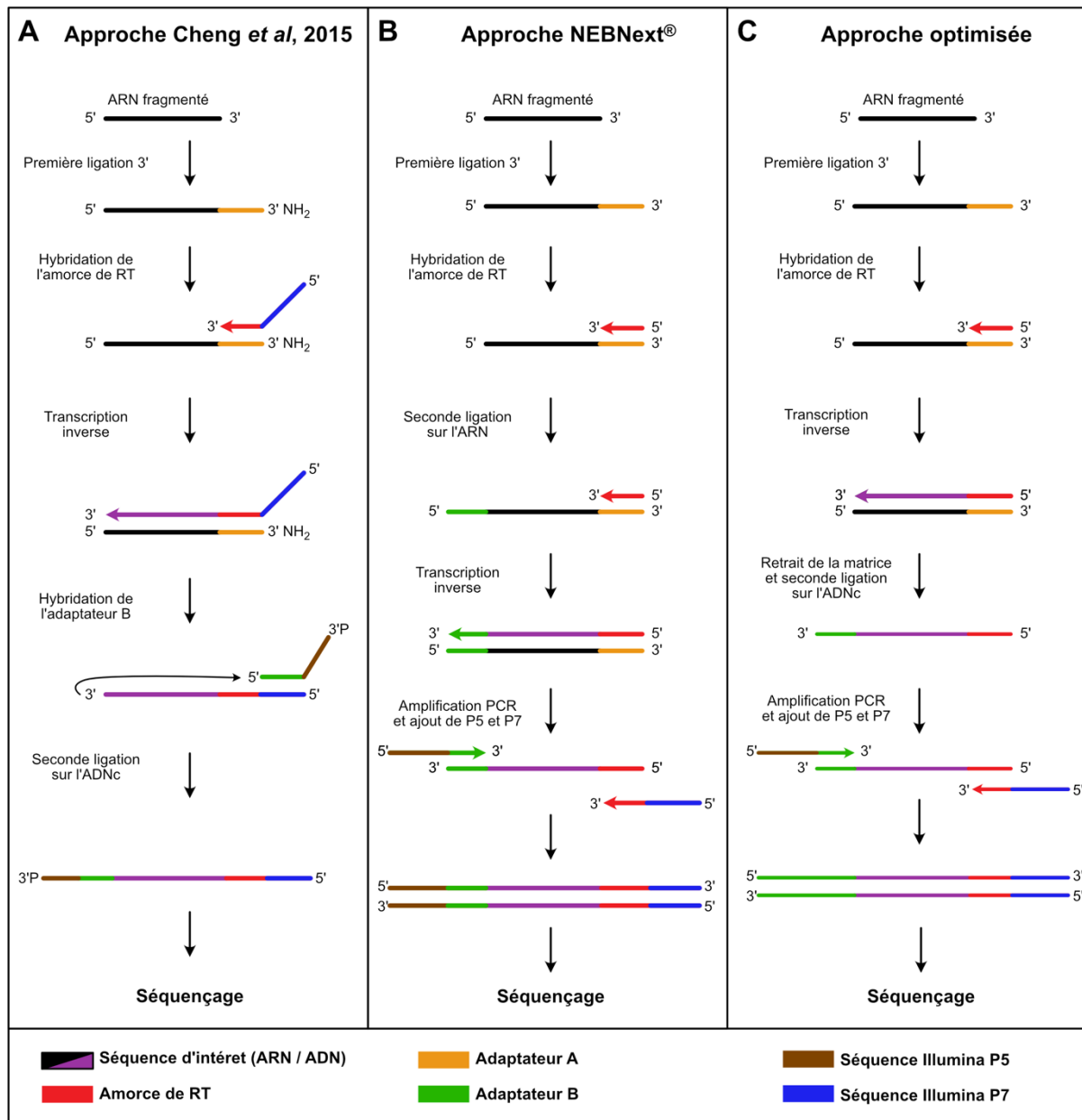
# Résultats

Lors de cette deuxième partie de mes travaux, je me suis donc intéressé à l'identification des interactions à longue distance impliquées dans le repliement de la région 5' terminale de l'ARNg du VIH-1, *in vitro*. Pour ce faire, j'ai réalisé la mise au point de l'approche de MOHCA-seq au laboratoire (**Figure 31**). J'ai ainsi pu développer et optimiser une approche de cartographie chimique multidimensionnelle destinée à l'étude d'un ARN de 604 nucléotides, soit un ARN trois fois plus long que ceux précédemment étudiés avec le MOHCA-seq. Pour ce faire, il a été nécessaire d'adapter différentes étapes de l'approche. En effet, le MOHCA-seq comporte différentes étapes cruciales qui peuvent résulter en une mauvaise interprétation des résultats dans le cas où elles ne sont pas réalisées correctement (**Figure 34**). Ces étapes sont [1] l'incorporation des nucléotides modifiés lors de la transcription *in vitro*, [2] la purification des ARN modifiés et [3] la préparation des banques de séquençage. Cependant, ces points n'ont pas été identifiés immédiatement. Lors de ma thèse, trois expériences préliminaires de MOHCA-seq ont été nécessaires à l'identification de ces différentes étapes délicates. Ainsi, la première approche, exclusivement basée sur le protocole développé par le laboratoire de Rhiju Das (**Figure 35A**), a permis de mettre en évidence l'importance de la méthode de préparation des banques de séquençage. La seconde tentative de MOHCA-seq, utilisant un kit NEBNext® Small RNA Library Prep Set for Illumina® (Multiplex Compatible) pour la préparation de la banque de séquence (**Figure 35B**), a montré l'importance du taux d'incorporation du nucléotide modifié et de la purification des ARN pour l'analyse des résultats. La troisième approche a également utilisé le kit NEBNext® après avoir corrigé les deux points précédents, mais n'a pas permis d'obtenir de résultats interprétables. Une nouvelle approche de préparation des banques de séquences a donc été envisagée. Ainsi, cette dernière approche s'est basée sur divers protocoles présents dans la littérature (260,295,298–301) principalement pour optimiser la préparation de la banque de séquences (**Figure 35C**).

Lors de ma thèse, j'ai donc progressivement détecté et corrigé certains biais inhérents à la stratégie du MOHCA-seq afin d'améliorer la qualité de résultats finaux.



**Figure 34** : Représentation schématique des biais de l'approche du MOHCA-seq. Les étapes sont identiques à celles présentées sur la **Figure 31**. La partie gauche de la figure représente les biais introduits par l'utilisation d'ARN tronqués ou dégradés avant l'étape de clivage. La partie droite met en avant les biais introduits par la présence de multiples nucléotides modifiés.



**Figure 35** : Représentation schématique des approches de préparation de banque de séquences réalisées. (A) L'approche initiale a strictement suivi la démarche expérimentale publiée par (260). (B) La seconde approche s'est basée sur le kit NEBNext®. (C) La dernière approche employée a été optimisée au laboratoire.

## I/ Approche 1: Cheng *et al.*, (2015)

Les premières tentatives de création de banque ont été réalisées en suivant le protocole publié pour le MOHCA-seq, en apportant le minimum d'adaptation possible (260) (**Figures 33 et 35A**). La préparation de la banque reposait sur la ligation d'un premier adaptateur de 17 nucléotides à l'extrémité 3' du fragment d'ARN ce qui permettait l'ajout de la séquence de la première amorce de séquençage lors de la transcription inverse. Cette séquence servait également à liguer la seconde amorce de séquençage par une pseudo-circulation de l'ADNc. Ainsi, j'ai suivi le protocole développé dans la publication et utilisé les différents réactifs recommandés aux concentrations préconisées par les auteurs. Les ARNs ont donc été synthétisés avec des ratios de nucléotides modifiés (2'NH<sub>2</sub>:2'OH) de 1:3, 2:5 et 1:2. Chaque ARN a été contrôlé sur gel après l'induction du clivage aux radicaux hydroxyles et présentait un profil caractéristique de dégradation tout en conservant une partie de l'ARN intact. Cependant, aucune amplification de la banque n'était visible à l'issue de la préparation, bien que chaque étape jusqu'à la transcription inverse ait été vérifiée. Il a donc été décidé de changer d'approche pour la préparation de la banque. En effet, le protocole initial était prévu pour des ARN de tailles autrement plus petites que celui étudié ici. De plus, depuis 2015, date de la publication, de nouvelles approches ont vu le jour pour le séquençage à haut débit et de nouveaux kits ont été développés pour faciliter la préparation des banques d'ADNc.

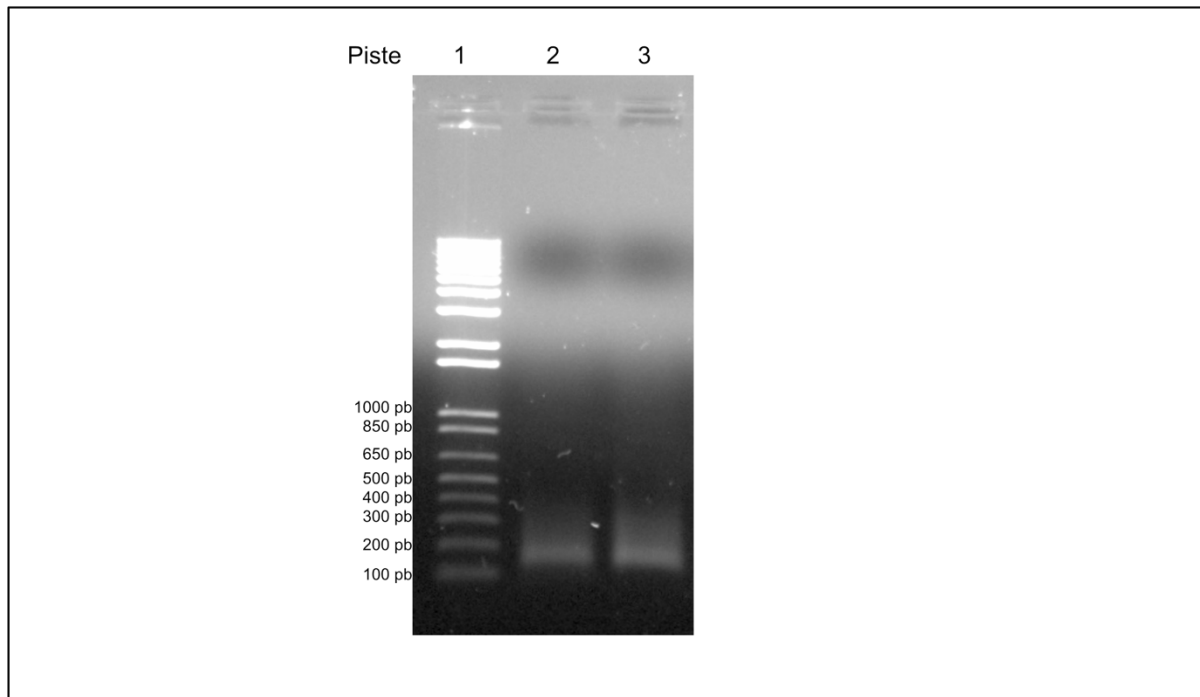
## II/ Approche 2: Kit NEBNext® Small RNA Library Prep Set for Illumina®

Après l'échec de la préparation de la librairie en suivant le protocole publié par Cheng *et al.*, (2015), nous avons opté pour l'utilisation du kit NEBNext® Small RNA Library Prep Set for Illumina® (Multiplex Compatible) (**Figure 35B**). En effet, ce kit est adapté à la préparation de petits ARN pour du séquençage à haut débit en mode paired-end pour des quantités de départ allant de 100 ng à 1 µg d'ARN totaux. La taille attendue pour les interactions à longue distance présentes dans les 604 premiers nucléotides du VIH-1 n'avait pas été définie avant le début de l'expérience. Cependant, nous pouvons supposer que le MOHCA-seq conduirait à la formation de fragments d'ARN allant jusqu'à au moins 370 nucléotides qui correspondent à l'interaction la plus longue identifiée à ce jour entre la boucle apicale de PolyA et la région



codante de *gag* (206). Le kit étant préconisé pour des ARN dont la taille est comprise entre 20 et 150 nucléotides, il était attendu qu'il ne permette pas d'identifier l'ensemble des interactions présentes. Cependant, ces expériences ont été supposées être suffisante pour servir de preuve de principe à l'application du MOHCA-seq à la région 5' terminale du VIH-1. Ainsi, j'ai appliqué l'approche de MOHCA-seq selon le protocole de Cheng *et al.*, (2015) pour les étapes précoces du protocole et au niveau de la préparation des banques de séquences j'ai utilisé le kit afin de valider notre protocole de MOHCA-seq. Deux conditions ont été réalisées en parallèle, avec le repliement de l'ARN dans un tampon favorisant la conformation monomérique ou dans un tampon favorisant la conformation dimérique.

Les premiers résultats obtenus par cette approche ont été bien plus encourageants que ceux du protocole original. En effet, l'analyse sur gel des produits PCR obtenus à la fin de la préparation de la banque a montré un léger smear pour les tailles d'amplicons allant de 120 pb jusqu'à 350 pb (**Figure 36**). Malgré la bande majoritaire correspondant aux dimères d'amorces à 120 pb, le léger smear correspond à des fragments théoriques d'ARN compris entre 20 et 230 nucléotides. Afin de confirmer la présence des séquences de l'ARNg du VIH-1, une petite fraction de la banque produite a été clonée dans un vecteur pJET1.2 (CloneJET PCR Cloning Kit). Le séquençage des amplicons clonés a permis de mettre en évidence la présence de séquences de la région 5' terminale de l'ARNg du VIH-1, flanquées par les séquences des adaptateurs 5' et 3' SR du kit NEBNext®. Ces résultats confirment donc l'efficacité de l'approche pour liguer les adaptateurs sur les fragments d'ARN issus du clivage aux radicaux hydroxyles.

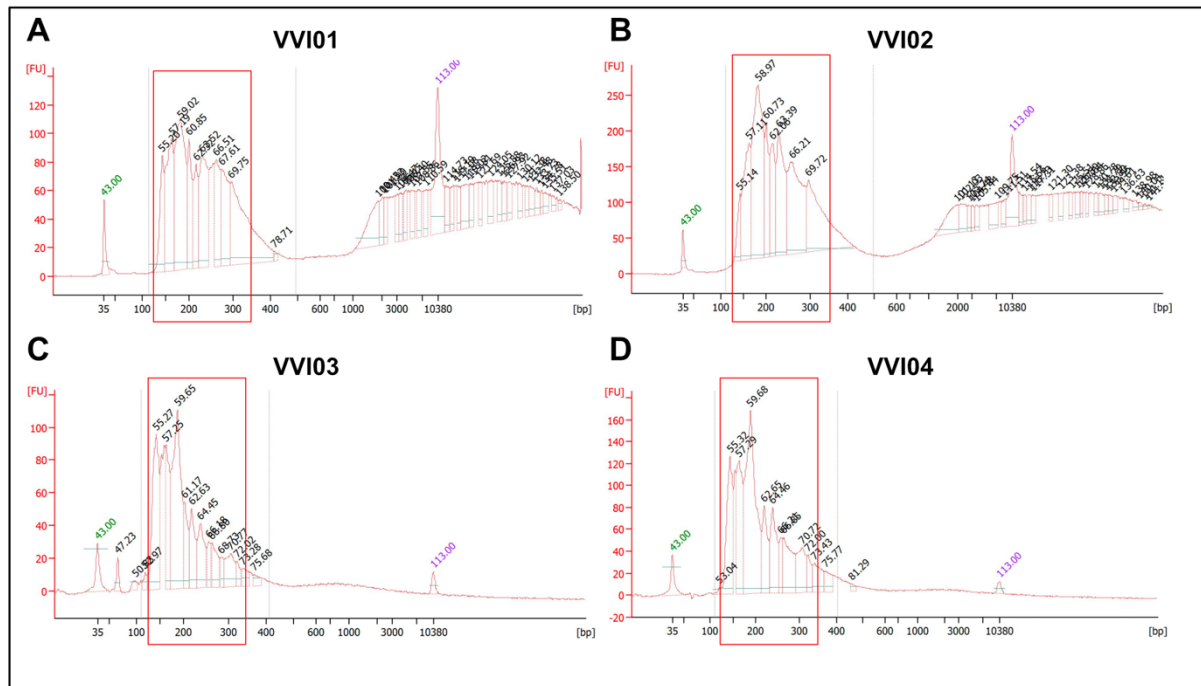


**Figure 36** : Analyse sur gel d’agarose des résultats préliminaires issus du kit NEBNext®. Les fragments d’ARN issus du clivage lors du MOHCA-seq ont été préparés à l’aide du kit NEBNext® Small RNA Library Prep Set for Illumina® (Multiplex Compatible) et amplifiés par PCR. Les pistes 2 et 3 correspondent aux échantillons repliés initialement dans le tampon favorisant la conformation monomérique ou dimérique de l’ARNg, respectivement. La piste 1 contient un marqueur de taille.

Une fois la validation de l’approche effectuée, les deux mêmes conditions ont été étudiées lors de la préparation de banque de séquençage. Ainsi, la région 5’ terminale de l’ARNg a été repliée dans un tampon favorisant soit la monomérisation (VVI01 et VVI03), soit la dimérisation (VVI02 et VVI04), juste avant que le clivage ne soit réalisé. Les fragments obtenus ont été séparés en deux lots pour être préparés en banques à l’aide du kit NEBNext® Small RNA Library Prep Set for Illumina® (Multiplex Compatible) par moi-même au laboratoire (VVI03 et VVI04) ou bien par la plateforme EpiRNA-Seq à Nancy (VVI01 et VVI02) qui a également effectué le séquençage. Chaque condition a été indexée lors de la ligation pour pouvoir multiplexer le séquençage. La plateforme a vérifié la qualité des banques formées à partir de chaque échantillon sur Bioanalyser (Agilent) et les résultats sont présentés en **Figure 37**. Globalement, les résultats obtenus sont cohérents avec ceux du gel présenté en **Figure 36**. En effet, les banques d’ADNc sont principalement comprises entre 150 pb et 350 pb (**Figure 37** – rectangles rouges). Cependant, les banques VVI01 et VVI02 présentent également un pic vers 1500 – 2000 pb qui n’est pas visible sur les banques VVI03 et VVI04. La présence

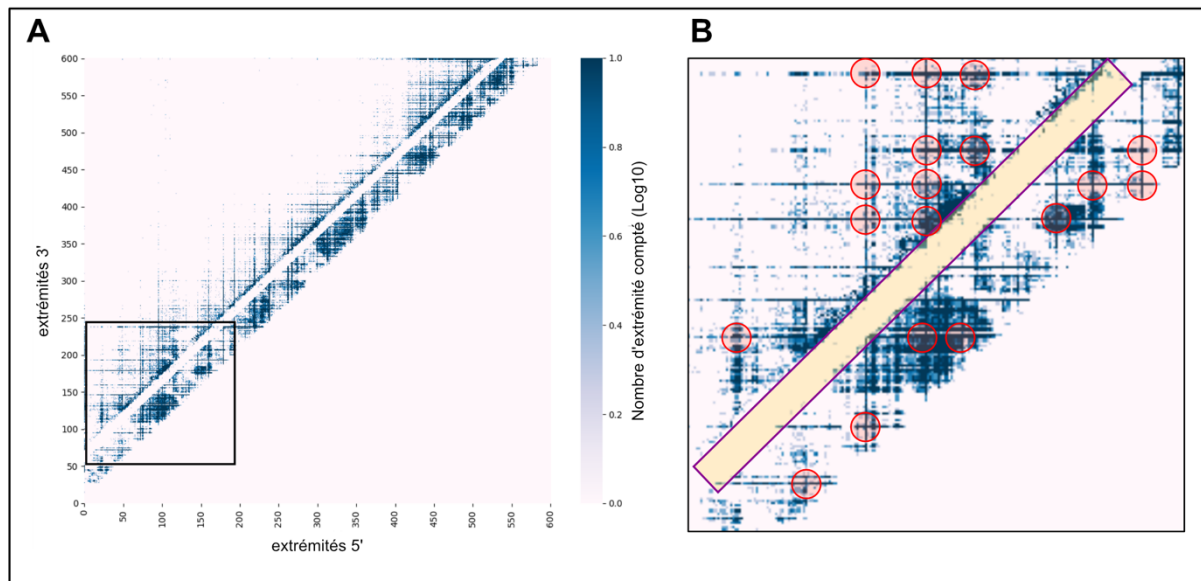
## AXE 2 – RESULTATS

de ce matériel de grandes tailles a été attribué à un artéfact apparu lors de l'analyse sur le Bioanalyzer (Agilent). Le service de la plateforme a conclu que la qualité des banques n'était pas impactée et le séquençage MiSeq en mode paired-end a été réalisé sur les 4 échantillons indexés. Le séquençage a conduit à l'obtention d'un total de 10,4 millions de lectures réparties entre les 4 banques de séquences.



**Figure 37** : Analyse des profils des banques de séquençage obtenus sur Bioanalyzer. Les ARN repliés dans des tampons favorisant la monomérisation (A et C) ou la dimérisation (B et D) ont été clivés et les fragments obtenus préparés à l'aide du kit NEBNext® à la plateforme EpiRNA-Seq de Nancy (A et B) ou au laboratoire (C et D). Les pics marqués en verts (35 pb) et violets (10 380 pb) sont des marqueurs de tailles internes. Les rectangles rouges entourent les fragments d'intérêt produits.

L'analyse préliminaire des résultats du séquençage a été réalisée en collaboration avec le Dr Laurence Despons et les données obtenues ont été exprimées sous forme de carte bidimensionnelle de la région 5' terminale de l'ARNg du VIH-1 (**Figure 38**). La répartition des fragments de la région 5' terminale séquencés est représentée en fonction des extrémités 5' et 3' de chacun d'entre eux. Ainsi chaque fragment d'ARN est représenté par un point dont l'abscisse correspond à la position de l'extrémité 5' sur la séquence de l'ARN et l'ordonnée correspond à la position de l'extrémité 3'.



**Figure 38** : Modélisation des données de séquençage à haut-débit. La position nucléotidique des extrémités de chaque fragment séquencé a été recueillie et représentée sur (A) la carte bidimensionnelle de la région 5' terminale de l'ARNg du VIH-1. (B) Le cadre noir entoure la région agrandie. L'ensemble de ces intersections n'est pas représenté. Le bandeau jaune central indique une région où les informations sont manquantes.

Cette analyse a permis de mettre en évidence trois biais dans notre approche. Premièrement, le séquençage en paired-end a révélé une écrasante majorité de petits fragments avec 99,27% de lectures de moins de 150 nucléotides (**Figure 38A**). Ce résultat, bien qu'attendu, reste un biais inhérent à l'utilisation d'un kit spécifique aux petits ARN. Deuxièmement, lors de l'analyse, des régions de clivage « excessif » sont apparues (**Figure 38B** – cercles rouges). Ces régions se manifestent sous forme de croix sur la carte, indiquant une sur-représentation des extrémités dans les séquences analysées. De plus, la majorité des fragments d'ARN retrouvés dans ces régions sont enrichis en adénines sur leurs deux extrémités. Ces résultats mis ensemble, semblent indiquer une sur-modification des ARN. En effet, l'incorporation de multiples nucléotides modifiés dans un unique transcrit favoriserait l'obtention de petits fragments dont les extrémités présenteraient majoritairement des adénines. Enfin, nous avons observé que les informations concernant les fragments de 70-75 nucléotides sont manquantes (**Figure 38B** – bandeau jaune). L'origine de ce biais a pu être relié à la nature de l'analyse bio-informatique et à la méthode de retrait des séquences des adaptateurs. Ainsi, cette approche nous a permis de repenser le MOHCA-seq appliqué à notre ARNg du VIH-1 et de corriger les biais identifiés.

### III/ Adaptation du MOHCA-seq à l'étude de la région 5' terminale de l'ARNg

#### III.1/ Taux d'incorporation des nucléotides modifiés et temps de clivage

Initialement, Cheng *et al.*, (2015) ont travaillé sur des ARN allant de 69 à 188 nts, contenant 20 à 40% d'adénines. Lors des synthèses *in vitro*, ils ont utilisé des ratios de nucléotides 2'NH<sub>2</sub> modifiés avec des nucléotides standards (2'NH<sub>2</sub>:2'OH) de 1:5, 1:2 et 5:4 sans observer de différences autres que celle d'intensité. Cependant, la taille de la région 5' terminale de l'ARNg du VIH-1 est de 604 nts et contient 180 adénines, soit 30% de la séquence. Ainsi, la probabilité d'incorporer de multiples nucléotides 2'NH<sub>2</sub> modifiés à ces ratios est plus importante. Il a donc été nécessaire d'optimiser cette étape pour permettre la synthèse d'une large majorité d'ARN ayant incorporé au plus, un unique nucléotide modifié. En effet, le principe de base du MOHCA-seq est d'induire plusieurs coupures dans un même brin d'ARN à partir d'une unique source radicalaire. Dans le cas de modifications multiples, différentes sources radicalaires sont présentes sur un même brin d'ARN (**Figure 34** – partie droite). Ces différentes sources peuvent donc induire des coupures excessives et ainsi conduire à la production de résultats apparentés à de faux-positifs. A l'inverse, une hypomodification des ARN n'a pas de conséquence aussi importante. Dans ce dernier cas, le risque principal est une perte d'information due au nombre trop faible de fragments d'intérêt à analyser par rapport à la totalité des ARN.

Afin de déterminer le meilleur taux de modifications pour mes expériences de MOHCA-seq, dans un premier temps, j'ai cherché à mettre en évidence la présence des nucléotides modifiés dans l'ARN et leur capacité à lier l'ITCB-EDTA. Pour ce faire, j'ai synthétisé des ARN 1-604 avec trois ratios de nucléotides 2'NH<sub>2</sub> (1:3, 2:5 et 1:2). Après avoir fixé l'ITCB-EDTA ou non (contrôle négatif), j'ai réalisé une réaction de transcription inverse selon le même protocole que pour le hSHAPE. La réaction de séquençage a été préparée en parallèle avec du ddTTP, incorporé en face des adénines lors de la transcription inverse, et permettant donc de valider la présence de l'ITCB-EDTA sur les adénines modifiées. Les ADNc synthétisés ont été analysés par électrophorèse capillaire et les résultats sont présentés en **Figure 39**. Les profils rouges correspondent aux ARN 2'NH<sub>2</sub> ayant fixé l'ITCB-EDTA et les profils bleus correspondent aux ARN 2'NH<sub>2</sub> sans ITCB-EDTA. Les profils verts et violets correspondent aux réactions de séquençage au ddTTP. Ainsi, les pics d'intensité fluorescente pour les profils de séquençage



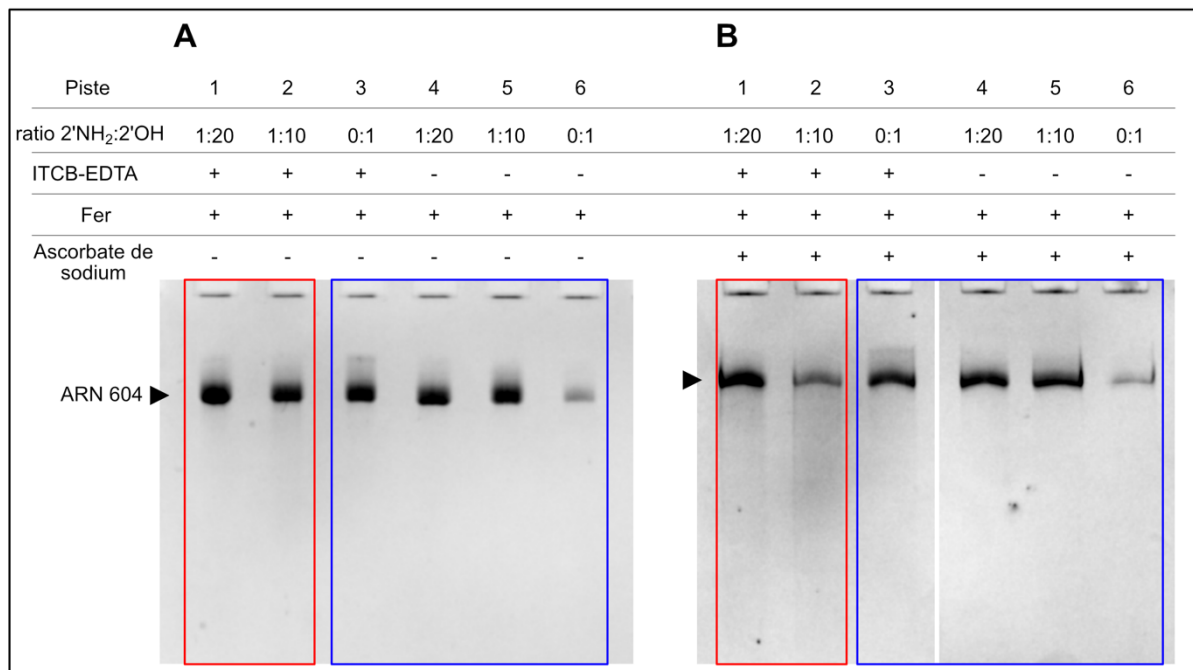
## AXE 2 – RESULTATS

$k \backslash \lambda$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1	1,2	1,5	2	2,5	3
0	0,905	0,819	0,741	0,670	0,607	0,549	0,497	0,449	0,407	0,368	0,301	0,223	0,135	0,082	0,050
1	0,090	0,164	0,222	0,268	0,303	0,329	0,348	0,359	0,366	0,368	0,361	0,335	0,271	0,205	0,149
2	0,005	0,016	0,033	0,054	0,076	0,099	0,122	0,144	0,165	0,184	0,217	0,251	0,271	0,257	0,224
3	0,000	0,001	0,003	0,007	0,013	0,020	0,028	0,038	0,049	0,061	0,087	0,126	0,180	0,214	0,224
4	0,000	0,000	0,000	0,001	0,002	0,003	0,005	0,008	0,011	0,015	0,026	0,047	0,090	0,134	0,168
5	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,002	0,003	0,006	0,014	0,036	0,067	0,101
6	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,004	0,012	0,028	0,050

**Tableau 10** : Table de la Loi de Poisson de paramètre  $\lambda$  et d'espérance  $k$ , représentant la probabilité d'incorporation de nucléotides 2'NH<sub>2</sub>. Les cases colorées en rouge montrent une probabilité d'incorporation supérieure à 15%, entre 0,5% et 15% pour les cases orange et à moins de 0,5% pour les cases vertes.

Le **Tableau 10** récapitule les probabilités d'occurrence d'un nucléotide modifié dans la séquence en supposant que cette incorporation suit la loi de Poisson. Ainsi, pour diminuer au maximum les risques d'incorporations multiples, il faut se placer dans des conditions où  $P(K \leq 1) \approx 1$  soit un  $\lambda = 0,1$ . Dans ces conditions, les probabilités d'incorporations multiples sont minimales et ~10% des ARN transcrits sont modifiés statistiquement une fois. Il faut donc une écrasante majorité d'ARN non modifiés pour éviter une sur-modification. Cependant, le paramètre  $\lambda$  ne se traduit pas en un ratio nucléotide 2'NH<sub>2</sub> contre nucléotide 2'OH. En effet, la polymérase n'utilise généralement pas les nucléotides modifiés comme des substrats équivalents aux nucléotides 2'OH naturels (302–304). De plus, la fixation de l'ITCB-EDTA sur les nucléotides modifiés et la chélation du fer par l'EDTA ne sont pas des réactions totales. Ainsi, pour déterminer un taux d'incorporation acceptable, je me suis donc basé sur un événement quantifiable. Dans le cas où le transcrit est modifié et fixe l'ITCB-EDTA, il peut dès lors chélater un ion Fe<sup>3+</sup> et induire un clivage sur l'ARN. J'ai donc transcrit des ARN à l'aide de différents ratios de nucléotides modifiés contre des nucléotides standards (1:2, 2:5, 1:3, 1:6, 1:10 et 1:20) avant d'induire les réactions de clivage. Seuls les ARNs transcrits avec des ratios 1:10 et 1:20 ont montré des profils de clivage cohérents avec ceux attendus dans le cas où la majorité des ARN ne porterait pas de modification, soit une grande majorité d'ARN non clivés (**Figure 40A** et **40B**). En effet, les résultats montrent que seuls les ARNs ayant incorporé du 2'NH<sub>2</sub> et fixé l'ITCB-EDTA sont capables de déclencher la production de radicaux hydroxyles en présence d'ascorbate de sodium (**Figure 40A** et **40B** – comparaison rectangles rouges). Les

autres ARN ne semblent pas subir de clivage (**Figure 40A** et **40B** – comparaison rectangles bleus).



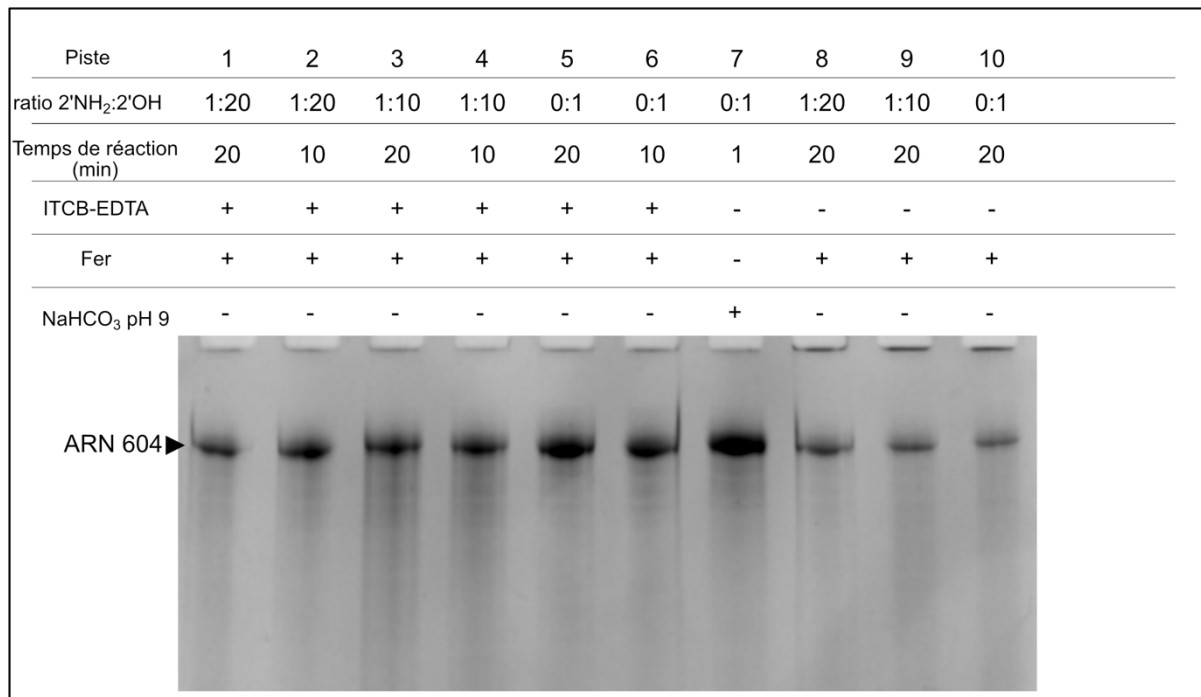
**Figure 40** : Validation de l'étape de clivage aux radicaux hydroxyles. Les ARN sont transcrits en présence de ratios 2'NH<sub>2</sub>:2'OH de 1:20 (pistes 1 et 4), 1:10 (pistes 2 et 5) ou 0:1 (pistes 3 et 6). L'ITCB-EDTA est fixé (pistes 1, 2 et 3) ou non (4, 5 et 6) avant que les ARN ne soient mis en présence de fer. Les ARN sont ensuite supplémentés d'ascorbate de sodium (B) ou non (A) pour induire un clivage. Les rectangles rouges encadrent les ARN où un clivage est attendu lorsque l'ascorbate est présent, les rectangles bleus encadrent ceux où le clivage ne devrait pas se produire. La flèche noire indique l'ARN de taille 604.

Une fois le protocole d'incorporation optimisé, j'ai vérifié l'influence du temps de clivage sur l'ARN. En effet, ce temps de réaction correspond à la période où les radicaux hydroxyles peuvent être produits. Ainsi, une durée plus ou moins longue pourrait avoir un impact sur le bruit de fond. Cette étape a été testée par les concepteurs de l'expérience sans y voir de différences entre 5, 10 et 30 min de réaction de clivage (260). Les auteurs ont donc opté pour un temps de clivage de 10 min. Cependant, nous avons également voulu nous assurer de la robustesse de ce protocole dans nos conditions. Ainsi, j'ai comparé les profils de différents ARN après 10 ou 20 min de réaction en présence d'ascorbate de sodium à 100 mM (**Figure 41**). Différents contrôles ont été réalisés pour vérifier la qualité des résultats. Ces contrôles comprennent des ARN synthétisés en absence de nucléotides modifiés 2'NH<sub>2</sub> (**Figure 41** - pistes 5, 6, 7 et 10) ou n'ayant pas été mis en contact avec de l'ITCB-EDTA (**Figure 41** - pistes 7 à 10). De plus, un contrôle de clivage à ~10% de la totalité des ARN a été



## AXE 2 – RESULTATS

réalisé par un traitement au bicarbonate de sodium ( $\text{NaHCO}_3$ ) à pH 9 suivi d'une incubation 1 min à  $90^\circ\text{C}$  (**Figure 41** – piste 7). Cette condition de clivage alcalin a été optimisée pour servir de contrôle de bruit de fond en cas de clivage aléatoire (**Annexe 2**). Les résultats, bien que peu concluants au regard des contrôles non clivés (**Figure 41** – pistes 5, 6, 8, 9 et 10), ne montrent pas non plus de différence significative dans le clivage de la bande principale des ARN modifiés (**Figure 41** – comparer les pistes 1-2 et 3-4), ce qui suggère un clivage similaire entre 10 et 20 min de réaction. Cependant, il n'est pas possible d'analyser plus en détails ces résultats du fait de la pauvre qualité des contrôles négatifs.



**Figure 41** : Vérification de l'impact du temps de réaction sur le clivage des ARN modifiés ou non. Le profil de clivage entre 20 min de réaction (pistes 1, 3 et 5) et 10 min de réaction (pistes 2, 4 et 6) est comparé pour les ARN synthétisés avec différents ratios de nucléotides 2'NH<sub>2</sub>:2'OH. Trois contrôles négatifs et un contrôle positif sont réalisés en parallèle. Les contrôles négatifs consistent en 3 ARN sans ITCB-EDTA (pistes 8, 9 et 10) et le contrôle positif est issu d'une réaction de clivage aléatoire due à un traitement au bicarbonate de sodium ( $\text{NaHCO}_3$ ) à pH 9 (piste 7) et une incubation à  $90^\circ\text{C}$  pendant 1 min. La flèche noire indique la position de l'ARN de taille 604 nts.

### III.2/ Purification des ARN modifiés

Une fois les ARN modifiés synthétisés, il est nécessaire d'éliminer les potentiels transcrits abortifs issus de la transcription. En effet, leur présence agirait de la même façon qu'une sur-modification des ARN. La présence d'une extrémité non native de l'ARN serait

perçue lors de l'analyse comme une résultante d'un clivage (**Figure 34** – partie gauche). Ainsi, une interaction à longue distance pourrait être faussement identifiée.

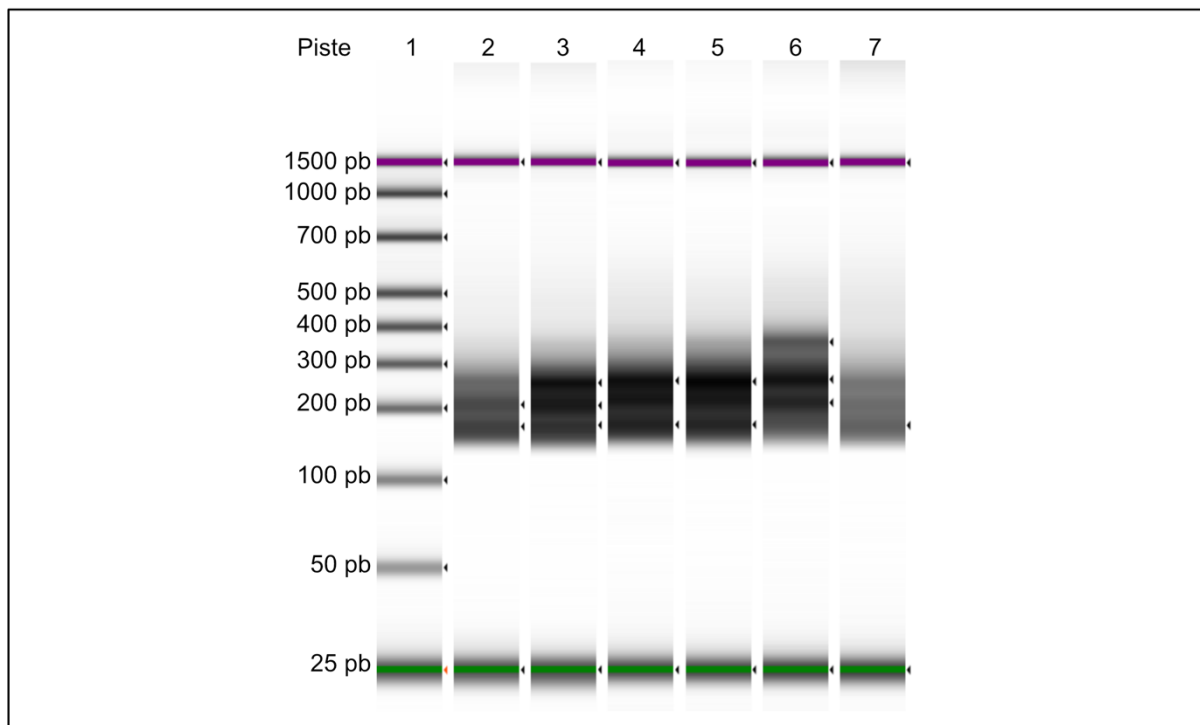
Initialement, la purification des ARN était réalisée sur une colonne TSK G4000SW, de façon identique à celle décrite dans le matériel et méthode de l'axe 1. Cependant, cette méthode de purification ne permet pas de retirer la totalité des produits abortifs présents dans la solution et une petite partie peut rester dans les ARN purifiés. De plus, les ARN dont la taille est proche de celle ciblée sont également conservés. La présence de ces transcrits abortifs est peu gênante dans le cadre du hSHAPE du fait de l'absence du site d'hybridation des oligonucléotides de transcription inverse en 3'. Cependant, dans le cas du MOHCA-seq, cette petite quantité d'ARN abortifs peut se révéler extrêmement délétère lors de l'analyse des résultats (**Figure 34** – partie gauche). En effet, dans des conditions où une minorité d'ARN sont modifiés sur 1 des 180 positions possibles, la présence d'une extrémité différente est perçue comme le signe d'un clivage par les radicaux hydroxyles. Ainsi, la présence de molécules d'ARN tronquées, de produits abortifs ou clivés est une source de biais considérable pour l'analyse des résultats et il a été nécessaire de s'affranchir de ce biais potentiel. Dans ce but, la méthode de purification des ARN a été modifiée et une purification à partir d'un gel de polyacrylamide à 6% d'acrylamide et 8 M d'urée a été choisie. Différents modes d'élution (passive ou active, broyage ou découpage du gel, variation du tampon d'élution, du temps et de la température) de l'ARN ont été testés avant d'opter pour une électro-élution des ARN hors des bandes de gels, ce qui permet un meilleur rendement de la purification. En effet, cette approche permet un rendement d'environ 40 - 50%, contre 15 - 30% pour les autres méthodes d'élutions testées.

### IV/ Approche 3: Kit NEBNext® Small RNA Library Prep Set for Illumina®

A la suite de l'optimisation des différentes étapes précoces du MOHCA-seq, les changements de protocole ont été appliqués au traitement des ARN tout au long de l'approche. Cependant, il était nécessaire de s'assurer que les conditions expérimentales définies étaient adaptées à l'étude de notre ARN. Pour ce faire, 6 conditions différentes ont été testées. Les deux premières conditions (VVI05 – VVI06) ont utilisé des ARN 2'NH<sub>2</sub> obtenus par transcription avec un ratio 1:20, modifiés à l'ITCB-EDTA et incubé 20 ou 10 min avec l'ascorbate de sodium à 100 mM, respectivement. Les deux conditions suivantes (VVI07 – VVI08) étaient identiques aux deux premières avec des ARN 2'NH<sub>2</sub> à un ratio 1:10. La condition

## AXE 2 – RESULTATS

VVI09 a utilisé un ARN non modifié dont le clivage alcalin a été réalisé par une incubation à 90°C pendant 1 min en présence de bicarbonate de sodium (NaHCO<sub>3</sub>) à pH 9. Cette condition sert de contrôle et correspond à environ 10% de clivage aléatoire comme montré en **Annexe 2**. La dernière condition (VVI10) a également été réalisée à l'aide d'un ARN non modifié et mis en présence d'ITCB-EDTA puis d'ascorbate de sodium à 100 mM pendant 20 min. L'ensemble des conditions a été réalisé dans un tampon de repliement favorisant la dimérisation. Les fragments d'ARN obtenus dans ces 6 conditions ont ensuite été envoyés à la plateforme EpiRNA-Seq de Nancy pour être convertis en banque de séquençage puis séquencés sur un appareil MiSeq en mode paired-end (75 pb). La plateforme a contrôlé la qualité des banques sur un Bioanalyser (Agilent) et les résultats sont présentés sur la **Figure 42**.



**Figure 42** : Gel théorique des résultats du contrôle au Bioanalyser pour les banques de séquences réalisées à l'aide du kit NEBNext®. Les échantillons VVI05 à VVI08 sont présents dans les pistes 2 à 5, respectivement. Les contrôles VVI09 et VVI10 ont été déposés dans les pistes 6 et 7, respectivement. La piste 1 correspond au marqueur de taille. Les bandes vertes (25 pb) et violettes (1 500 pb) sont des contrôles de taille internes.

La piste 1 correspond à l'échelle de taille, les pistes 2 à 5 contiennent les échantillons VVI05 à VVI08. Les pistes 6 et 7 contiennent les contrôles de clivage au bicarbonate de sodium (VVI09)

## AXE 2 – RESULTATS

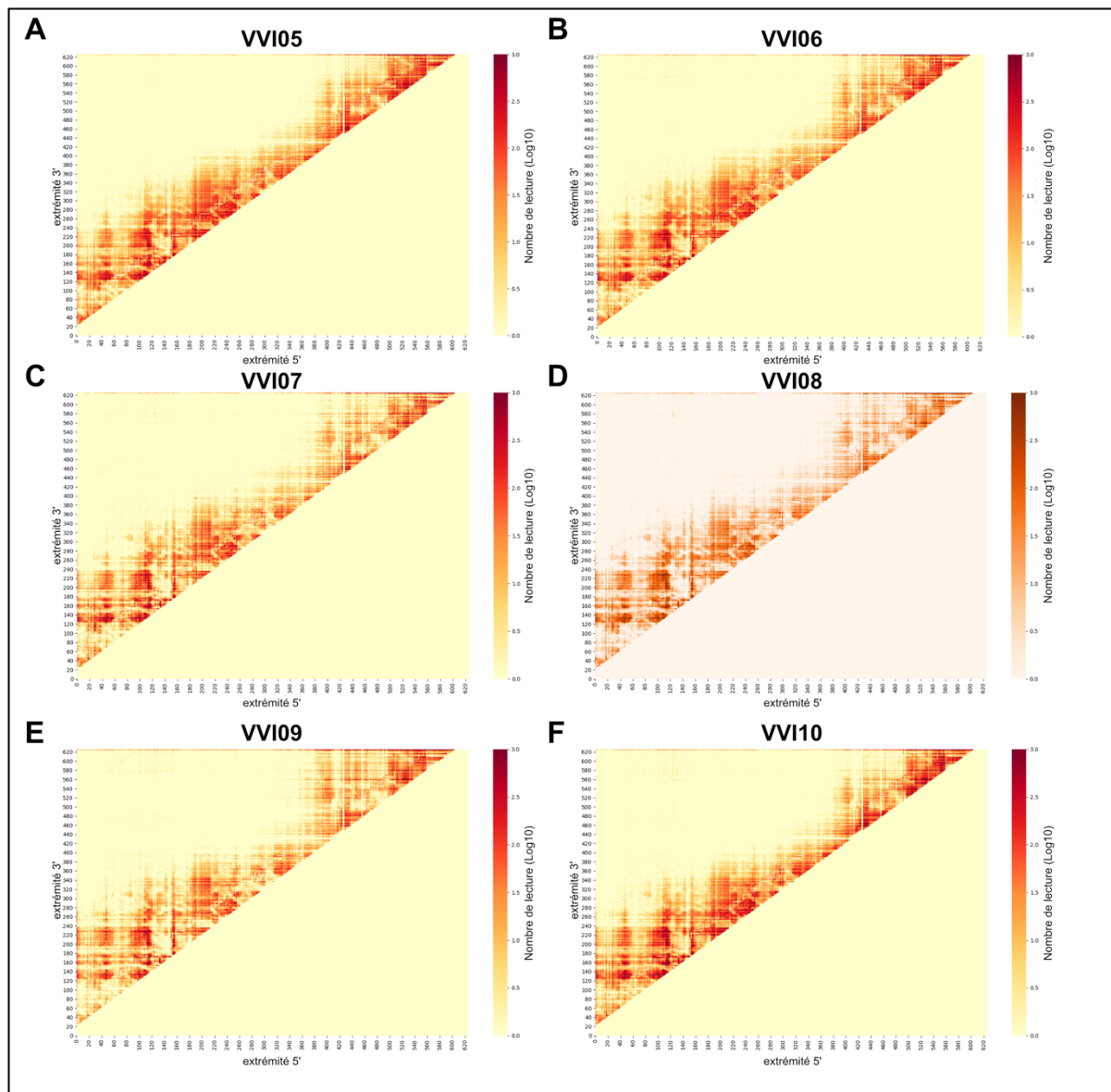
et le contrôle non clivé (VVI10), respectivement. Malgré l'utilisation d'un kit dédié à la préparation de petits ARN pour le séquençage, nous pensions pouvoir obtenir une petite partie de séquences correspondant à l'ARN non clivé à ~720 pb, au moins dans le contrôle non clivé VVI10. L'absence de cette bande au Bioanalyser est un indicateur que la banque produite dans nos conditions ne permet pas de détecter (ou très peu) les fragments de grande taille. De plus, la présence de bandes identiques pour l'ensemble des échantillons indique une possible contamination ou une fragilité de l'ARN sur des points particuliers. Cette fragilité n'ayant pas été détectée préalablement, il est peu probable qu'elle soit apparue pendant la création de la banque. Il paraît donc plus logique de parler de contamination. De plus, ces bandes contaminantes semblent séparées de façon régulière. Il a donc été spéculé qu'elles correspondaient à des concatémères d'oligonucléotides, sans pouvoir vérifier cette hypothèse. Malgré ces points d'incertitude, la plateforme a réalisé le séquençage de la banque.

Ce séquençage a permis l'obtention de 25 millions de lectures réparties sur les 6 conditions préparées. L'analyse des données issues de ce séquençage a également été réalisée par le Dr Laurence Despons. Cependant, une incohérence est directement apparue dans l'analyse préliminaire (**Tableau 11**). En effet, les concentrations des banques VVI06 à VVI09 sont supérieures au contrôle négatif VVI10, ce qui implique une quantité supérieure de matériel dans ces conditions que dans VVI10. Cette observation est confirmée sur le gel théorique (**Figure 42**) où la piste 7 est la moins intense. Il est donc attendu que le contrôle négatif corresponde au bruit de fond. Cependant, après le séquençage, la condition VVI10 est celle qui présente le plus de lectures avec 5,5 millions, soit près d'un million de plus que la seconde condition avec le plus de lectures de séquences (VVI05). Ces résultats suggèrent que la majorité des lectures sont dues au bruit de fond. Ces résultats préliminaires étaient peu encourageants par rapport à la qualité des banques créées et aux informations qui pouvaient en être tirées. Malgré cela, l'analyse a été poursuivie et les résultats ont été modélisés afin de valider la méthode d'analyse et d'identifier de nouveaux biais éventuels (**Figure 43**).

Condition	Concentration (ng/ $\mu$ L)	Lectures (millions)
VVI05	4,86	4,6
VVI06	6,22	4,3
VVI07	6,36	3,7
VVI08	6,42	3,1
VVI09	6,84	3,6
VVI10	5,28	5,5

**Tableau 11** : Résultats préliminaires de l’analyse des banques de séquences réalisées par la plateforme EpiRNA-Seq. Les concentrations en amplicons et le nombre de lectures obtenues après le séquençage sont rapportées pour chaque condition.

Les profils sont relativement similaires pour chaque condition. Ces résultats étaient attendus pour les conditions VVI05 à VVI08, car le repliement des ARN dans chaque condition avait été réalisé dans un unique tampon. De plus, les temps de clivage de 10 – 20 min et la différence entre les taux de nucléotides incorporés (1:20 et 1:10) ne semblaient pas impacter énormément les profils de clivage lors de la mise au point. Cependant, les profils attendus pour les conditions VVI09 et VVI10 n’était pas supposé être aussi similaires aux conditions VVI05 – VVI08. En effet, le contrôle VVI09 devrait correspondre au résultat d’un clivage aléatoire et le contrôle VVI10 à celui d’une absence de clivage. Face à ces résultats, nous avons tenté d’identifier la raison des similarités observées. Plusieurs hypothèses ont été formulées, comme précédemment, mais la seule conjecture considérant l’ensemble des données dans l’explication était celle de la contamination des banques. En effet, lors d’une étape précédant l’ajout des adaptateurs aux extrémités 5’ et 3’, un même événement de contamination a dû se produire dans l’ensemble des tubes. Finalement, l’analyse n’a pas été poursuivie, mais les données obtenues nous ont à nouveau aidé à affiner le protocole de MOHCA-seq. En effet, l’utilisation d’un kit spécialisé dans les petits ARN ne convient pas pour l’approche lorsqu’elle s’intéresse à des ARN de plusieurs centaines de nucléotides. Or, il n’existe pas, à ce jour et à notre connaissance, de kit permettant de préparer une banque composée de fragments d’ARN de tailles comprises entre 20 et 600 nucléotides et basé sur une autre méthode que celle du « random priming » qui est incompatible avec notre stratégie. Il a donc été décidé d’optimiser une approche de préparation de la banque de MOHCA-seq sans utiliser un kit commercial.



**Figure 43 :** Modélisation des données de séquençage à haut-débit sous forme de cartes bidimensionnelles. Les résultats de chacune des conditions expérimentales sont représentés.

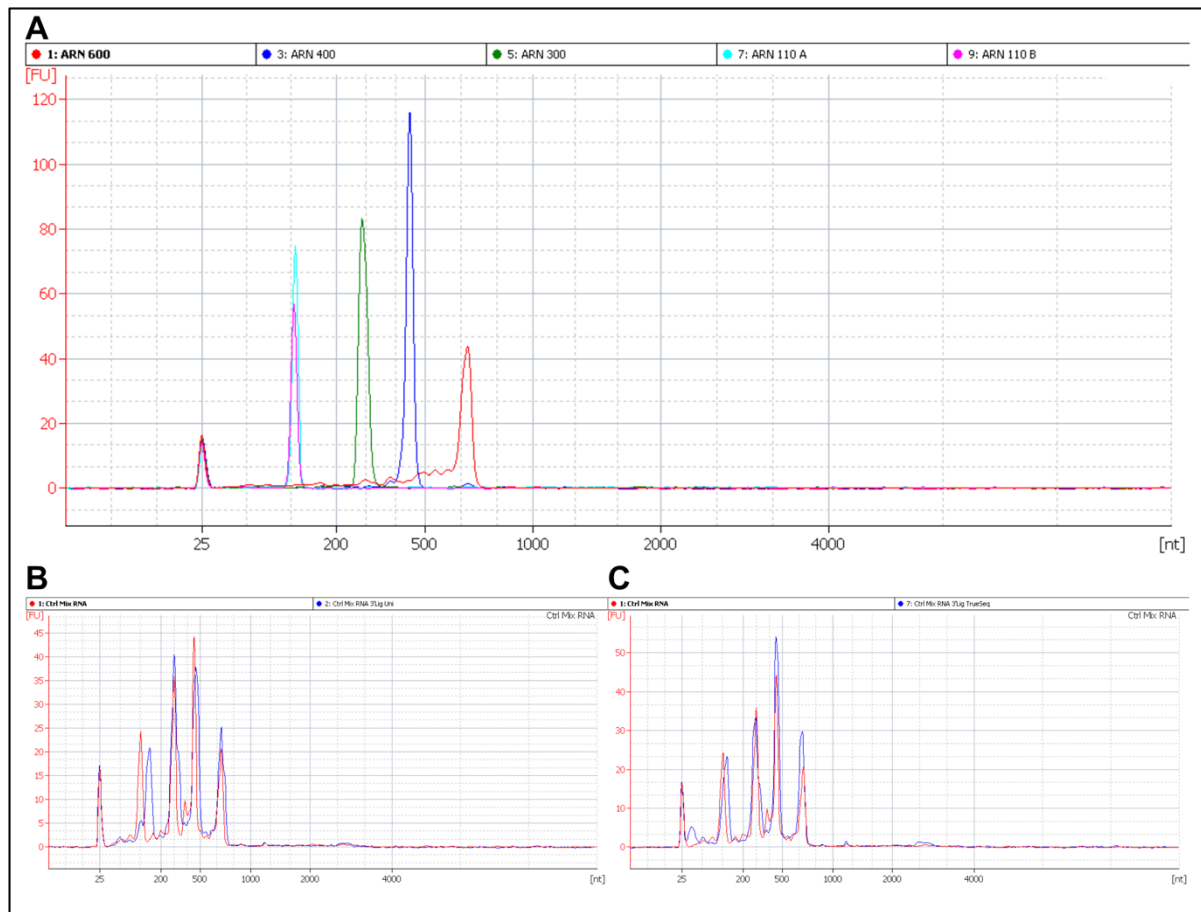
## V/ Mise au point de la préparation de la banque de séquences pour le MOHCA-seq

Le choix de l'approche pour la préparation des banques de séquences s'est basé sur la nécessité de préparer à la fois des petits et des grands ARN. Ainsi, lors de la mise au point de la méthode, différentes tailles d'ARN ont été testées. En effet, le clivage de la région 5' terminale de l'ARNg du VIH-1 peut théoriquement permettre l'obtention de fragments allant de quelques nucléotides à 604 nucléotides pour les ARN non clivés. De plus, du fait de l'hétérogénéité de tailles et d'extrémités dans le mélange, de nombreux biais peuvent être introduits dans la banque finale (298,299). Ainsi, j'ai décidé de réaliser les étapes d'optimisation en utilisant un mélange de quatre ARN de tailles différentes (110, 295, 400 et 604 nts) pour m'assurer que l'approche choisie permette de préparer au mieux les banques correspondant aux différentes tailles d'ARN. Pour les étapes précoces du protocole, j'ai mis en œuvre les méthodes précédemment validées, ensuite j'ai réalisé divers tests d'optimisation afin de valider chacune des étapes de la préparation de la banque (**Figure 35C**). Ainsi, les étapes de ligation en 3' des fragments d'ARN, de rétrotranscription, d'élimination de la matrice ARN et des oligonucléotides amorces de rétrotranscription, de ligation en 3' des ADNc et de PCR ont été optimisées.

### V.1/ Mise au point de la ligation en 3'

Dans un premier temps, je me suis assuré de la pureté des ARN utilisés pour les expériences de mise au point de la banque (**Figure 44A**). J'ai ensuite testé deux adaptateurs pour la ligation en 3' des ARN, l'adaptateur utilisé par Cheng *et al.*, (2015) et celui préconisé pour le séquençage Illumina® à l'aide des kit TruSeq (**Figures 44B** et **44C**). Dans les deux cas, les résultats ont montré que la ligation n'était pas efficace à 100%, mais un déplacement net est visible pour le pic correspondant à l'ARN de 110 nucléotides dans les deux conditions. Comme attendu ce déplacement se transforme en épaulement pour les ARN de plus grande taille (295, 400 et 604) dans le cas de l'adaptateur Universal miRNA Cloning Linker (NEB®) (**Figure 44B**). Dans le cas de l'adaptateur TruSeq, le déplacement est beaucoup moins visible (**Figure 44C**) mais globalement, les deux adaptateurs semblent être ligués en 3' des ARN testés.

## AXE 2 – RESULTATS



**Figure 44** : Contrôle de la qualité des ARN produits *in vitro* et validation de l'étape de ligation en 3' sur une puce RNA 6000 Nano au Bioanalyser. (A) La pureté des ARN de taille 110 (cyan et rose), 295 (vert), 400 (bleu) et 604 nts (rouge) a été vérifiée. (B et C) Les ARN ont ensuite été mélangé ensemble (profils rouges) et ligués en 3' (profils bleus) avec (B) un adaptateur Universal miRNA Cloning Linker de NEB® ou (C) un adaptateur TruSeq. Le premier pic de chaque électrophérogramme est un contrôle de taille interne.

### V.2/ Mise au point de la transcription inverse, de la seconde ligation et de l'amplification PCR

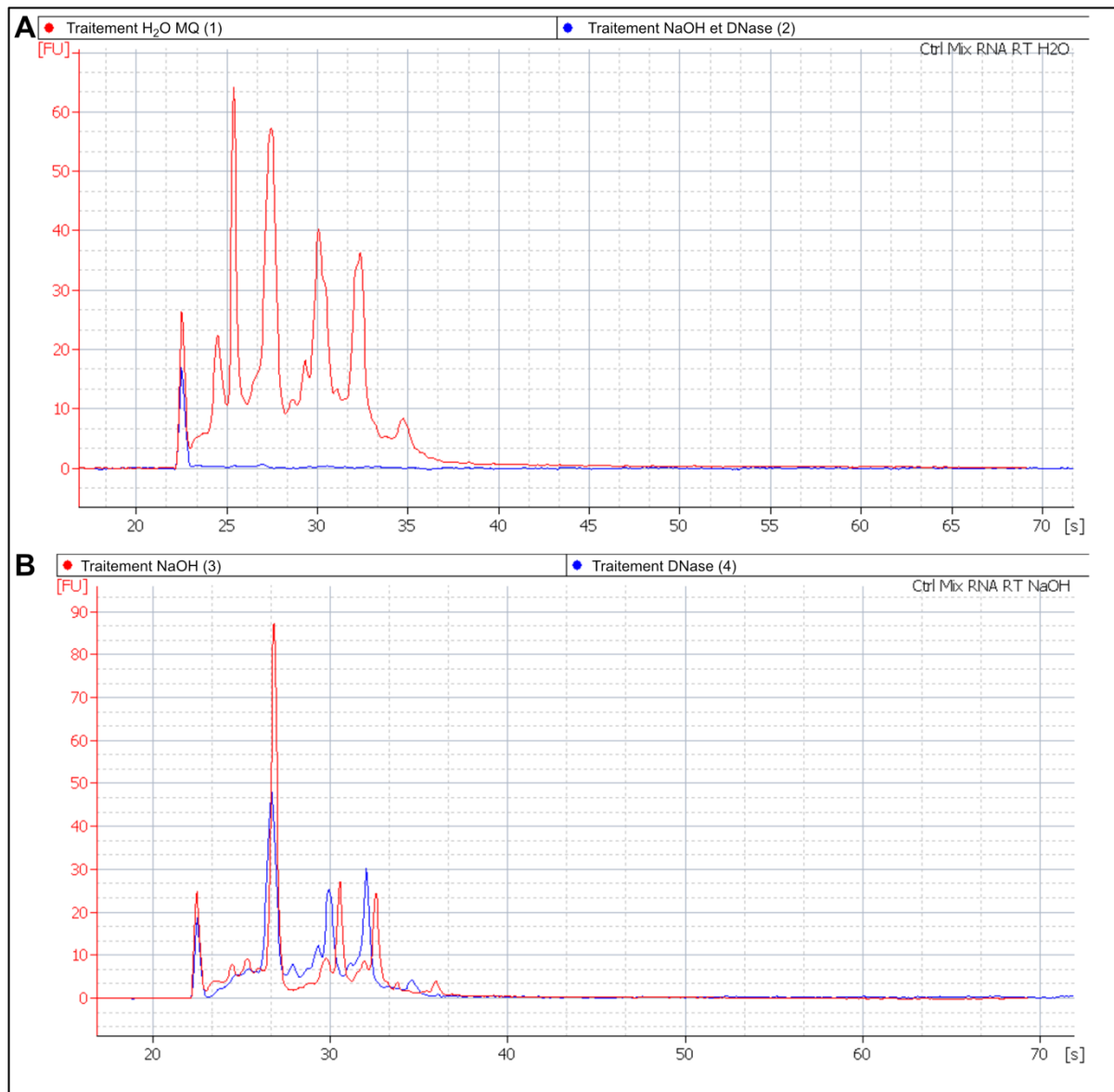
L'étape de transcription inverse en elle-même n'a nécessité que peu d'adaptation par rapport aux conditions préconisées pour la Reverse Transcriptase SuperScript IV (Invitrogen). Cependant, j'ai dû optimiser les conditions du retrait de la matrice ARN et de l'oligonucléotide amorce utilisé en excès lors de la transcription inverse. En effet, il est nécessaire d'éliminer la matrice ARN avant de réaliser la seconde ligation puisque dans le milieu réactionnel, certains ARN ne sont pas ligués en 3' et peuvent donc parasiter la seconde réaction de ligation. De même, la ligation du second adaptateur sur les oligonucléotides amorces utilisés pour la transcription inverse peuvent également conduire à une réaction parasite (305). Ces deux



espèces doivent donc être retirées pour s'assurer de liquer le second adaptateur sur un maximum d'ADNc issus de la transcription inverse et contenant donc la séquence du premier adaptateur.

Pour dégrader la matrice ARN, j'ai testé différents protocoles utilisant soit un mélange de RNase A (Ambion), de la RNase H (Invitrogen) ou bien de la soude (NaOH). Les tentatives enzymatiques n'ont pas donné de résultats très concluants, contrairement à l'approche NaOH. Ainsi, j'ai optimisé le protocole pour dégrader au maximum les ARN en solution sans affecter les ADNc. Les conditions finales de la dégradation des ARN sont présentées dans le Matériel & Méthodes. Pour vérifier l'effet du NaOH à la fois sur les ARN et les ADNc en solution, quatre conditions ont été testées. Après avoir ligué les 4 ARNs, une réaction de transcription inverse a été réalisée. Les complexes ARN/ADNc ont alors été traités (1) avec de H<sub>2</sub>O MQ, (2) uniquement au NaOH, (3) uniquement à la DNase ou (4) avec une combinaison NaOH et DNase. Les résultats sont présentés sur la **Figure 45**. Premièrement, la combinaison de traitements NaOH et DNase en combinaison élimine toute trace d'acide nucléique dans la solution (**Figure 45A** – profil bleu). Cependant, la DNase pourrait avoir un effet de dégradation de l'ARN matrice. Cette hypothèse a été rejetée après avoir testé la DNase et le NaOH séparément (**Figure 45B** – profils bleu et rouge, respectivement). De plus, cette expérience a montré que des ARN et des ADNc de tailles identiques ne migrent pas de la même manière lors d'un passage sur Bioanalyser. Ainsi, les résultats présentés montrent bien la spécificité de la DNase pour l'ADNc et celle du NaOH pour l'ARN.

## AXE 2 – RESULTATS

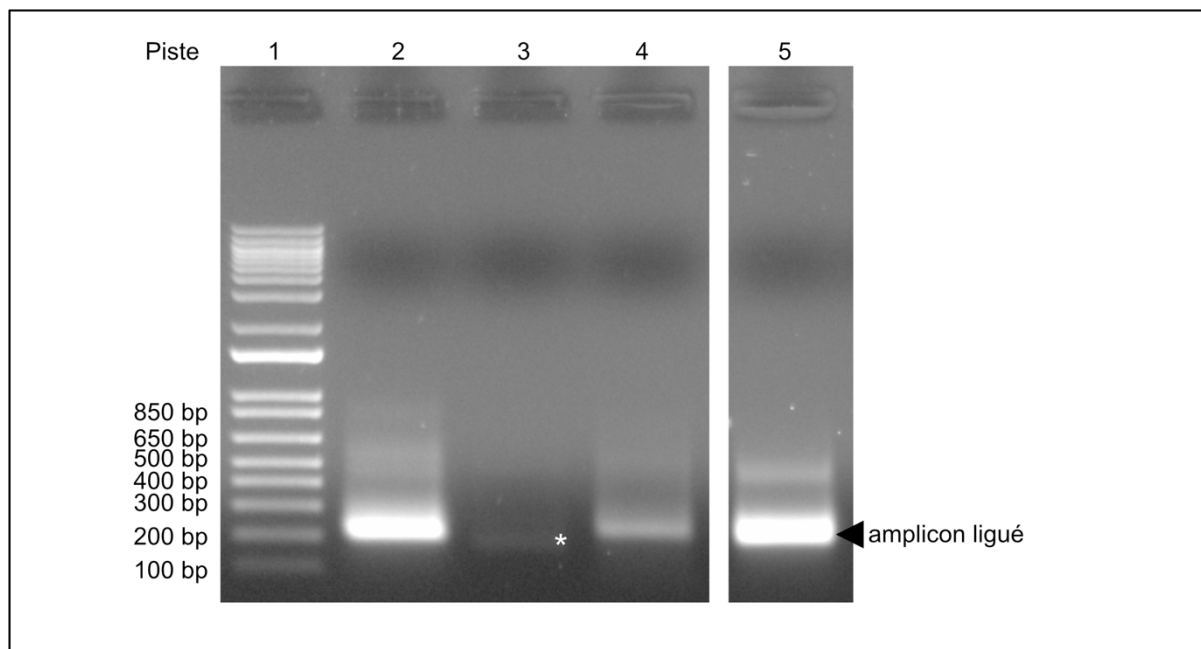


**Figure 45** : Validation sur une puce RNA 6000 Nano au Bioanalyser de l'utilisation de la soude (NaOH) pour la dégradation de l'ARN matriciel. L'effet du NaOH a été contrôlé dans quatre conditions différentes. Le mélange d'ARN a été rétrotranscrit et séparé en 4 pour subir différents traitements. Le mélange d'ARN et d'ADNc a donc été traité [1] uniquement à H<sub>2</sub>O MQ, [2] au NaOH puis à la DNase I, [3] uniquement au NaOH, [4] uniquement à la DNase I. (A) Comparaison des profils du mélange ARN/ADNc traité [1] uniquement à H<sub>2</sub>O MQ ou [2] au NaOH puis à la DNase I. (B) Comparaison des profils du mélange ARN/ADNc traité [3] uniquement au NaOH ou [4] uniquement à la DNase I.

Pour finaliser le protocole de transcription inverse et améliorer la qualité de la ligation en 3' des ADNc, j'ai choisi d'éliminer les oligonucléotides en excès dans le milieu à la fin de la réaction en introduisant une étape de dégradation par de l'exonucléase *d'Escherichia coli Exo1* selon le protocole développé par Enroth *et al.*, (2019). Ce traitement réalisé à la fin de la

## AXE 2 – RESULTATS

transcription inverse permet la dégradation des ADN simples brins en solution. Ainsi, à la fin de l'étape de rétrotranscription, les ADN néosynthétisés qui sont toujours sous forme d'hybride ARN/ADNc avec la matrice ARN ne seront pas dégradés alors que le traitement *ExoI* dégradera les oligonucléotides amorces utilisés en excès pour initier la transcription inverse. Afin de s'assurer que l'effet du traitement impactait positivement la seconde ligation, j'ai réalisé la transcription inverse d'un unique ARN en ADNc de 184 nucléotides puis j'ai effectué la ligation avec un adaptateur de 33 nucléotides. La ligation a été validée par une amplification PCR dont le produit théorique correspond à ~220 pb. Le mélange ARN/ADNc a donc été séparé en quatre conditions pour être traité soit [1] à l'*ExoI* avant d'être ligué en 3', [2] à l'*ExoI* puis au NaOH sans être ligué, [3] au NaOH avant d'être ligué ou [4] à l'*ExoI* puis au NaOH avant d'être ligué. Chaque condition a été amplifiée par PCR à l'aide d'amorces spécifiques et les résultats sont présentés en **Figure 46**. La condition non liguée (2 – piste 3) ne montre pas d'amplification mais une faible bande (astérisque blanc) peut être détecté légèrement sous 200 pb. Ce signal provient sûrement d'une amplification linéaire. Pour les conditions liguées, l'utilisation d'*ExoI* (1- piste 2 et 4- piste 5) améliore grandement l'amplification PCR des produits ligués et cela d'autant plus que la matrice ARN a été éliminée (4- piste 5), ce qui suggère que la seconde ligation a été affectée positivement par le traitement *ExoI*.

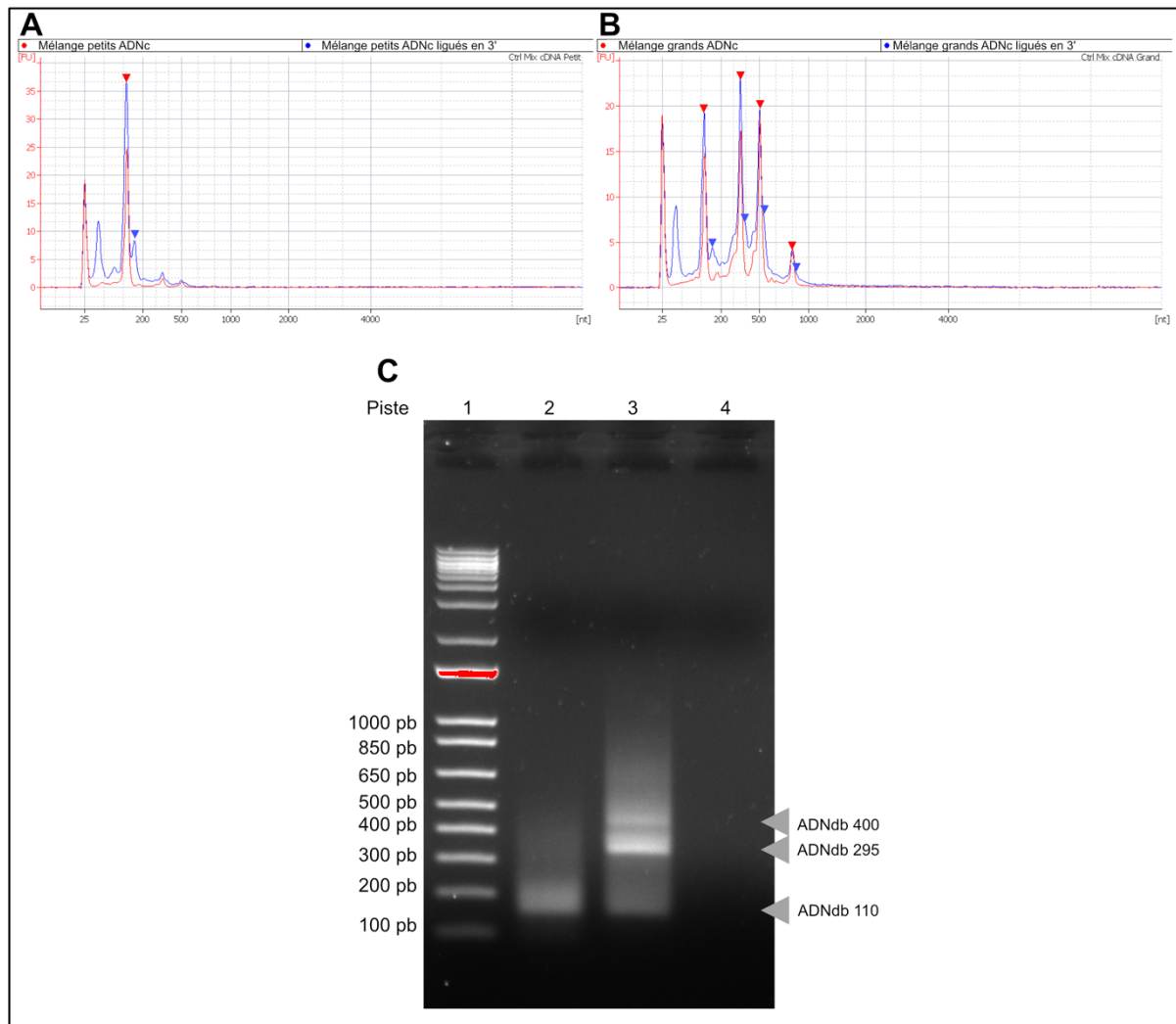


**Figure 46** : Vérification sur gel d'agarose de l'impact du traitement à l'*ExoI* sur la seconde ligation lors de la préparation de la banque. Un ARN a été rétrotranscrit en ADNc de 184 nucléotides et ligué ou non avec un adaptateur de 33 nucléotides. Les ADNc ligués ont ensuite été amplifiés par PCR et le produit attendu fait ~220 pb. Le mélange ARN/ADNc a donc été traité soit (1 – piste 2) à l'*ExoI* puis ligué en 3', (2 – piste 3) traité à

## AXE 2 – RESULTATS

l'ExoI puis NaOH sans être ligué, (3 – piste 4) traité au NaOH puis ligué en 3' ou (4 – piste 5) traité à l'ExoI puis au NaOH avant d'être ligué en 3'. L'astérisque blanc désigne une bande issue d'une amplification linéaire.

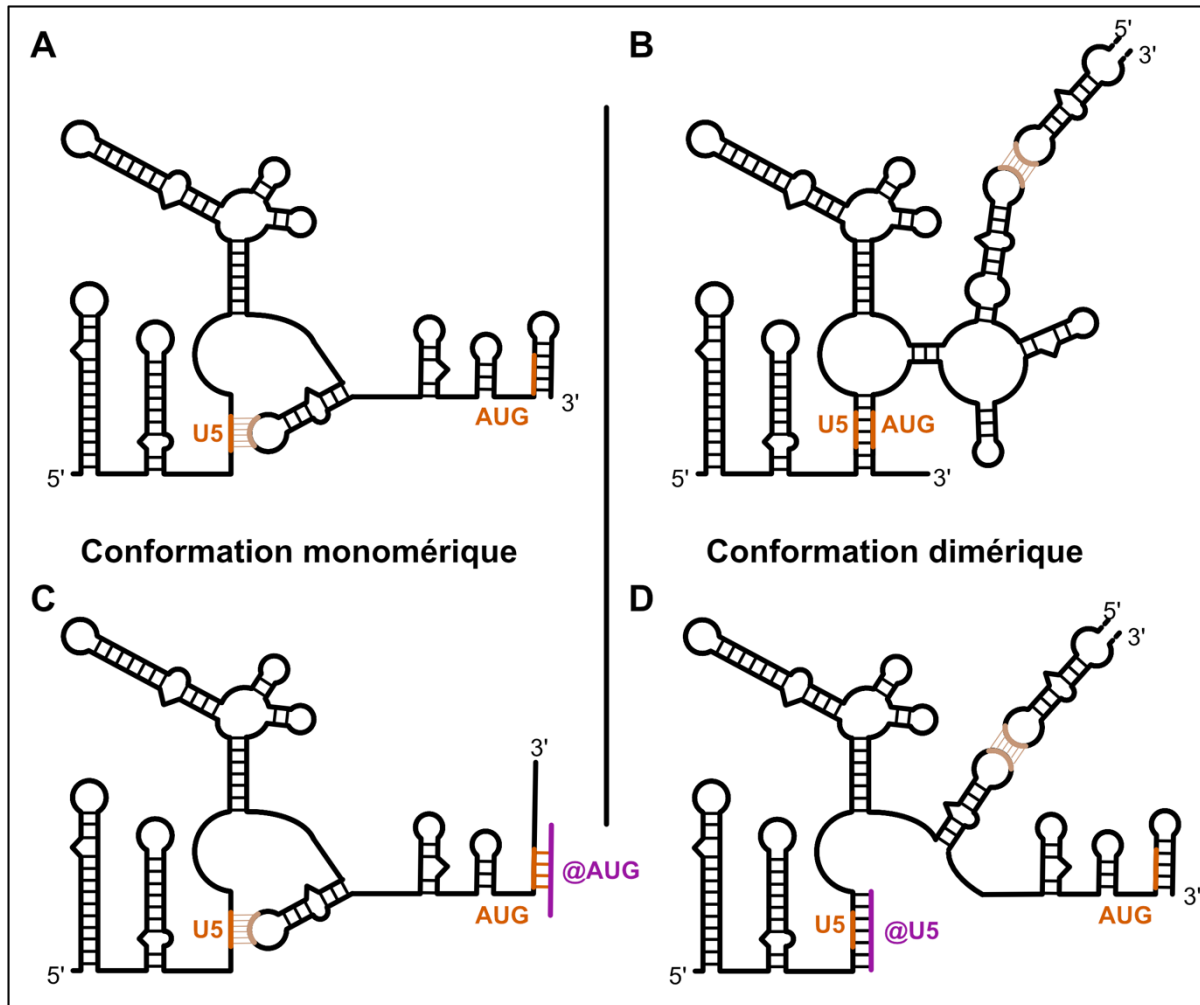
La mise au point l'étape de ligation du second adaptateur, nécessaire au séquençage, a marqué la dernière étape d'optimisation du protocole de préparation de banque de séquences. Le protocole initial était similaire à celui utilisé pour la première ligation et a été amélioré en suivant différents protocoles de la littérature (286,306–308). Comme précédemment, la vérification de la ligation du second adaptateur a été réalisée par un contrôle sur Bioanalyser et à l'aide d'une PCR ciblant les séquences des adaptateurs ligués. Les résultats préliminaires ont montré un biais important de la ligation, favorisant les petites espèces d'ADNc présentes dans le milieu. Pour tenter de s'affranchir de ce biais, l'expérience a été réalisée à l'identique avec une étape préalable de séparation du mélange des 4 ARN par une séparation sur des billes magnétiques SPRISelect®. Les résultats de la séparation sont présentés en **Figure 47**. Premièrement, nous pouvons observer une nette séparation des tailles entre les petits ARN (**Figure 47A**) et les grands ARN (**Figure 47B**). Cette différence est également visible après la PCR, entre les pistes 2 et 3 (**Figure 47C**). Un léger décalage au niveau des petits ADNc est relativement visible sur l'électrophérogramme, entre les courbes des ADNc non ligués (rouges) et des ADNc ligués (bleus) (**Figure 47A et 47B**). Cette différence suggère la ligation de l'adaptateur en 3' des ADNc et est confirmée par la PCR qui a permis l'amplification de matériel.



**Figure 47** : Validation de la séparation des ARN en fonction de la taille et de la seconde ligation sur les ADNc. Les données du Bioanalyser sur les mélanges de (A) petits et (B) grands ADNc montrent les profils de la séparation par les billes SPRiSelect. Les mélanges d'ADNc avant (profils rouges) et après (profils bleus) ligation en 3' sont comparés pour chaque taille d'ADNc. Les flèches bleues et rouges indiquent les produits ligués en 3' ou non, respectivement. (C) Les amplifications PCR des mélanges de petits (piste 2) et grands (piste 3) ADNc ligués valident les étapes de séparation et de ligations. La piste 4 correspond à un contrôle négatif où le mélange d'ADNc a été remplacé par un volume équivalent de H<sub>2</sub>O MQ. Les flèches grises indiquent les tailles attendues pour les ADNdb issus des ARN 110, 295 et 400 nts.

## VI/ Approche 4 : Préparation de banque de séquence adaptée à la région 5' de l'ARNg du VIH-1 pour un séquençage Illumina®

L'ensemble des optimisations du protocole initial a été incorporé dans le protocole final présenté en Matériel & Méthode. Chaque étape a été contrôlée en parallèle de la création de la banque pour s'assurer de la qualité de la préparation tout au long du protocole. Lors de l'expérience de MOHCA-seq, les ARN ont été repliés dans 5 conditions différentes, comprenant [1] de H<sub>2</sub>O MQ, [2] un tampon favorisant la monomérisation de l'ARN, [3] un tampon favorisant la dimérisation de l'ARN, [4] un tampon favorisant la dimérisation de l'ARN en présence d'un oligonucléotide antisens s'hybridant sur la région U5 (nucléotides 102 à 116) et [5] un tampon favorisant la dimérisation de l'ARN en présence d'un oligonucléotide antisens s'hybridant sur la région AUG (nucléotides 334 à 348). Les conditions en présence d'oligonucléotides antisens ont été réalisées dans le but de perturber des interactions potentielles entre ces deux régions (**Figure 48**). En effet, l'interaction U5:AUG a été proposée par l'équipe de Summers qui suggère également une influence de ces régions dans le processus de dimérisation (211). De plus, l'interactions entre U5 et AUG a été retrouvée dans les modèles issus de nos données hSHAPE en conditions favorisant la dimérisation (72).

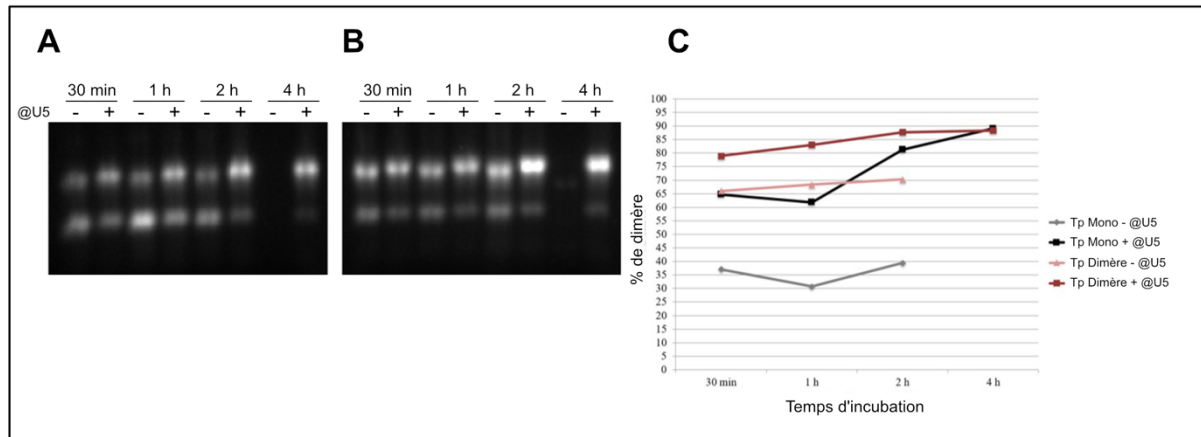


**Figure 48** : Schéma représentatif des structures attendues dans les conditions expérimentales testées en MOHCA-seq. Selon le modèle postulé par l'équipe Summers, (A) dans des conditions de faibles forces ioniques (tampon monomérique), l'ARN 1-604 devrait adopter une conformation majoritairement monomérique (U5:DIS). (B) Dans des conditions de forces ioniques supérieures (tampon dimérique), cette conformation devrait être dimérique (U5:AUG). L'utilisation d'oligonucléotides antisens complémentaires aux régions (C) @AUG et (D) @U5 devrait permettre de favoriser, respectivement, une conformation monomérique (U5:DIS) ou dimérique (U5:@U5) en tampon de forces ioniques élevées (tampon dimérique).

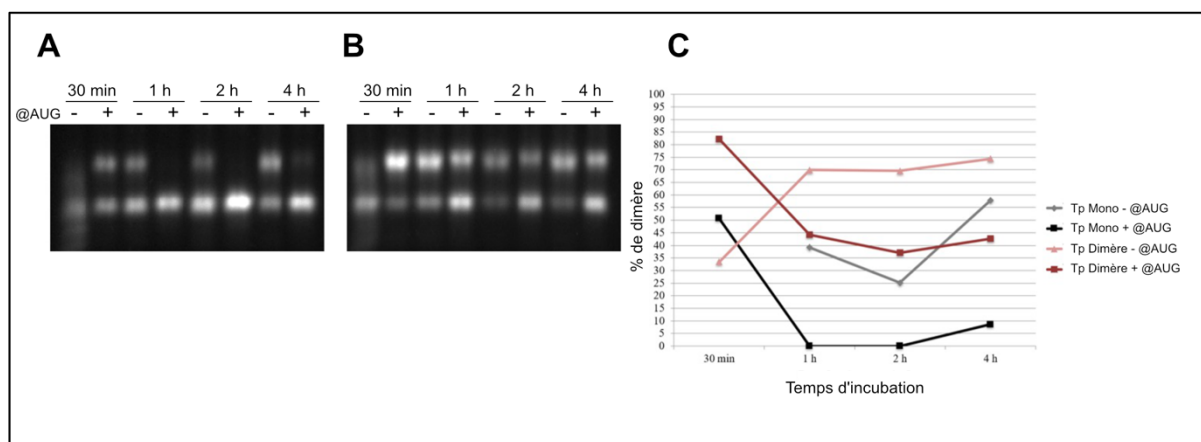
Ainsi, les conditions en présence d'oligonucléotides ont été mises au point au laboratoire avec Sarah Peterl, une étudiante en stage de licence (**Figures 49 et 50**). Les résultats de ses expériences suggèrent une influence des régions U5 et AUG dans le processus de dimérisation, comme proposé par l'équipe de Summers (211). En effet, une séquestration de la région U5 à l'aide d'un oligonucléotide antisens semble favoriser la dimérisation (**Figure 49**). En revanche, un oligonucléotide antisens ciblant la région AUG semble favoriser la monomérisation de l'ARN (**Figure 50**). Ainsi, une comparaison entre ces conditions en

## AXE 2 – RESULTATS

présence d'oligonucléotides antisens et les conditions où le tampon favorise soit la forme monomérique de l'ARN soit sa forme dimérique devrait permettre d'identifier le partenaire de chacune des deux régions.



**Figure 49** : Impact de l'oligonucléotide antisens U5 (@U5) sur la dimérisation. Les transcrits *in vitro* sont repliés dans (A) un tampon favorisant la monomérisation ou (B) la dimérisation, en présence (+) ou en absence (-) de l'oligonucléotide U5. L'ensemble des échantillons a été incubé à 37°C pendant 30 min, 1 h, 2 h ou 4 h. (C) Les courbes représentent la quantification relative des dimères d'ARN dans les différentes conditions.



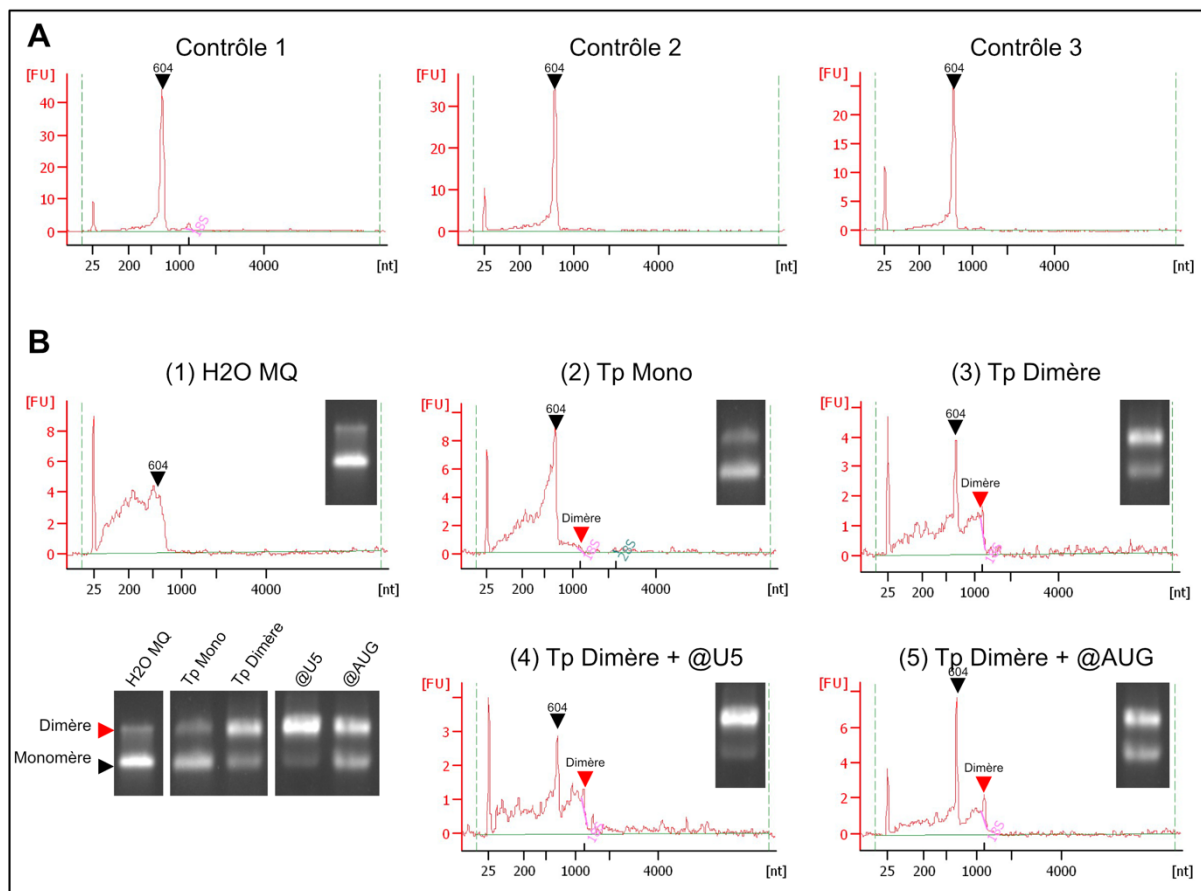
**Figure 50** : Impact de l'oligonucléotide antisens AUG (@AUG) sur la dimérisation. Les transcrits *in vitro* sont repliés dans (A) un tampon favorisant la monomérisation ou (B) la dimérisation, en présence (+) ou en absence (-) de l'oligonucléotide AUG. L'ensemble des échantillons a été incubé à 37°C pendant 30 min, 1 h, 2 h ou 4 h. (C) Les courbes représentent la quantification relative des dimères d'ARN dans les différentes conditions.

Afin de réaliser le protocole final de MOHCA-seq, dans un premier temps, j'ai vérifié la qualité de l'ARN ayant incorporé le 2'NH<sub>2</sub> après sa purification sur gel (**Figure 51A**). Les résultats obtenus montrent un ARN de 600 nucléotides jugé suffisamment pur pour poursuivre les expériences. Ainsi, l'ITCB-EDTA a été fixé aux groupements 2'NH<sub>2</sub> et du fer a été



## AXE 2 – RESULTATS

chélaté. Les différentes conditions de repliement de l'ARN ont été réalisées avant que l'ARN ne soit clivé par les radicaux hydroxyles (**Figure 51B**). Ainsi, l'ARN est majoritairement monomérique dans les conditions [1] et [2], et le taux d'ARN dimérique augmente progressivement entre les conditions [5], [3] et [4]. Pour le clivage, les résultats obtenus sur le Bioanalyser présentent des profils cohérents avec un clivage des ARN dans l'ensemble des conditions mais avec, comme attendu, une majorité des ARN qui semble ne pas avoir été clivé, sauf dans la condition de repliement dans H<sub>2</sub>O MQ.



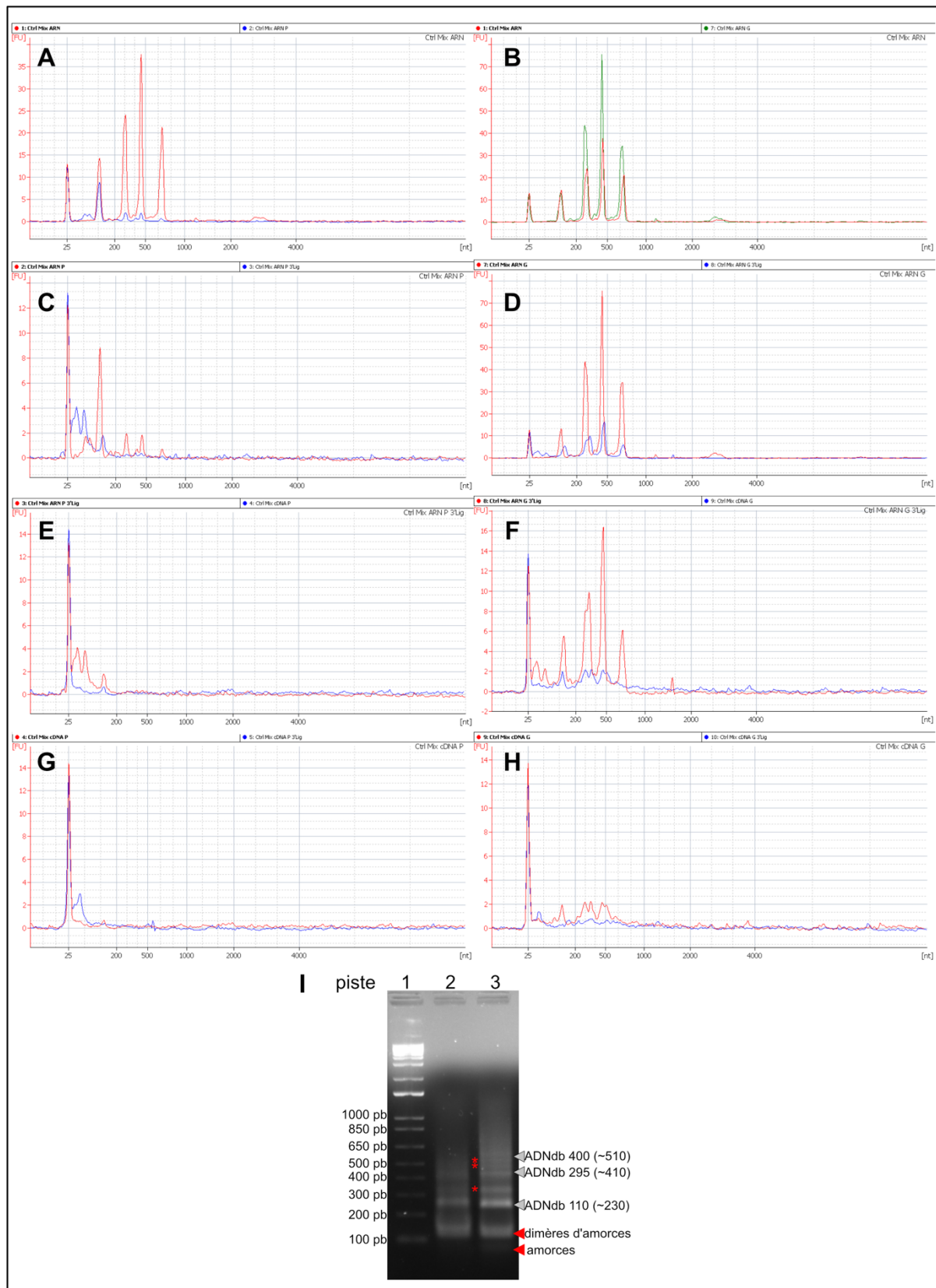
**Figure 51** : Vérification de la pureté de l'ARN modifié 2'NH<sub>2</sub> et des clivages aux radicaux hydroxyles lors du protocole de MOHCA-seq. (A) La pureté de l'ARN 2'NH<sub>2</sub> a été confirmée sur une puce RNA 6000 Nano au Bioanalyser à 3 reprises. (B) Les clivages aux radicaux hydroxyles ont été confirmés pour les 5 conditions d'intérêt sur une puce RNA 6000 Nano au Bioanalyser. Le taux de dimérisation des ARN modifiés dans chaque condition avant clivage ont été vérifiées sur un gel d'agarose 1%, TBM 0,5x.

Les différentes populations de fragments d'ARN ont ensuite été purifiées avant d'être déphosphorylées en 3'. A partir de ce moment, un contrôle externe a été ajouté à mes expériences pour vérifier que chaque étape de la préparation de la banque se déroule

## AXE 2 – RESULTATS

correctement. Ainsi, le mélange des 4 ARN utilisé pour la mise au point a été traité en parallèle des 5 conditions d'intérêt pour valider l'ensemble de la préparation de la banque. La séparation des populations de petits et grands fragments d'ARN a été réalisée avec les billes SPRISelect® (**Figure 52**). Les résultats obtenus pour notre échantillon contrôle contenant les 4 ARN indiquent que les grands fragments ont bien été exclus des mélanges comprenant les petits fragments d'ARN (**Figure 52A**), mais la réciproque n'est pas exacte.

## AXE 2 – RESULTATS



**Figure 52 :** Contrôles externes de chaque étape de la préparation des banques de séquences pour le MOHCA-seq. Premièrement, la séparation des (A) petits et (B) grands ARN est vérifiée. Puis, (C et D) l'étape de ligation de l'adaptateur 3' (TruSeq) sur les deux populations d'ARN est étudiée. (E et F) Les étapes de transcription inverse et (G et H) de ligation du second adaptateur (TruSeq) sont également contrôlées. La validation finale des

## AXE 2 – RESULTATS

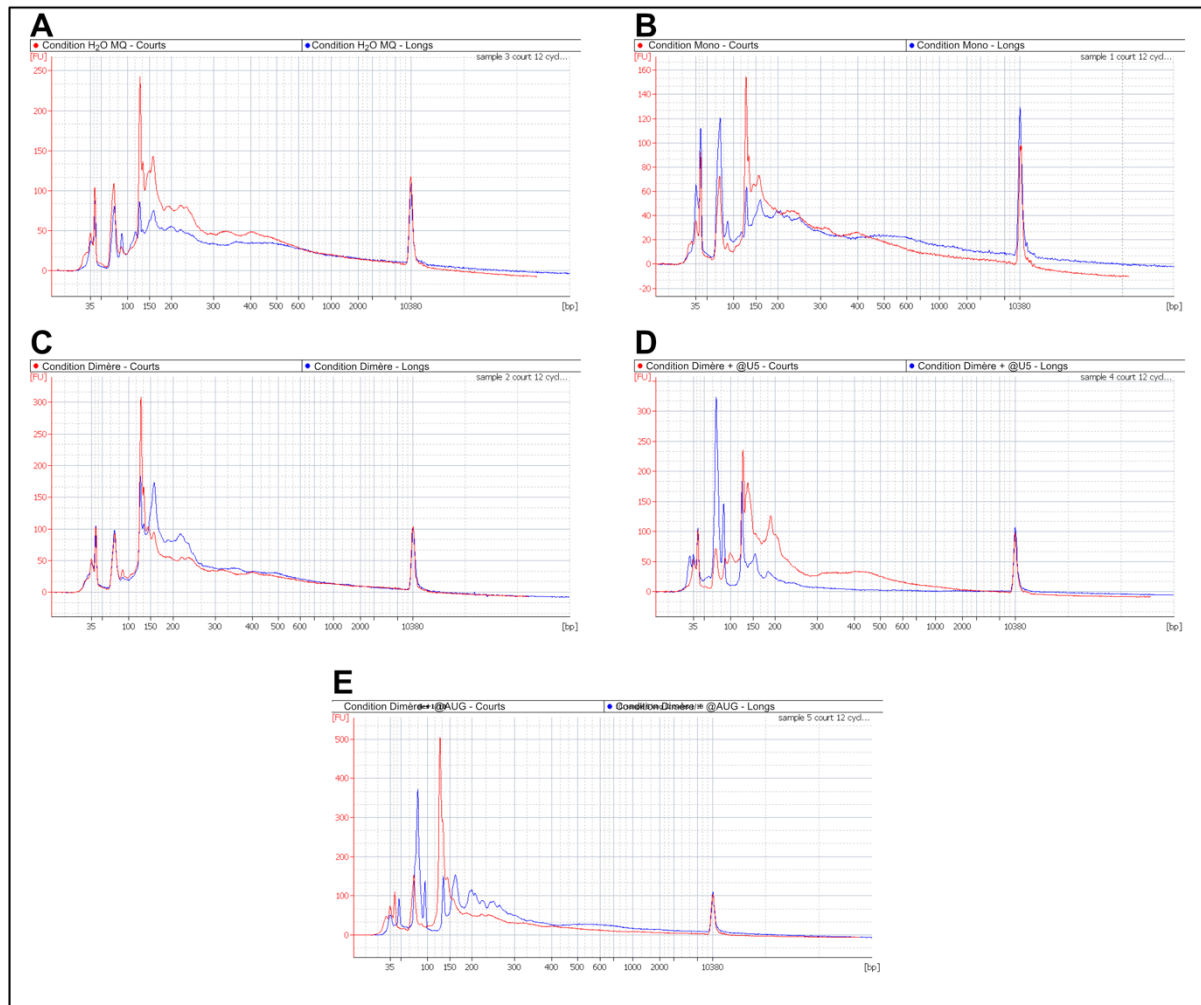
différentes étapes est réalisée par (I) amplification PCR. Les pistes 2 et 3 contiennent les ADNc issus des petits et des grands ARN, respectivement. Les flèches grises indiquent les amplicons issus des PCR. Les flèches rouges indiquent les bandes issues des amorces. Les astérisques rouges désignent les bandes contaminantes.

En effet, les petits ARN sont toujours présents dans la solution supposée contenir uniquement les grands ARN (**Figure 52B**). La préparation des différentes banques a tout de même été poursuivie en parallèle sur les deux lots de différente taille d'ARN et contrôlée à chaque étape. Cependant, la qualité des contrôles a rapidement décliné (**Figure 52C-H**) et il a été compliqué d'utiliser le contrôle externe pour juger de la préparation de la banque. En effet, les profils de Bioanalyser montrent un décalage entre les profils rouges et bleus, dû à la ligation du premier adaptateur (**Figure 52C et 52D**). En revanche, les autres résultats obtenus par le Bioanalyser permettent difficilement de conclure sur les événements de transcription inverse (**Figure 52E et 50F**) et n'apportent pas d'informations sur la seconde étape de ligation (**Figure 52G et 52H**). Ces résultats suggèrent que le matériel étudié a été perdu lors des différentes étapes des préparations. Cependant, l'amplification PCR du contrôle a permis de montrer que malgré la faible qualité des résultats obtenus sur Bioanalyser, la ligation des deux adaptateurs s'est bien déroulée, permettant d'amplifier des ADNc issus de fragments d'ARN allant jusqu'à 400 nucléotides (**Figure 52I**). La présence de bandes contaminantes (**Figure 52I** – astérisques rouges) pourrait correspondre à des amplicons n'ayant ligué qu'un adaptateur et étant amplifié de façon linéaire ou à des ARN étrangers ayant été ligués aux deux extrémités. Après avoir confirmé la ligation des deux adaptateurs, les différentes conditions, séparées en petits et grands fragments, ont été amplifiées par PCR et analysées sur le Bioanalyser (**Figure 53**). Cette étape de PCR permet d'amplifier la banque et d'ajouter les index et adaptateurs Illumina® nécessaires au séquençage. Les résultats montrent qu'une amplification a bien eu lieu dans l'ensemble des échantillons, mais qu'à l'inverse du contrôle externe, la séparation entre les petits et les grands fragments n'a pas été spécialement efficace. Les différentes banques ont donc été normalisées, rassemblées, déposées sur un gel NOVEX™ TBE-urée à 6% d'acrylamide et purifiées pour retirer les amorces et les dimères d'amorces PCR restant dans le milieu. Un dernier contrôle sur le Bioanalyser a été réalisé sur les banques avant et après purification. Les résultats mettent en évidence la disparition des amorces et des dimères d'amorces, initialement majoritaire (**Figure 54** - cadres) après la purification sur gel. Les échantillons ont donc été envoyés à la plateforme de séquençage de l'IBMP pour être séquencés sur un

## AXE 2 – RESULTATS

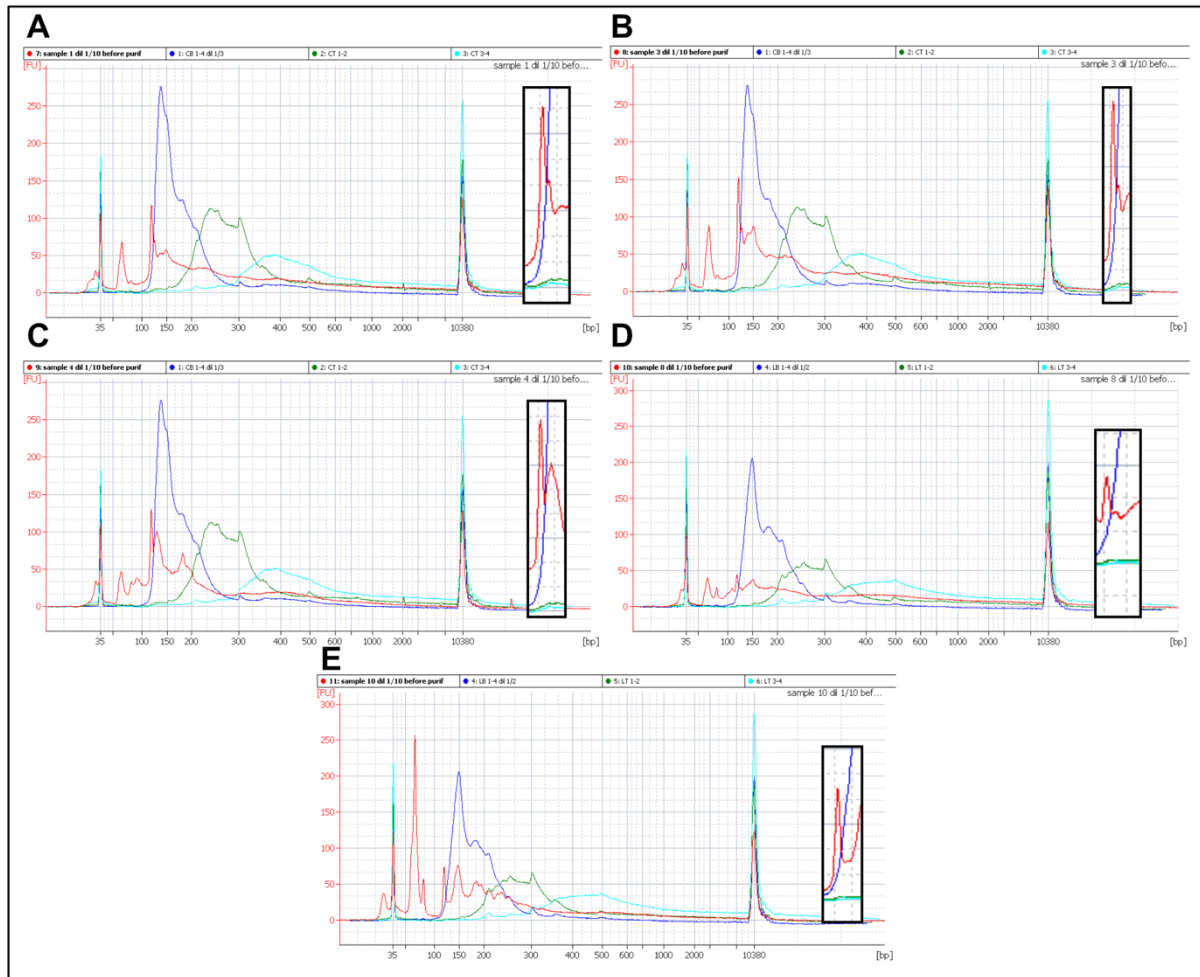
appareil MiSeq en mode paired-end de 75 nucléotides. Les ADNc double brins « longs » seront séquencés sur une piste différente de celle des ADNc double brins « courts » et « moyens » pour diminuer les risques de biais liés à l'amplification des courtes séquences.

Lors de l'écriture de ce manuscrit, nous n'avions pas encore obtenu les résultats du séquençage.



**Figure 53 :** Contrôle des amplifications PCR associées à chaque condition de MOHCA-seq. Les banques d'amplicons pour les conditions (A) H<sub>2</sub>O MQ, (B) Mono, (C) Dimère, (D) Dimère en présence d'oligonucléotide antisens U5 et (E) Dimère en présence d'oligonucléotide antisens AUG ont été contrôlées sur une puce HS DNA au Bioanalyser. Les profils rouges correspondent aux banques issues des fragments courts d'ARN et les profils bleus de celles issues des fragments longs d'ARN.

## AXE 2 – RESULTATS



**Figure 54** : Vérification de la qualité des banques de séquences après purification sur gel. Différentes dilutions ont été appliquées à chaque échantillons afin qu'ils puissent être lus sur une puce HS DNA au Bioanalyser. Les profils rouges correspondent aux banques de séquences avant la purification, contenant encore les amorces et les dimères d'amorces. Après la purification sur gel, les banques de séquences séparées en petits (profils bleus), moyens (profils verts) et grands (profils cyans) fragments ont été contrôlées pour s'assurer de leur pureté. Le cadre présent dans chaque analyse met en évidence le pic correspondant aux dimères d'amorces qui a été retiré lors de la purification.

## Discussion :

L'objectif de mon second axe de thèse s'est concentré sur l'identification des interactions à longue distance impliquées dans les repliements tridimensionnels de la région 5' terminale de l'ARNg du VIH-1. En effet, ces interactions sont responsables de l'organisation tridimensionnelle de l'ARN. Une identification expérimentale de régions interagissant ensemble et de celles proches dans l'espace permettrait donc d'améliorer les modèles structuraux existants. Ainsi, j'ai travaillé à la mise en place et à l'optimisation de l'approche du MOHCA-seq (260) (**Figure 33**). Cette mise en place s'est focalisée sur plusieurs points dont l'importance a été mise en évidence au cours des différentes tentatives successives. Ces points sont : le taux d'incorporation des nucléotides modifiés, la pureté des transcrits et la préparation des banques de séquençage.

Ainsi, j'ai commencé par m'assurer que les nucléotides modifiés 2'NH<sub>2</sub> étaient correctement incorporés dans les transcrits synthétisés *in vitro*. Cette étape a été réalisée par une méthode d'arrêt de la reverse transcription où j'ai mis en évidence la gêne stérique induite par la fixation de l'ITCB-EDTA à l'ARN au niveau de l'hydroxyle 2' des adénosines (**Figure 39**). Cependant, il était également possible que la fixation de la molécule ne se fasse pas sur le 2'NH<sub>2</sub> des nucléotides modifiés, mais sur le groupement amine situé sur le C6 des adénines, bien que cette position soit moins réactive. Cette hypothèse a été réfutée lors des expériences de clivage réalisées avec les ARNs transcrits à partir de différents ratios de nucléotides modifiés. En effet, les ARN n'ayant incorporé aucun nucléotides 2'NH<sub>2</sub> mais traités avec l'ITCB-EDTA, le fer et l'ascorbate de sodium n'ont pas subi de clivage suite à la production de radicaux hydroxyles (**Figure 40B** – rectangles bleus). Ces résultats indiquent bien que l'ITCB-EDTA se fixe seulement aux groupements 2'NH<sub>2</sub> des nucléotides modifiés et que cette molécule est nécessaire au clivage. De même, ces expériences montrent que le fer ne reste pas présent après les étapes de purification, que ce soit sur l'ARN ou en solution. Il ne devrait donc pas y avoir de clivage aspécifique dû à la présence d'ion non fixés.

Lors de l'optimisation de la purification des ARN, le choix de la purification sur gel de polyacrylamide a été motivé par la nécessité d'obtenir le produit le plus pur possible, en perdant le minimum de matériel. Dans cette optique, différentes méthodes d'élution ont été testées. La comparaison des rendements a montré que l'électro-élution était l'approche la

plus adaptée à nos expériences. Cependant, pour diminuer les risques liés à l'électro-élution, j'ai décidé de réaliser cette purification en amont de la fixation de l'ITCB-EDTA et de la chélation du fer. En effet, différents ions divalents sont connus pour leurs capacités à hydrolyser les acides nucléiques dépendamment du pH, de la température et du temps (309). Afin de diminuer les risques d'un clivage précoce de l'ARN, avant qu'il soit replié selon les conditions d'intérêt, il a été décidé de ne fixer l'ITCB-EDTA et le fer qu'après la purification et l'électro-élution. Ainsi, le courant électrique, la température et la présence d'EDTA dans le tampon de migration n'affectent pas les conditions expérimentales.

Ces différentes optimisations sont le résultat de deux tentatives de MOHCA-seq n'ayant pas permis l'obtention d'informations structurales sur la région 5' terminale de l'ARNg du VIH-1. Cependant, elles ont été nécessaires pour mettre en avant des points à améliorer au niveau de la méthodologie.

La troisième tentative de réalisation de l'approche de MOHCA-seq a été réalisée jusqu'à l'étape de clivage des ARN, la préparation des banques de séquences ayant été confiée à la plateforme EpiRNA-Seq de Nancy. Cependant, les données obtenues semblaient majoritairement provenir d'une contamination dont la source n'a pas été identifiée (**Figures 42 et 43**). Elles n'ont donc pas été analysées dans le but d'obtenir des informations structurales sur l'ARN, mais à nouveau, pour trouver des points à améliorer. Ainsi, l'incapacité du kit NEBNext® à préparer des fragments de plus de 150 nucléotides a été mise en évidence pour la seconde fois. Or, l'interaction à longue distance la plus longue identifiée à ce jour dans la littérature dans la région 5' terminale de l'ARNg du VIH-1 fait ~370 nucléotides (206). Il a donc été décidé de ne pas utiliser un kit commercial et de réaliser la préparation des banques de séquences à l'aide de protocoles adaptés à la taille de nos fragments d'intérêt. De plus, l'analyse initiale des lectures de séquences montrait également un biais lié à la perte d'information pour les fragments d'environ 70 nucléotides (**Figures 38 et 43**). La source du biais a été identifiée dans la méthode utilisée pour l'alignement des séquences à la suite du « trimming » ou retrait des séquences d'adaptateurs aux extrémités des lectures. En effet, deux méthodes existent pour l'alignement avec le programme Bowtie2, « local » ou « end-to-end », qui assignent des scores en fonction de la qualité des alignements avec la séquence fournie (310). La méthode « end-to-end » utilisée par défaut vise à aligner l'ensemble des nucléotides lus, sans omission aucune. Dans le cas des fragments approchant de 70 nucléotides, la séquence de l'adaptateur n'est pas retirée car elle est trop courte pour être



reconnue. Cependant, elle est suffisamment longue pour venir perturber l'alignement avec la séquence source, faisant drastiquement diminuer le score d'alignement. Ainsi, l'ensemble des séquences autour de 70 nucléotides est éliminé de l'analyse finale. Pour contourner ce biais, il a été décidé de réaliser un alignement de séquences en mode « end-to-end », classiquement utilisé, puis de reprendre l'ensemble des alignements dont le score était trop bas pour les réaligner avec le mode « local ». Ainsi, la fusion des deux jeux de données a permis d'obtenir les informations fournies par l'un ou l'autre des deux modes d'alignement et une partie des informations manquantes a été retrouvée (**Figure 43**).

Les dernières étapes d'optimisation ont été nécessaires à l'obtention d'un protocole validé pour la préparation des banques de séquences aux tailles nous intéressant. En effet, chaque étape de la préparation a pu être contrôlée et vérifiée (**Figure 52**), à l'inverse de celle du kit. Les différents contrôles nous ont permis de mettre en évidence, entre autres, un biais par rapport à la taille des fragments d'ARN et des ADNc ligués, avec une préférence pour les acides nucléiques de petites tailles (**Figure 44B et 44C**). Ce biais, bien que connu dans la littérature et les approches de préparation de banque, reste également important à prendre en compte dans l'analyse des résultats finaux et leur interprétation. En effet, dans le cas d'un lissage des données à l'aide de filtres supprimant le bruit de fond, il sera important de considérer ce biais pour ne pas supprimer des informations par rapport aux fragments de grandes tailles. En outre, les différents contrôles réalisés au cours de la préparation de la banque ont montré qu'ils n'étaient pas toujours représentatifs des processus enzymatiques et chimiques se déroulant à chaque étape (**Figure 52**). En effet, il n'a pas été possible de visualiser correctement les effets des étapes de transcription inverse et de ligation en 3' des ADNc. Pourtant, l'amplification PCR des différents fragments indique que ces étapes ont été réalisées, sans quoi l'amplification n'aurait pas pu avoir lieu.

Finalement, lors de ma thèse, j'ai pu achever la mise en place d'une approche de cartographie chimique multidimensionnelle en solution et l'appliquer à l'étude de la région 5' terminale de l'ARNg du VIH-1 dans différentes conditions. Parmi ces conditions, deux d'entre elles utilisent des oligonucléotides antisens ciblant les régions U5 ou AUG. L'idée sous-jacente à l'utilisation de ces deux conditions étant de perturber les interactions réalisées par les régions U5 et AUG et d'évaluer l'impact de la fixation des oligonucléotides sur la dimérisation sans pour autant modifier la séquence nucléotidique de l'ARN (**Figure 48**). En

effet, selon les travaux de l'équipe de Summers (211,214,311,312), les régions U5 et AUG interagissent ensemble lorsque l'ARN est sous forme dimérique. En revanche, en conformation monomérique, la région U5 interagirait préférentiellement avec le DIS. Nos travaux sur la capacité de l'ARN à dimériser ou non en présence des oligonucléotides antisens U5 et AUG concordent avec cette hypothèse puisque le blocage de la région U5 amplifie le taux de dimérisation de l'ARNg (**Figure 49**) et que celui de la région AUG augmente le taux de monomère dans la solution (**Figure 50**). Cependant, ces résultats confirment uniquement l'impact des deux régions sur la dimérisation de l'ARN, et non les partenaires d'interactions impliqués dans cette modulation. Ainsi, ces conditions en présence d'oligonucléotides antisens ont été ajoutées aux conditions initiales pour identifier, à terme, les partenaires de chacune de ces régions.

Les 5 conditions d'intérêt ont donc été transmises à la plateforme de l'Institut de Biologie Moléculaire des Plantes (IBMP) à Strasbourg pour être séquencées. Cependant, le séquençage d'échantillons contenant différentes tailles peut également introduire un biais au niveau de l'amplification. En effet, les petits amplicons sont plus facilement et fréquemment amplifiés que les grands amplicons. Pour minimiser l'impact de la taille des amplicons, les échantillons ont été séparés en 2 pistes de séquençage en fonction de leurs tailles (**Figures 53 et 54**). Une piste de séquençage est chargée avec les petits et moyens fragments tandis qu'une autre sera chargée avec les grands fragments. Ainsi, la préférence pour les séquences courtes devrait moins impacter le séquençage des grands amplicons. De même, nous avons choisi de séquencer nos échantillons sur un appareil MiSeq et non pas un HiSeq. En effet, les séquençages HiSeq amplifient préférentiellement les tailles de fragments bien précises et moins les populations de taille hétérogène qui nous intéressent dans notre étude.

Pour autant, les résultats des séquençages n'ont pas encore été obtenus au moment où cette thèse est rédigée. Il n'a donc pas été possible d'en réaliser l'analyse, l'interprétation ou la discussion critique. Cependant, les contrôles internes qui permettront d'estimer la qualité des résultats peuvent déjà être discutés. En effet, plusieurs conditions ont déjà été testées, et pour chacune d'entre elles, le taux de dimérisation a été quantifié (**Figures 51B**). Ces résultats montrent une diversité de taux de conformères entre chaque condition. Il est donc attendu que les résultats varient également et que plusieurs profils de cartes bidimensionnelles soient obtenus, à l'inverse des banques précédentes (**Figure 43**). De plus, certaines interactions à longue distance ont été rapportées dans la littérature (205,206) et

## AXE 2 – DISCUSSION

pourraient servir de contrôles pour valider ou non l'approche. Finalement, la comparaison entre les conditions en présence d'oligonucléotides antisens (U5, AUG ou autre) et en absence de ces derniers pourrait montrer une perte d'interaction au niveau des régions ciblées, à moins que ce ne soit des interactions non canoniques.



# Discussion générale, perspectives et conclusion

Lors de ma thèse, j'ai optimisé la technique de cartographie chimique en solution hSHAPE pour l'étude de complexe ribonucléoprotéiques. Ce travail m'a permis par la suite d'étudier l'évolution de la structure de la région 5' terminale de l'ARNg du VIH-1 complexée aux protéines virales contenant le domaine NC qui sont présentes au cours de l'étape de maturation du virion (72). J'ai également mis en place une méthode de cartographie qui permet d'identifier des interactions à longue distance dans un ARN de 600 nucléotides, le MOHCA-seq.

La première partie de mon travail a été axée sur la formation et l'étude des complexes entre la région 5' terminale de l'ARNg du VIH-1 et les protéines virales contenant le domaine de nucléocapside (NC). La méthode, déjà en application au laboratoire, a dû être modifiée afin de convenir aux complexes ribonucléoprotéiques étudiés. En effet, les expériences initiales réalisées ont montré que ces complexes ne semblaient pas se former correctement dans toutes mes conditions et qu'une optimisation était nécessaire pour la protéine Pr55<sup>Gag</sup>. Cette étape de mise au point réalisée, j'ai pu établir un protocole permettant l'étude *in vitro* de la structure adoptée par la région 5' de l'ARNg dans plusieurs conditions. En effet, l'utilisation d'une approche de hSHAPE a permis d'obtenir des mesures quantitatives de la flexibilité des nucléotides lorsque l'ARNg était seul ou bien complexé avec différentes protéines issues de la maturation protéolytique. Ces expériences ont permis de mettre en évidence la capacité de chacune de ces protéines à agir sur la structure de l'ARN, ainsi que l'implication différentielle des domaines en doigts de zinc en fonction du contexte protéique. Les différentes activités chaperon de l'ARN des protéines Pr55<sup>Gag</sup>, NCp15, NCp9 et NCp7 montrent des effets distincts et croissants de chacune de ces protéines sur l'ARN. En effet, les protéines Pr55<sup>Gag</sup> et NCp15 n'ont que très peu d'effets de remodelage de la structure de l'ARNg. En revanche, les protéines NCp9 et NCp7 semblent induire des réarrangements conséquents par leur fixation à l'ARNg. Ces données peuvent être extrapolées et suggérer que la structure de la région 5' terminale de l'ARNg du VIH-1 est progressivement réarrangée dans les particules virales au cours de maturation. De plus, ces réarrangements semblent modifier les interactions à longue

## DISCUSSION GENERALE, PERSPECTIVES ET CONCLUSIONS

distance. En effet, les interactions U5:AUG et CU:GA qui sont présentes dans les modèles structuraux construits avec les données issues des ARN en complexe avec Pr55<sup>Gag</sup> et NCp15 traités à la protéinase K sont perdues pour les ARN en complexe avec NCp9 et NCp7 dans les mêmes conditions. En outre, les expériences avec l'AT-2 ont également permis de réaffirmer l'évolution de l'implication et des fonctions des domaines en doigts de zinc des protéines virales contenant le domaine NC (236,240,241). En effet, le traitement par l'AT-2 ne semble pas affecter le maintien de la fixation des protéines Pr55<sup>Gag</sup> et NCp9 alors qu'elle induit des déstabilisations marquées des complexes ARN-NCp15 et ARN-NCp7. Ces observations peuvent être interprétées de plusieurs façons. Premièrement, il est possible que la structure des protéines Pr55<sup>Gag</sup>, NCp15, NCp9 et NCp7 permette une accessibilité différentielle de l'AT-2 aux domaines en doigts de zinc lorsqu'ils sont en complexe avec l'ARN. Cette différence structurale pourrait expliquer la variation d'efficacité de l'AT-2 en fonction des protéines ciblées. Deuxièmement, les domaines en doigts de zinc ne sont pas les seules régions capables d'interagir avec les acides nucléiques. Ainsi, dans le cas du VIH-1, une région flexible et basique, pouvant lier les acides nucléiques de manière aspécifique, est localisé entre les deux doigts de zincs du domaine de NC (313,314). De même, le domaine MA présent dans Pr55<sup>Gag</sup> est également connu pour interagir avec les acides nucléiques (315). Il est donc possible que d'autres régions soient capables de conserver l'interaction entre l'ARN et les protéines, bien que la spécificité et l'affinité de cette interaction puissent être impactées. En effet, les domaines en doigts de zinc ont été montrés comme nécessaires à la fixation spécifique des protéines contenant le domaine NC (316). En revanche, le maintien de cette fixation peut être dépendant d'autres domaines protéiques. Finalement, il ne peut pas être exclu que la fixation de l'AT-2 ait un impact sur la structure de certaines des protéines, expliquant la différence d'effet de l'AT-2 selon les complexes ribonucléoprotéiques étudiés. L'AT-2 agit en modifiant les cystéines des domaines en doigts de zinc par une liaison covalente. Cette liaison empêche l'organisation du domaine autour d'un ion Zn<sup>2+</sup> et est donc supposée inactiver la fixation des protéines aux acides nucléiques. Cependant, il est possible que la liaison de l'AT-2 induise également une déstructuration locale ou globale de certaines protéines, impactant de ce fait leur capacité de fixation aux acides nucléiques.

Cependant, bien que ces expériences nous aient permis d'obtenir une vision globale des effets induits par la fixation des différentes protéines, des améliorations sont possibles.

## DISCUSSION GENERALE, PERSPECTIVES ET CONCLUSIONS

En effet, mimer l'étape de maturation *in vitro* pour mettre en évidence les réarrangements complexes se déroulant dans la particule virale pourrait permettre une meilleure compréhension des mécanismes régulant l'infectivité virale. En effet, nos conditions expérimentales ont isolé l'effet de chaque protéine sur l'ARN, sans prendre en compte les modifications potentielles induites par les protéines précédentes de la cascade ou l'effet de la présence des autres domaines clivés. Pour prendre en compte ces paramètres, il est imaginable de réaliser les expériences de formation de complexes entre l'ARN et Pr55<sup>Gag</sup>, puis d'ajouter la protéase virale (PR) pour induire la maturation de la protéine Pr55<sup>Gag</sup>. En effet, PR est déjà connue pour être active *in vitro* (317). Ainsi, les effets de la fixation de chaque protéine contenant le domaine de NC pourraient être évalués. De plus, une cinétique de la maturation permettrait de mettre en évidence les effets de chaque protéine, puisque les clivages sont successifs et dépendants du temps (64,318). La protéolyse pourrait également être contrôlée en parallèle par western blot, ce qui permettrait de suivre l'évolution de la maturation. Une approche plus complexe mais n'utilisant pas la protéase virale est également envisageable. Ainsi, le suivi d'une partie de l'évolution de la maturation serait réalisé par l'ajout successif des différentes protéines contenant le domaine de NC avec le retrait des protéines précédentes par un traitement à la protéinase K. Cependant, la mise au point de cette approche risque d'être trop lourde et peu reproductible du fait de la nécessité d'éliminer la protéinase K après chaque étape de maturation. A l'inverse de l'approche *in vitro*, il peut être intéressant de réaliser les expériences de cartographie chimique directement sur les particules virales (219), en présence de partenaires supplémentaires, et avec des mutants de la protéines Pr55<sup>Gag</sup> permettant de bloquer la maturation à certaines étapes spécifiques (64).

Outre les améliorations possibles de l'approche mimant l'étape de maturation virale, le protocole de hSHAPE que j'ai mis au point pour l'étude des complexes pourrait également être appliqué à l'étude de l'impact de la formation du complexe ternaire ARNg – ARNt<sup>Lys,3</sup> – Reverse transcriptase virale (RT) sur la structure de l'ARNg. Ce type d'expérience a déjà été rapporté dans la littérature par le passé, avec par exemple, les 311 premiers nucléotides étudiés par cartographie enzymatique en présence de l'ARNt<sup>Lys,3</sup> et la NCp7 (319). Plus récemment, les 350 premiers nucléotides de l'ARNg couplés à l'ARNt<sup>Lys,3</sup> et à la NCp7 ou bien un mutant incapable de dimériser couplé à l'intégrase et NCp7 ont été étudiés par FRET et XL-SHAPE, respectivement (244,245). De plus, l'ARN pourrait être structuré en présence de différents oligonucléotides antisens afin de perturber les interactions dans l'ARN (206) comme

## DISCUSSION GENERALE, PERSPECTIVES ET CONCLUSIONS

cela a été effectué pour certaines expériences de MOHCA-seq lors de ma thèse. Ces données expérimentales permettraient d'obtenir des informations supplémentaires sur les interactions impliquées dans la structure de l'ARN étudié, afin de valider et de préciser certains des modèles actuellement rapportés dans la littérature.

Enfin, l'analyse des données obtenues par hSHAPE lors de ma thèse aurait pu être légèrement améliorée. En effet, comme toutes données issues de mesures, les réactivités de hSHAPE présentent un bruit de fond qui peut biaiser, au moins partiellement, les données récoltées et donc les conclusions tirées de leur interprétation. Afin de diminuer l'impact de ce bruit de fond et permettre l'extraction de données de meilleure qualité, des outils mathématiques tels que le filtre Kalman (320) peuvent être appliqués aux jeux de données SHAPE (321). Ce filtre a d'ailleurs prouvé son efficacité sur le traitement de données biologiques (322). Ainsi, j'ai tenté avec l'aide d'Erwan Quignon, un autre doctorant du laboratoire, de l'implémenter pour l'analyse de nos répliques. Malheureusement, nous n'avons pas réussi à automatiser correctement l'application du filtre à l'ensemble de nos jeux de données et l'analyse n'a pas été poursuivie.

La seconde partie de mon travail de thèse s'est concentrée sur l'optimisation de l'approche de MOHCA-seq et en particulier des étapes de préparation des banques de séquences. En effet, les protocoles explorés initialement, commerciaux ou non, se sont révélés inadaptés à l'application du MOHCA-seq sur un ARN de 600 nucléotides et ont dû être optimisés. Lors de ces optimisations, de nombreux biais ont été mis en évidence et même si une partie a pu être éliminée, il ne nous a pas été possible, par manque de temps, de les corriger tous autant que souhaité. Une des difficultés dans le traitement de ces biais vient du fait que certains n'ont été identifiés qu'à la suite du séquençage de nos banques. C'est le cas, par exemple, du biais de la méthode d'alignement des lectures. En revanche, certains biais avaient été anticipés, comme celui favorisant la ligation des petits fragments au dépend des grands. Afin de réduire ce biais, j'ai séparé les fragments d'ARN en fonction de leur taille (petits et grands), laissant ainsi une possibilité aux grands fragments d'être également ligés aux adaptateurs pour le séquençage. Cependant, la séparation sur billes n'étant pas totale, nous avons pu observer la présence de petits fragments d'ARN dans la partie supposée ne contenir que des grands fragments. Par conséquent, nous avons une sur-représentation des ADNdb de petite taille dans nos échantillons. A moins d'améliorer la séparation sur billes, il pourrait être



## DISCUSSION GENERALE, PERSPECTIVES ET CONCLUSIONS

plus intéressant de réaliser la séparation par purification sur gel, ce qui permettrait de s'assurer de liguer les petits et les grands fragments séparément même si la perte de matériel au niveau de cette purification pourrait également introduire un biais. Quoiqu'il en soit, les résultats de l'analyse au Bioanalyser montrent tout de même que des grands fragments ont été obtenus par notre méthode. Ainsi, il devrait être possible d'obtenir des informations qualitatives sur les interactions à longue distance. Un autre biais à prendre en compte est la tendance de certaines approches de séquençage à favoriser les séquences courtes. Afin d'augmenter la probabilité d'obtenir des lectures correspondant à des séquences longues, le séquençage a été réalisé sur un appareil MiSeq et sur deux pistes différentes, l'une contenant les séquences courtes et l'autre les séquences longues. Cependant, nous sommes conscients que ces deux derniers biais peuvent causer un large déséquilibre au niveau des lectures de séquences obtenues à la fin du séquençage. Il sera donc important de traiter le bruit de fond présent dans les résultats du séquençage de façon adaptée aux tailles des lectures filtrées. Dans le cas contraire, les données produites ne seront pas représentatives des interactions à longues distances impliquées dans la structure étudiée.

Le séquençage à haut débit des banques que j'ai préparées devrait nous permettre d'obtenir des informations sur les repliements structuraux mis en place dans la région 5' terminale de l'ARNg du VIH-1. Ces données structurales seront ensuite éprouvées sur les modèles structuraux existants, en particulier sur ceux présents au laboratoire et obtenus par hSHAPE dans des conditions similaires à celles utilisées pour le MOHCA-seq. Les différents repliements identifiés devront par la suite être validés. Pour ce faire, les interactions à longue distance pourront être testées par mutagenèse dirigée. Ainsi, un des partenaires de l'interaction sera muté pour perturber l'interaction dans un premier temps, puis le second partenaire sera muté à son tour pour rétablir l'interaction. Ces mutants perturbateurs et compensatoires pourront également être étudiés en hSHAPE pour évaluer l'impact de ces mutations sur la structure globale de l'ARN (323). En parallèle de la mutagenèse, une approche basée sur l'utilisation d'oligonucléotides antisens pourra être mise au point. En effet, quelques études préliminaires ont été réalisées lors de ma thèse. Le principe était d'utiliser de courts oligonucléotides pour perturber des interactions WC-WC et valider le partenaire de l'interaction en étudiant les réactivités SHAPE associées. Ainsi, différentes approches permettraient une vérification des données obtenues par le MOHCA-seq.

## DISCUSSION GENERALE, PERSPECTIVES ET CONCLUSIONS

Finalement, les deux approches mises au point lors de ma thèse pourraient également être combinées. En effet, le MOHCA-seq n'est pas limité à l'étude des ARN seuls. Certains ARN ont été analysés en présence de cofacteurs (260). Il est donc envisageable que l'approche soit étendue à l'étude de complexes ribonucléoprotéiques, ou au moins à des ARN déprotéinisés, comme ceux étudiés dans le cas des réarrangements structuraux induits par les protéines contenant le domaine de NC (72).

Pour conclure, les résultats obtenus au cours de ma thèse ont permis de mieux comprendre l'impact des protéines virales impliquées dans la maturation sur les réarrangements structuraux de la région 5' terminale de l'ARNg du VIH-1. Les données obtenues pour l'ARN seul, combinées aux informations qui pourraient être obtenus par le MOHCA-seq devraient permettre de modéliser le premier modèle 3D de cette région soutenu par des données expérimentales. Les informations obtenues permettraient ainsi de clarifier et peut être d'expliquer une partie des divergences structurales observées dans la littérature. En effet, bien que le but premier du MOHCA-seq soit l'identification des régions proches dans l'espace et donc des interactions à longue distance, l'approche devrait également permettre d'obtenir des informations sur les interactions à courte distance, comme celle trouvées dans les tiges-boucles articulant la structure secondaire. Une première étape sera donc de filtrer l'ensemble des modèles correspondant aux données secondaires extraites. Par la suite, il faudra évaluer la capacité de ces modèles à former les interactions à longue distance identifiées. Pour ce faire, des modèles tridimensionnels peuvent être produits *in silico* via le logiciel FARFAR2 (296,297) ou par Intelligence Artificielle avec DeepFoldRNA (324) et comparés avec les données obtenues par le MOHCA-seq. Ainsi, les modèles 3D n'étant pas capables de former les interactions identifiées pourront être éliminés. Finalement, il est espéré que les données produites permettent d'affiner la perception de la structure de la région 5' terminale de l'ARNg du VIH-1, pour améliorer la compréhension des mécanismes moléculaires impliquant cette structure.





# Bibliographie

1. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science*. 20 mai 1983;220(4599):868-71.
2. Coffin JM. Structure and Classification of Retroviruses. In: Levy JA, éditeur. *The Retroviridae* [Internet]. Boston, MA: Springer US; 1992 [cité 15 oct 2022]. p. 19-49. Disponible sur: [http://link.springer.com/10.1007/978-1-4615-3372-6\\_2](http://link.springer.com/10.1007/978-1-4615-3372-6_2)
3. Archer J, Robertson DL. Understanding the diversification of HIV-1 groups M and O. *AIDS*. 20 août 2007;21(13):1693-700.
4. Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. The Challenge of HIV-1 Subtype Diversity. *N Engl J Med*. 10 avr 2008;358(15):1590-602.
5. Briggs JAG, Simon MN, Gross I, Kräusslich HG, Fuller SD, Vogt VM, et al. The stoichiometry of Gag protein in HIV-1. *Nat Struct Mol Biol*. juill 2004;11(7):672-5.
6. Jouvenet N, Simon SM, Bieniasz PD. Visualizing HIV-1 assembly. *J Mol Biol*. 22 juill 2011;410(4):501-11.
7. Bukrinskaya AG. HIV-1 assembly and maturation. *Arch Virol*. 1 juin 2004;149(6):1067-82.
8. Frankel AD, Young JAT. HIV-1: Fifteen Proteins and an RNA. *Annu Rev Biochem*. juin 1998;67(1):1-25.
9. Welker R, Hohenberg H, Tessmer U, Huckhagel C, Kräusslich HG. Biochemical and Structural Analysis of Isolated Mature Cores of Human Immunodeficiency Virus Type 1. *J Virol*. févr 2000;74(3):1168-77.
10. Mallery DL, Kleinpeter AB, Renner N, Faysal KMR, Novikova M, Kiss L, et al. A stable immature lattice packages IP<sub>6</sub> for HIV capsid maturation. *Sci Adv*. 12 mars 2021;7(11):eabe4716.
11. Wain-Hobson S, Sonigo P, Danos O, Cole S, Alizon M. Nucleotide sequence of the AIDS virus, LAV. *Cell*. janv 1985;40(1):9-17.
12. Psallidopoulos MC, Schnittman SM, Thompson LM, Baseler M, Fauci AS, Lane HC, et al. Integrated proviral human immunodeficiency virus type 1 is present in CD4+ peripheral blood lymphocytes in healthy seropositive individuals. *J Virol*. nov 1989;63(11):4626-31.
13. Rice AP, Mathews MB. Transcriptional but not translational regulation of HIV-1 by the tat gene product. *Nature*. avr 1988;332(6164):551-3.
14. Clever JL, Miranda, D, Parslow TG. RNA Structure and Packaging Signals in the 5' Leader Region of the Human Immunodeficiency Virus Type 1 Genome. *J Virol*. 1 déc 2002;76(23):12381-7.
15. Cassan E, Arigon-Chifolleau AM, Mesnard JM, Gross A, Gascuel O. Concomitant emergence of the antisense protein gene of HIV-1 and of the pandemic. *Proc Natl Acad Sci*. 11 oct 2016;113(41):11537-42.
16. Pavesi A, Romerio F. Extending the Coding Potential of Viral Genomes with Overlapping Antisense ORFs: A Case for the De Novo Creation of the Gene Encoding the Antisense Protein ASP of HIV-1. *Viruses*. 14 janv 2022;14(1):146.
17. Swanson CM, Malim MH. SnapShot: HIV-1 Proteins. *Cell*. mai 2008;133(4):742-742.e1.
18. McCune JM, Rabin LB, Feinberg MB, Lieberman M, Kosek JC, Reyes GR, et al.

## BIBLIOGRAPHIE

- Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. *Cell*. avr 1988;53(1):55-67.
19. Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, et al. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res*. 1 nov 2012;40(20):10345-55.
  20. Nguyen Quang N, Goudey S, Ségéral E, Mohammad A, Lemoine S, Blugeon C, et al. Dynamic nanopore long-read sequencing analysis of HIV-1 splicing events during the early steps of infection. *Retrovirology*. déc 2020;17(1):25.
  21. Stupfler B, Verriez C, Gallois-Montbrun S, Marquet R, Paillart JC. Degradation-Independent Inhibition of APOBEC3G by the HIV-1 Vif Protein. *Viruses*. 3 avr 2021;13(4):617.
  22. Sato K, Gee P, Koyanagi Y. Vpu and BST2: Still Not There Yet? *Front Microbiol* [Internet]. 2012 [cité 17 oct 2022];3. Disponible sur: <http://journal.frontiersin.org/article/10.3389/fmicb.2012.00131/abstract>
  23. Joseph SB, Arrildt KT, Sturdevant CB, Swanstrom R. HIV-1 target cells in the CNS. *J Neurovirol*. juin 2015;21(3):276-89.
  24. Falkenhagen A, Joshi S. HIV Entry and Its Inhibition by Bifunctional Antiviral Proteins. *Mol Ther - Nucleic Acids*. déc 2018;13:347-64.
  25. Shen Q, Wu C, Freniere C, Tripler TN, Xiong Y. Nuclear Import of HIV-1. *Viruses*. 8 nov 2021;13(11):2242.
  26. Burdick RC, Li C, Munshi M, Rawson JMO, Nagashima K, Hu WS, et al. HIV-1 uncoats in the nucleus near sites of integration. *Proc Natl Acad Sci*. 10 mars 2020;117(10):5486-93.
  27. Li C, Burdick RC, Nagashima K, Hu WS, Pathak VK. HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. *Proc Natl Acad Sci*. 9 mars 2021;118(10):e2019467118.
  28. Freed EO. HIV-1 assembly, release and maturation. *Nat Rev Microbiol*. août 2015;13(8):484-96.
  29. Shaik MM, Peng H, Lu J, Rits-Volloch S, Xu C, Liao M, et al. Structural basis of coreceptor recognition by HIV-1 envelope spike. *Nature*. janv 2019;565(7739):318-23.
  30. Ingram Z, Taylor M, Okland G, Martin R, Hulme AE. Characterization of HIV-1 uncoating in human microglial cell lines. *Virol J*. déc 2020;17(1):31.
  31. Suzuki Y, Craigie R. The road to chromatin — nuclear entry of retroviruses. *Nat Rev Microbiol*. mars 2007;5(3):187-96.
  32. Francis AC, Melikyan GB. Single HIV-1 Imaging Reveals Progression of Infection through CA-Dependent Steps of Docking at the Nuclear Pore, Uncoating, and Nuclear Transport. *Cell Host Microbe*. avr 2018;23(4):536-548.e6.
  33. Yamashita M, Engelman AN. Capsid-Dependent Host Factors in HIV-1 Infection. *Trends Microbiol*. sept 2017;25(9):741-55.
  34. Bhattacharya A, Alam SL, Fricke T, Zadrozny K, Sedzicki J, Taylor AB, et al. Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. *Proc Natl Acad Sci*. 30 déc 2014;111(52):18625-30.
  35. Fernandez J, Machado AK, Lyonnais S, Chamontin C, Gärtner K, Léger T, et al. Transportin-1 binds to the HIV-1 capsid via a nuclear localization signal and triggers uncoating. *Nat Microbiol*. nov 2019;4(11):1840-50.
  36. Maertens GN, Cook NJ, Wang W, Hare S, Gupta SS, Öztop I, et al. Structural basis for

## BIBLIOGRAPHIE

- nuclear import of splicing factors by human Transportin 3. *Proc Natl Acad Sci.* 18 févr 2014;111(7):2728-33.
37. Hu WS, Hughes SH. HIV-1 Reverse Transcription. *Cold Spring Harb Perspect Med.* 1 oct 2012;2(10):a006882-a006882.
38. Krupkin M, Jackson LN, Ha B, Puglisi EV. Advances in understanding the initiation of HIV-1 reverse transcription. *Curr Opin Struct Biol.* déc 2020;65:175-83.
39. Menéndez-Arias L, Sebastián-Martín A, Álvarez M. Viral reverse transcriptases. *Virus Res.* avr 2017;234:153-76.
40. Thierry S, Thierry E, Subra F, Deprez E, Leh H, Bury-Moné S, et al. Opposite transcriptional regulation of integrated vs unintegrated HIV genomes by the NF- $\kappa$ B pathway. *Sci Rep.* mai 2016;6(1):25678.
41. Engelman A, Cherepanov P. The Lentiviral Integrase Binding Protein LEDGF/p75 and HIV-1 Replication. *Finlay BB, éditeur. PLoS Pathog.* 28 mars 2008;4(3):e1000046.
42. Freed EO. HIV-1 replication. *Somat Cell Mol Genet.* 2001;26(1/6):13-33.
43. Zhang Z, Klatt A, Gilmour DS, Henderson AJ. Negative Elongation Factor NELF Represses Human Immunodeficiency Virus Transcription by Pausing the RNA Polymerase II Complex. *J Biol Chem.* juin 2007;282(23):16981-8.
44. Yamaguchi Y, Inukai N, Narita T, Wada T, Handa H. Evidence that Negative Elongation Factor Represses Transcription Elongation through Binding to a DRB Sensitivity-Inducing Factor/RNA Polymerase II Complex and RNA. *Mol Cell Biol.* mai 2002;22(9):2918-27.
45. Cullen BR. Nuclear RNA export. *J Cell Sci.* 15 févr 2003;116(4):587-97.
46. Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature.* mars 1989;338(6212):254-7.
47. Mann DA, Mikaélian I, Zimmel RW, Green SM, Lowe AD, Kimura T, et al. A Molecular Rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression. *J Mol Biol.* août 1994;241(2):193-207.
48. Chen J, Nikolaitchik O, Singh J, Wright A, Bencsics CE, Coffin JM, et al. High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc Natl Acad Sci.* 11 août 2009;106(32):13535-40.
49. Bernacchi S, Abd El-Wahab EW, Dubois N, Hijnen M, Smyth RP, Mak J, et al. HIV-1 Pr55(Gag) binds genomic and spliced RNAs with different affinity and stoichiometry. *RNA Biol.* 2017;14(1):90-103.
50. Durand S, Seigneuret F, Burlaud-Gaillard J, Lemoine R, Tassi MF, Moreau A, et al. Quantitative analysis of the formation of nucleoprotein complexes between HIV-1 Gag protein and genomic RNA using transmission electron microscopy. *J Biol Chem.* janv 2022;298(1):101500.
51. Harrison GP, Lever AM. The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J Virol.* juill 1992;66(7):4144-53.
52. Hayashi T, Shioda T, Iwakura Y, Shibuta H. RNA packaging signal of human immunodeficiency virus type 1. *Virology.* juin 1992;188(2):590-9.
53. De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF. Structure of the HIV-1 Nucleocapsid Protein Bound to the SL3  $\Psi$ -RNA Recognition Element. *Science.* 16 janv 1998;279(5349):384-8.

## BIBLIOGRAPHIE

54. Clever JL, Parslow TG. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol.* mai 1997;71(5):3407-14.
55. Maddison B, Marya P, Heaphy S. HIV-1 Gag binds specifically to RNA stem-loops in the 5' leader sequence. *Biochim Biophys Acta BBA - Gene Struct Expr.* juill 1998;1398(3):305-14.
56. Rong L, Russell RS, Hu J, Laughrea M, Wainberg MA, Liang C. Deletion of stem-loop 3 is compensated by second-site mutations within the Gag protein of human immunodeficiency virus type 1. *Virology.* sept 2003;314(1):221-8.
57. Smyth RP, Despons L, Huili G, Bernacchi S, Hijnen M, Mak J, et al. Mutational interference mapping experiment (MIME) for studying RNA structure and function. *Nat Methods.* 3 août 2015;12:866.
58. Mailler E, Bernacchi S, Marquet R, Paillart JC, Vivet-Boudou V, Smyth RP. The Life-Cycle of the HIV-1 Gag-RNA Complex. *Freed EO, éditeur. Viruses.* sept 2016;8(9):248.
59. Comas-Garcia M, Davis S, Rein A. On the Selective Packaging of Genomic RNA by HIV-1. *Viruses.* 12 sept 2016;8(9):246.
60. Dubois N, Khoo KK, Ghossein S, Seissler T, Wolff P, McKinstry WJ, et al. The C-terminal p6 domain of the HIV-1 Pr55<sup>Gag</sup> precursor is required for specific binding to the genomic RNA. *RNA Biol.* 3 juill 2018;15(7):923-36.
61. Brügger B, Glass B, Haberkant P, Leibrecht I, Wieland FT, Kräusslich HG. The HIV lipidome: A raft with an unusual composition. *Proc Natl Acad Sci.* 21 févr 2006;103(8):2641-6.
62. Göttlinger HG, Dorfman T, Sodroski JG, Haseltine WA. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci.* 15 avr 1991;88(8):3195-9.
63. Dubé M, Bhusan Roy B, Guiot-Guillain P, Binette J, Mercier J, Chiasson A, et al. Antagonism of Tetherin Restriction of HIV-1 Release by Vpu Involves Binding and Sequestration of the Restriction Factor in a Perinuclear Compartment. *Luban J, éditeur. PLoS Pathog.* 8 avr 2010;6(4):e1000856.
64. Ohishi M, Nakano T, Sakuragi S, Shioda T, Sano K, Sakuragi J i. The relationship between HIV-1 genome RNA dimerization, virion maturation and infectivity. *Nucleic Acids Res.* 1 avr 2011;39(8):3404-17.
65. Harrison JJEK, Passos DO, Bruhn JF, Bauman JD, Tuberty L, DeStefano JJ, et al. Cryo-EM structure of the HIV-1 Pol polyprotein provides insights into virion maturation. *Sci Adv.* 8 juill 2022;8(27):eabn9874.
66. Song R, Kafaie J, Yang L, Laughrea M. HIV-1 Viral RNA Is Selected in the Form of Monomers that Dimerize in a Three-step Protease-dependent Process; the DIS of Stem-Loop 1 Initiates Viral RNA Dimerization. *J Mol Biol.* août 2007;371(4):1084-98.
67. Jalalirad M, Laughrea M. Formation of immature and mature genomic RNA dimers in wild-type and protease-inactive HIV-1: Differential roles of the Gag polyprotein, nucleocapsid proteins NCp15, NCp9, NCp7, and the dimerization initiation site. *Virology.* nov 2010;407(2):225-36.
68. Paillart JC, Berthoux L, Ottmann M, Darlix JL, Marquet R, Ehresmann B, et al. A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J Virol.* déc 1996;70(12):8348-54.
69. Berkhout B, Das AT, van Wamel JLB. The Native Structure of the Human Immunodeficiency Virus Type 1 RNA Genome Is Required for the First Strand Transfer



## BIBLIOGRAPHIE

- of Reverse Transcription. *Virology*. sept 1998;249(2):211-8.
70. Rein A. RNA Packaging in HIV. *Trends Microbiol*. mai 2019;S0966842X19300964.
  71. Yedavalli VSRK, Jeang KT. Trimethylguanosine capping selectively promotes expression of Rev-dependent HIV-1 RNAs. *Proc Natl Acad Sci*. 17 août 2010;107(33):14787-92.
  72. Gilmer O, Mailler E, Paillart JC, Mouhand A, Tisné C, Mak J, et al. Structural maturation of the HIV-1 RNA 5' untranslated region by Pr55<sup>Gag</sup> and its maturation products. *RNA Biol*. 31 déc 2022;19(1):191-205.
  73. Mouzakis KD, Lang AL, Vander Meulen KA, Easterday PD, Butcher SE. HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome. *Nucleic Acids Res*. févr 2013;41(3):1901-13.
  74. Rausch J, Grice S. HIV Rev Assembly on the Rev Response Element (RRE): A Structural Perspective. *Viruses*. 12 juin 2015;7(6):3053-75.
  75. Kao SY, Calman AF, Luciw PA, Peterlin BM. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature*. déc 1987;330(6147):489-93.
  76. Berkhout B, Silverman RH, Jeang KT. Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell*. oct 1989;59(2):273-82.
  77. Selby MJ, Bain ES, Luciw PA, Peterlin BM. Structure, sequence, and position of the stem-loop in tar determine transcriptional elongation by tat through the HIV-1 long terminal repeat. *Genes Dev*. avr 1989;3(4):547-58.
  78. Aboul-ela F, Karn J, Varani G. The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein. *J Mol Biol*. oct 1995;253(2):313-32.
  79. Wei P, Garber ME, Fang SM, Fischer WH, Jones KA. A Novel CDK9-Associated C-Type Cyclin Interacts Directly with HIV-1 Tat and Mediates Its High-Affinity, Loop-Specific Binding to TAR RNA. *Cell*. févr 1998;92(4):451-62.
  80. Mousseau G, Valente S. Strategies to Block HIV Transcription: Focus on Small Molecule Tat Inhibitors. *Biology*. 19 nov 2012;1(3):668-97.
  81. Jeeninga RE, Hoogenkamp M, Armand-Ugon M, de Baar M, Verhoef K, Berkhout B. Functional Differences between the Long Terminal Repeat Transcriptional Promoters of Human Immunodeficiency Virus Type 1 Subtypes A through G. *J Virol*. 15 avr 2000;74(8):3740-51.
  82. Montano MA, Novitsky VA, Blackard JT, Cho NL, Katzenstein DA, Essex M. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. *J Virol*. nov 1997;71(11):8657-65.
  83. Klasens BIF, Thiesen M, Virtanen A, Berkhout B. The ability of the HIV-1 AAUAAA signal to bind polyadenylation factors is controlled by local RNA structure. *Nucleic Acids Res*. 1 janv 1999;27(2):446-54.
  84. Smyth RP, Smith MR, Jousset AC, Despons L, Laumond G, Decoville T, et al. In cell mutational interference mapping experiment (in cell MIME) identifies the 5' polyadenylation signal as a dual regulator of HIV-1 genomic RNA production and packaging. *Nucleic Acids Res*. 18 mai 2018;46(9):e57-e57.
  85. Ashe MP, Furger A, Proudfoot NJ. Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. *RNA*. févr 2000;6(2):170-7.
  86. Mueller N, Klaver B, Berkhout B, Das AT. Human immunodeficiency virus type 1 splicing at the major splice donor site is controlled by highly conserved RNA sequence

## BIBLIOGRAPHIE

- and structural elements. *J Gen Virol.* 1 nov 2015;96(11):3389-95.
87. Abbink TEM, Berkhout B. RNA Structure Modulates Splicing Efficiency at the Human Immunodeficiency Virus Type 1 Major Splice Donor. *J Virol.* 15 mars 2008;82(6):3090-8.
  88. Mueller N, van Bel N, Berkhout B, Das AT. HIV-1 splicing at the major splice donor site is restricted by RNA structure. *Virology.* nov 2014;468-470:609-20.
  89. Shoemaker CJ, Green R. Translation drives mRNA quality control. *Nat Struct Mol Biol.* juin 2012;19(6):594-601.
  90. Zapp ML, Green MR. Sequence-specific RNA binding by the HIV-1 Rev protein. *Nature.* déc 1989;342(6250):714-6.
  91. Vercruyssen T, Daelemans D. HIV-1 Rev Multimerization: Mechanism and Insights. *Curr HIV Res.* 31 avr 2014;11(8):623-34.
  92. Rao S, Amorim R, Niu M, Breton Y, Tremblay MJ, Mouland AJ. Host mRNA decay proteins influence HIV-1 replication and viral gene expression in primary monocyte-derived macrophages. *Retrovirology.* déc 2019;16(1):3.
  93. Banerjee A, Benjamin R, Balakrishnan K, Ghosh P, Banerjee S. Human protein Staufen-2 promotes HIV-1 proliferation by positively regulating RNA export activity of viral protein Rev. *Retrovirology.* déc 2014;11(1):18.
  94. Ajamian L, Abel K, Rao S, Vyboh K, García-de-Gracia F, Soto-Rifo R, et al. HIV-1 Recruits UPF1 but Excludes UPF2 to Promote Nucleocytoplasmic Export of the Genomic RNA. *Biomolecules.* 20 oct 2015;5(4):2808-39.
  95. Guerrero S, Batisse J, Libre C, Bernacchi S, Marquet R, Paillart JC. HIV-1 Replication and the Cellular Eukaryotic Translation Apparatus. *Viruses.* 19 janv 2015;7(1):199-218.
  96. Anderson JL, Johnson AT, Howard JL, Purcell DFJ. Both Linear and Discontinuous Ribosome Scanning Are Used for Translation Initiation from Bicistronic Human Immunodeficiency Virus Type 1 *env* mRNAs. *J Virol.* mai 2007;81(9):4664-76.
  97. Krummheuer J, Johnson AT, Hauber I, Kammler S, Anderson JL, Hauber J, et al. A minimal uORF within the HIV-1 vpu leader allows efficient translation initiation at the downstream *env* AUG. *Virology.* juill 2007;363(2):261-71.
  98. Schwartz S, Felber BK, Pavlakis GN. Mechanism of translation of monocistronic and multicistronic human immunodeficiency virus type 1 mRNAs. *Mol Cell Biol.* janv 1992;12(1):207-19.
  99. Brierley I, Dos Ramos FJ. Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res.* juill 2006;119(1):29-42.
  100. Giedroc DP, Cornish PV. Frameshifting RNA pseudoknots: Structure and mechanism. *Virus Res.* févr 2009;139(2):193-208.
  101. Low JT, Garcia-Miranda P, Mouzakis KD, Gorelick RJ, Butcher SE, Weeks KM. Structure and Dynamics of the HIV-1 Frameshift Element RNA. *Biochemistry.* 8 juill 2014;53(26):4282-91.
  102. Brasey A, Lopez-Lastra M, Ohlmann T, Beerens N, Berkhout B, Darlix JL, et al. The Leader of Human Immunodeficiency Virus Type 1 Genomic RNA Harbors an Internal Ribosome Entry Segment That Is Active during the G<sub>2</sub>/M Phase of the Cell Cycle. *J Virol.* avr 2003;77(7):3939-49.
  103. Buck CB, Shen X, Egan MA, Pierson TC, Walker CM, Siliciano RF. The Human Immunodeficiency Virus Type 1 *gag* Gene Encodes an Internal Ribosome Entry Site. *J Virol.* janv 2001;75(1):181-91.

## BIBLIOGRAPHIE

104. Deforges J, de Breyne S, Ameer M, Ulryck N, Chamond N, Saaidi A, et al. Two ribosome recruitment sites direct multiple translation events within HIV1 Gag open reading frame. *Nucleic Acids Res.* 7 juill 2017;45(12):7382-400.
105. Sorokin II, Vassilenko KS, Terenin IM, Kalinina NO, Agol VI, Dmitriev SE. Non-Canonical Translation Initiation Mechanisms Employed by Eukaryotic Viral mRNAs. *Biochem Mosc.* sept 2021;86(9):1060-94.
106. Boeras I, Seufzer B, Brady S, Rendahl A, Heng X, Boris-Lawrie K. The basal translation rate of authentic HIV-1 RNA is regulated by 5'UTR nt-pairings at junction of R and U5. *Sci Rep.* déc 2017;7(1):6902.
107. Chen J, Liu Y, Wu B, Nikolaitchik OA, Mohan PR, Chen J, et al. Visualizing the translation and packaging of HIV-1 full-length RNA. *Proc Natl Acad Sci.* 17 mars 2020;117(11):6145-55.
108. Ferrer M, Clerté C, Chamontin C, Basyuk E, Lainé S, Hottin J, et al. Imaging HIV-1 RNA dimerization in cells by multicolor super-resolution and fluctuation microscopies. *Nucleic Acids Res.* 19 sept 2016;44(16):7922-34.
109. Chen J, Rahman SA, Nikolaitchik OA, Grunwald D, Sardo L, Burdick RC, et al. HIV-1 RNA genome dimerizes on the plasma membrane in the presence of Gag protein. *Proc Natl Acad Sci [Internet].* 12 janv 2016 [cité 16 oct 2022];113(2). Disponible sur: <https://pnas.org/doi/full/10.1073/pnas.1518572113>
110. Skripkin E, Paillart JC, Marquet R, Ehresmann B, Ehresmann C. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc Natl Acad Sci.* 24 mai 1994;91(11):4945-9.
111. Dubois N, Marquet R, Paillart JC, Bernacchi S. Retroviral RNA Dimerization: From Structure to Functions. *Front Microbiol.* 2018;9:527.
112. Bieniasz P, Telesnitsky A. Multiple, Switchable Protein:RNA Interactions Regulate Human Immunodeficiency Virus Type 1 Assembly. *Annu Rev Virol.* 2018;5(1):165-83.
113. Keane SC, Summers MF. NMR Studies of the Structure and Function of the HIV-1 5'-Leader. *Viruses.* 21 2016;8(12).
114. Lerner G, Weaver N, Anokhin B, Spearman P. Advances in HIV-1 Assembly. *Viruses.* 26 févr 2022;14(3):478.
115. Luban J, Goff SP. Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. *J Virol.* juin 1994;68(6):3784-93.
116. Lu K, Heng X, Summers MF. Structural Determinants and Mechanism of HIV-1 Genome Packaging. *Struct Mol Biol HIV.* 22 juill 2011;410(4):609-33.
117. Ingemarsdotter CK, Zeng J, Long Z, Lever AM, Kenyon JC. An RNA-binding compound that stabilizes the HIV-1 gRNA packaging signal structure and specifically blocks HIV-1 RNA encapsidation. *Retrovirology.* 2018;15:25.
118. Abd El-Wahab EW, Smyth RP, Mailler E, Bernacchi S, Vivet-Boudou V, Hijnen M, et al. Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat Commun.* 2 juill 2014;5:4304.
119. Nikolaitchik OA, Somoulay X, Rawson JMO, Yoo JA, Pathak VK, Hu WS. Unpaired Guanosines in the 5' Untranslated Region of HIV-1 RNA Act Synergistically To Mediate Genome Packaging. *Simon V, éditeur. J Virol.* 12 août 2020;94(21):e00439-20, /jvi/94/21/JVI.00439-20.atom.
120. Ding P, Kharytonchik S, Waller A, Mbaekwe U, Basappa S, Kuo N, et al. Identification of the initial nucleocapsid recognition element in the HIV-1 RNA packaging signal. *Proc Natl Acad Sci.* 28 juill 2020;117(30):17737-46.

## BIBLIOGRAPHIE

121. Telesnitsky A, Wolin S. The Host RNAs in Retroviral Particles. *Viruses*. 19 août 2016;8(8):235.
122. Song Z, Gremminger T, Singh G, Cheng Y, Li J, Qiu L, et al. The three-way junction structure of the HIV-1 PBS-segment binds host enzyme important for viral infectivity. *Nucleic Acids Res*. 4 juin 2021;49(10):5925-42.
123. Yao RW, Wang Y, Chen LL. Cellular functions of long noncoding RNAs. *Nat Cell Biol*. mai 2019;21(5):542-51.
124. Wang KC, Chang HY. Molecular Mechanisms of Long Noncoding RNAs. *Mol Cell*. sept 2011;43(6):904-14.
125. Cantara WA, Olson ED, Musier-Forsyth K. Progress and outlook in structural biology of large viral RNAs. *Virus Res*. nov 2014;193:24-38.
126. Wang XW, Liu CX, Chen LL, Zhang QC. RNA structure probing uncovers RNA structure-dependent biological functions. *Nat Chem Biol*. juill 2021;17(7):755-66.
127. Getz M, Sun X, Casiano-Negrone A, Zhang Q, Al-Hashimi HM. Review NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. *Biopolymers*. 5 août 2007;86(5-6):384-402.
128. Li B, Cao Y, Westhof E, Miao Z. Advances in RNA 3D Structure Modeling Using Experimental Data. *Front Genet*. 26 oct 2020;11:574485.
129. Shi Y. A Glimpse of Structural Biology through X-Ray Crystallography. *Cell*. nov 2014;159(5):995-1014.
130. Yoffe AM, Prinsen P, Gopal A, Knobler CM, Gelbart WM, Ben-Shaul A. Predicting the sizes of large RNA molecules. *Proc Natl Acad Sci*. 21 oct 2008;105(42):16153-8.
131. Vinayagam D, Mager T, Apelbaum A, Bothe A, Merino F, Hofnagel O, et al. Electron cryo-microscopy structure of the canonical TRPC4 ion channel. *eLife*. 2 mai 2018;7:e36615.
132. Zhang K, Li S, Kappel K, Pintilie G, Su Z, Mou TC, et al. Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nat Commun*. déc 2019;10(1):5511.
133. Zhang K, Keane SC, Su Z, Irobalieva RN, Chen M, Van V, et al. Structure of the 30 kDa HIV-1 RNA Dimerization Signal by a Hybrid Cryo-EM, NMR, and Molecular Dynamics Approach. *Structure*. 6 mars 2018;26(3):490-498.e3.
134. Gopal A, Zhou ZH, Knobler CM, Gelbart WM. Visualizing large RNA molecules in solution. *RNA*. févr 2012;18(2):284-99.
135. Chaloner PA. Two-dimensional NMR methods for establishing molecular connectivity; a chemist's guide to experiment selection, performance and interpretation. *J Organomet Chem*. avr 1990;386(2):C16-7.
136. Hermann T, Westhof E. Non-Watson-Crick base pairs in RNA-protein recognition. *Chem Biol*. déc 1999;6(12):R335-43.
137. Bothe JR, Nikolova EN, Eichhorn CD, Chugh J, Hansen AL, Al-Hashimi HM. Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat Methods*. nov 2011;8(11):919-31.
138. Fang X, Stagno JR, Bhandari YR, Zuo X, Wang YX. Small-angle X-ray scattering: a bridge between RNA secondary structures and three-dimensional topological structures. *Curr Opin Struct Biol*. févr 2015;30:147-60.
139. Lipfert J, Herschlag D, Doniach S. Riboswitch Conformations Revealed by Small-Angle X-Ray Scattering. In: Serganov A, éditeur. *Riboswitches* [Internet]. Totowa, NJ: Humana Press; 2009 [cité 16 oct 2022]. p. 141-59. (Methods in Molecular Biology; vol. 540). Disponible sur: [http://link.springer.com/10.1007/978-1-59745-558-9\\_11](http://link.springer.com/10.1007/978-1-59745-558-9_11)

## BIBLIOGRAPHIE

140. Gilmer O, Quignon E, Jousset AC, Paillart JC, Marquet R, Vivet-Boudou V. Chemical and Enzymatic Probing of Viral RNAs: From Infancy to Maturity and Beyond. *Viruses*. 22 sept 2021;13(10):1894.
141. Bender W, Davidson N. Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules. *Cell*. avr 1976;7(4):595-607.
142. Coffin JM, Hageman TC, Maxam AM, Haseltine WA. Structure of the genome of moloney murine leukemia virus: a terminally redundant sequence. *Cell*. avr 1978;13(4):761-73.
143. Darlix JL, Spahr PF, Bromley PA. Analysis of rous sarcoma virus (RSV) RNA structure by means of specific nucleases. *Virology*. oct 1978;90(2):317-29.
144. Swanstrom R, Varmus HE, Bishop JM. Nucleotide sequence of the 5' noncoding region and part of the gag gene of Rous sarcoma virus. *J Virol*. févr 1982;41(2):535-41.
145. Okamoto T, Wong-Staal F. Demonstration of virus-specific transcriptional activator(s) in cells infected with HTLV-III by an in vitro cell-free system. *Cell*. oct 1986;47(1):29-35.
146. Parkin NT, Cohen EA, Darveau A, Rosen C, Haseltine W, Sonenberg N. Mutational analysis of the 5' non-coding region of human immunodeficiency virus type 1: effects of secondary structure on translation. *EMBO J*. sept 1988;7(9):2831-7.
147. Hanly SM, Rimsky LT, Malim MH, Kim JH, Hauber J, Duc Dodon M, et al. Comparative analysis of the HTLV-I Rex and HIV-1 Rev trans-regulatory proteins and their RNA response elements. *Genes Dev*. oct 1989;3(10):1534-44.
148. Seiki M, Inoue J, Hidaka M, Yoshida M. Two cis-acting elements responsible for posttranscriptional trans-regulation of gene expression of human T-cell leukemia virus type I. *Proc Natl Acad Sci*. oct 1988;85(19):7124-8.
149. Barat C, Lullien V, Schatz O, Keith G, Nugeyre MT, Grüninger-Leitch F, et al. HIV-1 reverse transcriptase specifically interacts with the anticodon domain of its cognate primer tRNA. *EMBO J*. nov 1989;8(11):3279-85.
150. Darlix JL, Gabus C, Nugeyre MT, Clavel F, Barré-Sinoussi F. Cis elements and Trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J Mol Biol*. déc 1990;216(3):689-99.
151. Lever A, Gottlinger H, Haseltine W, Sodroski J. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J Virol*. sept 1989;63(9):4085-7.
152. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*. 6 août 2009;460(7256):711-6.
153. Tinoco I, Borer PN, Dengler B, Levine MD, Uhlenbeck OC, Crothers DM, et al. Improved Estimation of Secondary Structure in Ribonucleic Acids. *Nature New Biol*. nov 1973;246(150):40-1.
154. Okamoto T, Benter T, Josephs SF, Reza Sadaie M, Wong-Staal F. Transcriptional activation from the long-terminal repeat of human immunodeficiency virus in Vitro. *Virology*. août 1990;177(2):606-14.
155. Muesing M, Smith DH, Capon DJ. Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein. *Cell*. févr 1987;48(4):691-701.
156. Colvin RA, Garcia-Blanco MA. Unusual structure of the human immunodeficiency virus type 1 trans-activation response element. *J Virol*. févr 1992;66(2):930-5.

## BIBLIOGRAPHIE

157. Tao J, Frankel AD. Specific binding of arginine to TAR RNA. *Proc Natl Acad Sci.* avr 1992;89(7):2723-6.
158. Baudin F, Marquet R, Isel C, Darlix JL, Ehresmann B, Ehresmann C. Functional Sites in the 5' Region of Human Immunodeficiency Virus Type 1 RNA Form Defined Structural Domains. *J Mol Biol.* janv 1993;229(2):382-97.
159. Delling U, Reid LS, Barnett RW, Ma MY, Climie S, Sumner-Smith M, et al. Conserved nucleotides in the TAR RNA stem of human immunodeficiency virus type 1 are critical for Tat binding and trans activation: model for TAR RNA tertiary structure. *J Virol.* mai 1992;66(5):3018-25.
160. Puglisi JD, Tan R, Calnan BJ, Frankel AD, Williamson JR. Conformation of the TAR RNA-Arginine Complex by NMR Spectroscopy. *Science.* 3 juill 1992;257(5066):76-80.
161. Colvin RA, White SW, Garcia-Blanco MA, Hoffman DW. Structural features of an RNA containing the CUGGGA loop of the human immunodeficiency virus type 1 trans-activation response element. *Biochemistry.* 1 févr 1993;32(4):1105-12.
162. Jaeger JA, Tinoco I. An NMR study of the HIV-1 TAR element hairpin. *Biochemistry.* nov 1993;32(46):12522-30.
163. Salsler W. Globin mRNA Sequences: Analysis of Base Pairing and Evolutionary Implications. *Cold Spring Harb Symp Quant Biol.* 1 janv 1978;42(0):985-1002.
164. Cech TR, Tanner NK, Tinoco I, Weir BR, Zuker M, Perlman PS. Secondary structure of the Tetrahymena ribosomal RNA intervening sequence: structural homology with fungal mitochondrial intervening sequences. *Proc Natl Acad Sci.* juill 1983;80(13):3903-7.
165. Malim MH, Tiley LS, McCarn DF, Rusche JR, Hauber J, Cullen BR. HIV-1 structural gene expression requires binding of the rev trans-activator to its RNA target sequence. *Cell.* févr 1990;60(4):675-83.
166. Borer PN, Dengler B, Tinoco I, Uhlenbeck OC. Stability of ribonucleic acid double-stranded helices. *J Mol Biol.* juill 1974;86(4):843-53.
167. Kimura T, Ohyama A. Interaction with the Rev Response Element along an Extended Stem I Duplex Structure Is Required to Complete Human Immunodeficiency Virus Type 1 rev-Mediated trans-Activation In Vivo. *J Biochem (Tokyo).* mai 1994;115(5):945-52.
168. Bartel DP, Zapp ML, Green MR, Szostak JW. HIV-1 rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell.* nov 1991;67(3):529-36.
169. Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, et al.  $\alpha$  Helix-RNA Major Groove Recognition in an HIV-1 Rev Peptide-RRE RNA Complex. *Science.* 13 sept 1996;273(5281):1547-51.
170. Charpentier B, Stutz F, Rosbash M. A dynamic in Vivo view of the HIV-IRev-RRE interaction. *J Mol Biol.* mars 1997;266(5):950-62.
171. Holland SM, Ahmad N, Maitra RK, Wingfield P, Venkatesan S. Human immunodeficiency virus rev protein recognizes a target sequence in rev-responsive element RNA within the context of RNA secondary structure. *J Virol.* déc 1990;64(12):5966-75.
172. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature.* janv 1988;331(6153):280-3.
173. Parkin NT, Chamorro M, Varmus HE. Human immunodeficiency virus type 1 gag-pol frameshifting is dependent on downstream mRNA secondary structure:

## BIBLIOGRAPHIE

- demonstration by expression in vivo. *J Virol.* août 1992;66(8):5147-51.
174. Shu-Yun Le, Shapiro BA, Chen JH, Nussinov R, Maizel JV. RNA pseudoknots downstream of the frameshift sites of retroviruses. *Genet Anal Biomol Eng.* nov 1991;8(7):191-205.
  175. Dulude D. Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res.* 1 déc 2002;30(23):5094-102.
  176. Isel C, Marquet R, Keith G, Ehresmann C, Ehresmann B. Modified nucleotides of tRNA(3Lys) modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription. *J Biol Chem.* 5 déc 1993;268(34):25269-72.
  177. Isel C, Ehresmann C, Keith G, Ehresmann B, Marquet R. Initiation of Reverse Transcription of HIV-1: Secondary Structure of the HIV-1 RNA/tRNA<sup>Lys</sup> Loop-Loop Complex. *J Mol Biol.* mars 1995;247(2):236-50.
  178. Skripkin E, Isel C, Marquet R, Ehresmann B, Ehresmann C. Psoralen Crosslinking between Human Immunodeficiency Virus Type 1 RNA and Primer tRNA<sup>3Lys</sup>. *Nucleic Acids Res.* 1 févr 1996;24(3):509-14.
  179. Marquet R, Baudin F, Gabus C, Darlix JL, Mougél M, Ehresmann C, et al. Dimerization of human immunodeficiency virus (type 1) RNA: stimulation by cations and possible mechanism. *Nucleic Acids Res.* 1991;19(9):2349-57.
  180. Sundquist WI, Heaphy S. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proc Natl Acad Sci.* 15 avr 1993;90(8):3393-7.
  181. Laughrea M, Jette L. A 19-Nucleotide Sequence Upstream of the 5' Major Splice Donor Is Part of the Dimerization Domain of Human Immunodeficiency Virus 1 Genomic RNA. *Biochemistry.* 15 nov 1994;33(45):13464-74.
  182. Muriaux D, Girard PM, Bonnet-Mathonière B, Paoletti J. Dimerization of HIV-1 RNA at Low Ionic Strength. *J Biol Chem.* avr 1995;270(14):8209-16.
  183. Paillart JC, Skripkin E, Ehresmann B, Ehresmann C, Marquet R. 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.* 28 mai 1996;93(11):5572-7.
  184. Ennifar E, Walter P, Ehresmann B, Ehresmann C, Dumas P. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol.* 1 déc 2001;8(12):1064-8.
  185. Alford RL, Honda S, Lawrence CB, Belmont JW. RNA secondary structure analysis of the packaging signal for Moloney murine leukemia virus. *Virology.* août 1991;183(2):611-9.
  186. Bender MA, Palmer TD, Gelinas RE, Miller AD. Evidence that the packaging signal of Moloney murine leukemia virus extends into the gag region. *J Virol.* mai 1987;61(5):1639-46.
  187. Pugatsch T, Stacey DW. Identification of a sequence likely to be required for avian retroviral packaging. *Virology.* 30 juill 1983;128(2):505-11.
  188. Sorge J, Ricci W, Hughes SH. cis-Acting RNA packaging locus in the 115-nucleotide direct repeat of Rous sarcoma virus. *J Virol.* déc 1983;48(3):667-75.
  189. Gorelick RJ, Henderson LE, Hanser JP, Rein A. Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a « zinc finger-like » protein sequence. *Proc Natl Acad Sci.* nov 1988;85(22):8420-4.

## BIBLIOGRAPHIE

190. Méric C, Goff SP. Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys-His box of the nucleocapsid protein. *J Virol.* avr 1989;63(4):1558-68.
191. Greatorex J, Gallego J, Varani G, Lever A. Structure and Stability of Wild-type and Mutant RNA Internal Loops from the SL-1 Domain of the HIV-1 Packaging Signal. *J Mol Biol.* sept 2002;322(3):543-57.
192. Zeffman A, Hassard S, Varani G, Lever A. The major HIV-1 packaging signal is an extended bulged stem loop whose structure is altered on interaction with the gag polyprotein. Edited by I. Tinoco. *J Mol Biol.* 7 avr 2000;297(4):877-93.
193. Han K, Kim HJ. Prediction of common folding structures of homologous RNAs. *Nucleic Acids Res.* 1993;21(5):1251-7.
194. Sakaguchi K, Zambrano N, Baldwin ET, Shapiro BA, Erickson JW, Omichinski JG, et al. Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. *Proc Natl Acad Sci.* juin 1993;90(11):5219-23.
195. Clever J, Sasseti C, Parslow TG. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J Virol.* avr 1995;69(4):2101-9.
196. Hayashi T, Ueno Y, Okamoto T. Elucidation of a conserved RNA stem-loop structure in the packaging signal of human immunodeficiency virus type 1. *FEBS Lett.* 26 juill 1993;327(2):213-8.
197. Rizvi TA, Panganiban AT. Simian immunodeficiency virus RNA is efficiently encapsidated by human immunodeficiency virus type 1 particles. *J Virol.* mai 1993;67(5):2681-8.
198. Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics.* 1 août 2009;25(15):1974-5.
199. Berkhout B. Structure and Function of the Human Immunodeficiency Virus Leader RNA. In: *Progress in Nucleic Acid Research and Molecular Biology* [Internet]. Elsevier; 1996 [cité 16 oct 2022]. p. 1-34. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S0079660308603591>
200. McBride MS, Schwartz MD, Panganiban AT. Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation. *J Virol.* juin 1997;71(6):4544-54.
201. Damgaard CK, Dyhr-Mikkelsen H, Kjems J. Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions. *Nucleic Acids Res.* 15 août 1998;26(16):3667-76.
202. Huthoff H, Berkhout B. Two alternating structures of the HIV-1 leader RNA. *RNA.* janv 2001;7(1):143-57.
203. Berkhout B, Van Wamel JLB. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA.* févr 2000;6(2):282-95.
204. Huthoff H, Berkhout B. Mutations in the TAR hairpin affect the equilibrium between alternative conformations of the HIV-1 leader RNA. *Nucleic Acids Res.* 15 juin 2001;29(12):2594-600.
205. Abbink TEM, Berkhout B. A Novel Long Distance Base-pairing Interaction in Human Immunodeficiency Virus Type 1 RNA Occludes the Gag Start Codon. *J Biol Chem.* 28 mars 2003;278(13):11601-11.
206. Paillart JC, Skripkin E, Ehresmann B, Ehresmann C, Marquet R. In Vitro Evidence for a Long Range Pseudoknot in the 5'-Untranslated and Matrix Coding Regions of HIV-1



## BIBLIOGRAPHIE

- Genomic RNA. *J Biol Chem.* févr 2002;277(8):5995-6004.
207. Damgaard CK, Andersen ES, Knudsen B, Gorodkin J, Kjems J. RNA Interactions in the 5' Region of the HIV-1 Genome. *J Mol Biol.* févr 2004;336(2):369-79.
208. Abbink TEM, Ooms M, Haasnoot PCJ, Berkhout B. The HIV-1 Leader RNA Conformational Switch Regulates RNA Dimerization but Does Not Regulate mRNA Translation. *Biochemistry.* 1 juin 2005;44(25):9058-66.
209. Ye L, Gribling-Burrer AS, Bohn P, Kibe A, Börtlein C, Ambi UB, et al. Short- and long-range interactions in the HIV-1 5' UTR regulate genome dimerization and packaging. *Nat Struct Mol Biol.* avr 2022;29(4):306-19.
210. Kasprzak W, Binderwald E, Shapiro BA. Structural polymorphism of the HIV-1 leader region explored by computational methods. *Nucleic Acids Res.* 9 déc 2005;33(22):7151-63.
211. Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchik S, et al. NMR Detection of Structures in the HIV-1 5'-Leader RNA That Regulate Genome Packaging. *Science.* 14 oct 2011;334(6053):242-5.
212. Spriggs S, Garyu L, Connor R, Summers MF. Potential Intra- and Intermolecular Interactions Involving the Unique-5' Region of the HIV-1 5'-UTR. *Biochemistry.* 9 déc 2008;47(49):13064-73.
213. Kenyon JC, Prestwood LJ, Le Grice SFJ, Lever AML. In-gel probing of individual RNA conformers within a mixed population reveals a dimerization structural switch in the HIV-1 leader. *Nucleic Acids Res.* 1 oct 2013;41(18):e174-e174.
214. Keane SC, Heng X, Lu K, Kharytonchik S, Ramakrishnan V, Carter G, et al. Structure of the HIV-1 RNA packaging signal. *Science.* 22 mai 2015;348(6237):917-21.
215. Kharytonchik S, Monti S, Smaldino PJ, Van V, Bolden NC, Brown JD, et al. Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proc Natl Acad Sci.* 22 nov 2016;113(47):13378-83.
216. Brown JD, Kharytonchik S, Chaudry I, Iyer AS, Carter H, Becker G, et al. Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science.* 24 avr 2020;368(6489):413-7.
217. Obayashi CM, Shinohara Y, Masuda T, Kawai G. Influence of the 5'-terminal sequences on the 5'-UTR structure of HIV-1 genomic RNA. *Sci Rep.* déc 2021;11(1):10920.
218. Ding P, Summers MF. Sequestering the 5'-cap for viral RNA packaging. *BioEssays.* 13 sept 2022;2200104.
219. Paillart JC, Dettenhofer M, Yu X fang, Ehresmann C, Ehresmann B, Marquet R. First Snapshots of the HIV-1 RNA Structure in Infected Cells and in Virions. *J Biol Chem.* 12 nov 2004;279(46):48397-403.
220. Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* 29 avr 2008;6(4):e96.
221. Kladow W, VanLang CC, Cordero P, Das R. Understanding the Errors of SHAPE-Directed RNA Structure Modeling. *Biochemistry.* 20 sept 2011;50(37):8049-56.
222. Vamva E, Griffiths A, Vink CA, Lever AML, Kenyon JC. A novel role for *gag* as a *cis*-acting element regulating RNA structure, dimerization and packaging in HIV-1 lentiviral vectors. *Nucleic Acids Res.* 11 janv 2022;50(1):430-48.
223. Bell NM, Lever AML. HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol.* mars 2013;21(3):136-44.

## BIBLIOGRAPHIE

224. Hanson HM, Willkomm NA, Yang H, Mansky LM. Human Retrovirus Genomic RNA Packaging. *Viruses*. 19 mai 2022;14(5):1094.
225. Amarasinghe GK, De Guzman RN, Turner RB, Chancellor KJ, Wu ZR, Summers MF. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the  $\Psi$ -RNA packaging signal. implications for genome recognition 1 Edited by P. Wright. *J Mol Biol*. août 2000;301(2):491-511.
226. Amarasinghe GK, De Guzman RN, Turner RB, Summers MF. NMR structure of stem-loop SL2 of the HIV-1  $\Psi$  RNA packaging signal reveals a novel A-U-A base-triple platform 1 Edited by I. Tinoco. *J Mol Biol*. mai 2000;299(1):145-56.
227. Amarasinghe GK, Zhou J, Miskimon M, Chancellor KJ, McDonald JA, Matthews AG, et al. Stem-loop SL4 of the HIV-1  $\Psi$  RNA packaging signal exhibits weak affinity for the nucleocapsid protein. structural studies and implications for genome recognition. *J Mol Biol*. déc 2001;314(5):961-70.
228. Kutluay SB, Zang T, Blanco-Melo D, Powell C, Jannain D, Errando M, et al. Global Changes in the RNA Binding Specificity of HIV-1 Gag Regulate Virion Genesis. *Cell*. nov 2014;159(5):1096-109.
229. Zargarian L, Tisé C, Barraud P, Xu X, Morellet N, René B, et al. Dynamics of Linker Residues Modulate the Nucleic Acid Binding Properties of the HIV-1 Nucleocapsid Protein Zinc Fingers. Mayer C, éditeur. *PLoS ONE*. 16 juill 2014;9(7):e102150.
230. Comas-Garcia M, Kroupa T, Datta SA, Harvin DP, Hu WS, Rein A. Efficient support of virus-like particle assembly by the HIV-1 packaging signal. *eLife*. 2 août 2018;7:e38438.
231. Nikolaitchik OA, Dilley KA, Fu W, Gorelick RJ, Tai SHS, Soheilian F, et al. Dimeric RNA Recognition Regulates HIV-1 Genome Packaging. Krausslich HG, éditeur. *PLoS Pathog*. 21 mars 2013;9(3):e1003249.
232. Dilley KA, Nikolaitchik OA, Galli A, Burdick RC, Levine L, Li K, et al. Interactions between HIV-1 Gag and Viral RNA Genome Enhance Virion Assembly. Beemon KL, éditeur. *J Virol*. 15 août 2017;91(16):e02319-16.
233. Comas-Garcia M, Datta SA, Baker L, Varma R, Gudla PR, Rein A. Dissection of specific binding of HIV-1 Gag to the « packaging signal » in viral RNA. *eLife*. 20 juill 2017;6:e27055.
234. Sarni S, Biswas B, Liu S, Olson ED, Kitzrow JP, Rein A, et al. HIV-1 Gag protein with or without p6 specifically dimerizes on the viral RNA packaging signal. *J Biol Chem*. oct 2020;295(42):14391-401.
235. Carlson LA, Bai Y, Keane SC, Doudna JA, Hurley JH. Reconstitution of selective HIV-1 RNA packaging in vitro by membrane-bound Gag assemblies. *eLife*. 25 juin 2016;5:e14663.
236. McCauley MJ, Rouzina I, Li J, Núñez ME, Williams MC. Significant Differences in RNA Structure Destabilization by HIV-1 Gag $\Delta$ p6 and NCp7 Proteins. *Viruses*. 25 avr 2020;12(5):484.
237. Kafaie J, Dolatshahi M, Ajamian L, Song R, Moulard AJ, Rouiller I, et al. Role of capsid sequence and immature nucleocapsid proteins p9 and p15 in Human Immunodeficiency Virus type 1 genomic RNA dimerization. *Virology*. mars 2009;385(1):233-44.
238. Wang W, Naiyer N, Mitra M, Li J, Williams MC, Rouzina I, et al. Distinct nucleic acid interaction properties of HIV-1 nucleocapsid protein precursor NCp15 explain reduced viral infectivity. *Nucleic Acids Res*. 17 juin 2014;42(11):7145-59.

## BIBLIOGRAPHIE

239. Coren LV, Thomas JA, Chertova E, Sowder RC, Gagliardi TD, Gorelick RJ, et al. Mutational Analysis of the C-Terminal Gag Cleavage Sites in Human Immunodeficiency Virus Type 1. *J Virol.* 15 sept 2007;81(18):10047-54.
240. Cruceanu M. Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res.* 30 janv 2006;34(2):593-605.
241. Mirambeau G, Lyonnais S, Coulaud D, Hameau L, Lafosse S, Jeusset J, et al. Transmission Electron Microscopy Reveals an Optimal HIV-1 Nucleocapsid Aggregation with Single-stranded Nucleic Acids and the Mature HIV-1 Nucleocapsid Protein. *J Mol Biol.* déc 2006;364(3):496-511.
242. Mirambeau G, Lyonnais S, Coulaud D, Hameau L, Lafosse S, Jeusset J, et al. HIV-1 Protease and Reverse Transcriptase Control the Architecture of Their Nucleocapsid Partner. Egli M, éditeur. *PLoS ONE.* 22 août 2007;2(8):e669.
243. Racine PJ, Chamontin C, de Rocquigny H, Bernacchi S, Paillart JC, Mougel M. Requirements for nucleocapsid-mediated regulation of reverse transcription during the late steps of HIV-1 assembly. *Sci Rep.* juin 2016;6(1):27536.
244. Liu S, Koneru PC, Li W, Pathirage C, Engelman AN, Kvaratskhelia M, et al. HIV-1 integrase binding to genomic RNA 5'-UTR induces local structural changes *in vitro* and *in vivo*. *Retrovirology.* déc 2021;18(1):37.
245. Brigham BS, Kitzrow JP, Reyes JPC, Musier-Forsyth K, Munro JB. Intrinsic conformational dynamics of the HIV-1 genomic RNA 5'UTR. *Proc Natl Acad Sci.* 21 mai 2019;116(21):10372-81.
246. Muriaux D, Fossé P, Paoletti J. A Kissing Complex Together with a Stable Dimer Is Involved in the HIV-1 <sub>LAI</sub> RNA Dimerization Process *in Vitro*. *Biochemistry.* 1 janv 1996;35(15):5075-82.
247. Kenyon JC, Prestwood LJ, Lever AML. A novel combined RNA-protein interaction analysis distinguishes HIV-1 Gag protein binding sites from structural change in the viral RNA leader. *Sci Rep.* nov 2015;5(1):14369.
248. Guo C, Yao X, Wang K, Wang J, Wang Y. Comparison of HIV-1 Gag and NCp7 in their selectivity for package signal, affinity for stem-loop 3, and Zn<sup>2+</sup> content. *Biochimie.* déc 2020;179:135-45.
249. Wang Y, Guo C, Wang X, Xu L, Li R, Wang J. The Zinc Content of HIV-1 NCp7 Affects Its Selectivity for Packaging Signal and Affinity for Stem-Loop 3. *Viruses.* 24 sept 2021;13(10):1922.
250. Fu W, Gorelick RJ, Rein A. Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J Virol.* août 1994;68(8):5013-8.
251. Mirambeau G, Lyonnais S, Gorelick RJ. Features, processing states, and heterologous protein interactions in the modulation of the retroviral nucleocapsid protein function. *RNA Biol.* nov 2010;7(6):724-34.
252. Rossio JL, Esser MT, Suryanarayana K, Schneider DK, Bess JW, Vasquez GM, et al. Inactivation of Human Immunodeficiency Virus Type 1 Infectivity with Preservation of Conformational and Functional Integrity of Virion Surface Proteins. *J Virol.* oct 1998;72(10):7992-8001.
253. Gherghe C, Lombo T, Leonard CW, Datta SAK, Bess JW, Gorelick RJ, et al. Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. *Proc Natl Acad Sci.* 9 nov 2010;107(45):19248-53.
254. McKinstry WJ, Hijnen M, Tanwar HS, Sparrow LG, Nagarajan S, Pham ST, et al.

## BIBLIOGRAPHIE

- Expression and purification of soluble recombinant full length HIV-1 Pr55Gag protein in *Escherichia coli*. *Protein Expr Purif.* août 2014;100:10-8.
255. Lee BM, De Guzman RN, Turner BG, Tjandra N, Summers MF. Dynamical behavior of the HIV-1 nucleocapsid protein. *J Mol Biol.* juin 1998;279(3):633-49.
256. Mouhand A, Belfetmi A, Catala M, Larue V, Zargarian L, Brachet F, et al. Modulation of the HIV nucleocapsid dynamics finely tunes its RNA-binding properties during virion genesis. *Nucleic Acids Res.* 12 oct 2018;46(18):9699-710.
257. Karabiber F, McGinnis JL, Favorov OV, Weeks KM. QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA.* 1 janv 2013;19(1):63-73.
258. Sleiman D, Bernacchi S, Xavier Guerrero S, Brachet F, Larue V, Paillart JC, et al. Characterization of RNA binding and chaperoning activities of HIV-1 Vif protein: Importance of the C-terminal unstructured tail. *RNA Biol.* juill 2014;11(7):906-20.
259. Henriët S, Richer D, Bernacchi S, Decroly E, Vigne R, Ehresmann B, et al. Cooperative and Specific Binding of Vif to the 5' Region of HIV-1 Genomic RNA. *J Mol Biol.* nov 2005;354(1):55-72.
260. Cheng CY, Chou FC, Kladwang W, Tian S, Cordero P, Das R. Consistent global structures of complex RNA states through multidimensional chemical mapping. Nilsen TW, éditeur. *eLife.* 2015;4:e07600.
261. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *J Am Chem Soc.* 1 mars 2005;127(12):4223-31.
262. Mortimer SA, Weeks KM. Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat Protoc.* oct 2009;4(10):1413-21.
263. Grohman JK, Gorelick RJ, Lickwar CR, Lieb JD, Bower BD, Znosko BM, et al. A Guanosine-Centric Mechanism for RNA Chaperone Function. *Science.* 12 avr 2013;340(6129):190-5.
264. D'Souza AR, Jayaraman D, Long Z, Zeng J, Prestwood LJ, Chan C, et al. HIV-1 Packaging Visualised by In-Gel SHAPE. *Viruses.* 29 nov 2021;13(12):2389.
265. Yin J, Chen R, Liu C. Nucleic acid induced protein aggregation and its role in biology and pathology. *Front Biosci.* 2009;14(1):5084.
266. Lotz GP, Legleiter J. The role of amyloidogenic protein oligomerization in neurodegenerative disease. *J Mol Med.* juin 2013;91(6):653-64.
267. Nalavade R, Griesche N, Ryan DP, Hildebrand S, Krauß S. Mechanisms of RNA-induced toxicity in CAG repeat disorders. *Cell Death Dis.* août 2013;4(8):e752-e752.
268. Nikolaitchik OA, Liu S, Kitzrow JP, Liu Y, Rawson JMO, Shakya S, et al. Selective packaging of HIV-1 RNA genome is guided by the stability of 5' untranslated region polyA stem. *Proc Natl Acad Sci.* 14 déc 2021;118(50):e2114494118.
269. Sakuragi S, Kotani O, Yokoyama M, Shioda T, Sato H, Sakuragi J ichi. Identification of a Novel Cis-Acting Regulator of HIV-1 Genome Packaging. *Int J Mol Sci.* 26 mars 2021;22(7):3435.
270. Stephenson JD, Li H, Kenyon JC, Symmons M, Klenerman D, Lever AM. Three-Dimensional RNA Structure of the Major HIV-1 Packaging Signal Region. *Struct England*1993. 4 juin 2013;21(6):951-62.
271. Jones CP, Cantara WA, Olson ED, Musier-Forsyth K. Small-angle X-ray scattering-derived structure of the HIV-1 5' UTR reveals 3D tRNA mimicry. *Proc Natl Acad Sci.* 4

## BIBLIOGRAPHIE

- mars 2014;111(9):3395-400.
272. Sakuragi J ichi, Ode H, Sakuragi S, Shioda T, Sato H. A proposal for a new HIV-1 DLS structural model. *Nucleic Acids Res.* juin 2012;40(11):5012-22.
273. Osmer PS, Singh G, Boris-Lawrie K. A New Approach to 3D Modeling of Inhomogeneous Populations of Viral Regulatory RNA. *Viruses.* 29 sept 2020;12(10):1108.
274. Comandur R, Olson ED, Musier-Forsyth K. Conservation of tRNA mimicry in the 5'-untranslated region of distinct HIV-1 subtypes. *RNA.* déc 2017;23(12):1850-9.
275. Cech TR, Steitz JA. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell.* mars 2014;157(1):77-94.
276. Vicens Q, Kieft JS. Thoughts on how to think (and talk) about RNA structure. *Proc Natl Acad Sci.* 26 avr 2022;119(17):e2112677119.
277. Noller HF. RNA Structure: Reading the Ribosome. *Science.* 2 sept 2005;309(5740):1508-14.
278. Ganser LR, Kelly ML, Herschlag D, Al-Hashimi HM. The roles of structural dynamics in the cellular functions of RNAs. *Nat Rev Mol Cell Biol.* août 2019;20(8):474-89.
279. Solayman M, Litfin T, Singh J, Paliwal K, Zhou Y, Zhan J. Probing RNA structures and functions by solvent accessibility: an overview from experimental and computational perspectives. *Brief Bioinform.* 13 mai 2022;23(3):bbac112.
280. Yesselman JD, Denny SK, Bisaria N, Herschlag D, Greenleaf WJ, Das R. Sequence-dependent RNA helix conformational preferences predictably impact tertiary structure formation. *Proc Natl Acad Sci.* 20 août 2019;116(34):16847-55.
281. Tian S, Das R. RNA structure through multidimensional chemical mapping. *Q Rev Biophys.* 2016;49:e7.
282. Mailler E, Paillart J, Marquet R, Smyth RP, Vivet-Boudou V. The evolution of RNA structural probing methods: From gels to next-generation sequencing. *WIREs RNA* [Internet]. mars 2019 [cité 16 oct 2022];10(2). Disponible sur: <https://onlinelibrary.wiley.com/doi/10.1002/wrna.1518>
283. Sclavi B, Zaychikov E, Rogozina A, Walther F, Buckle M, Heumann H. Real-time characterization of intermediates in the pathway to open complex formation by *Escherichia coli* RNA polymerase at the T7A1 promoter. *Proc Natl Acad Sci.* 29 mars 2005;102(13):4706-11.
284. Sclavi B, Sullivan M, Chance MR, Brenowitz M, Woodson SA. RNA Folding at Millisecond Intervals by Synchrotron Hydroxyl Radical Footprinting. *Science.* 20 mars 1998;279(5358):1940-3.
285. Feng C, Chan D, Joseph J, Muuronen M, Coldren WH, Dai N, et al. Light-activated chemical probing of nucleobase solvent accessibility inside cells. *Nat Chem Biol.* 1 mars 2018;14(3):276-83.
286. Kielpinski LJ, Vinther J. Massive parallel-sequencing-based hydroxyl radical probing of RNA accessibility. *Nucleic Acids Res.* 1 avr 2014;42(8):e70-e70.
287. Ehresmann C, Baudin F, Mougél M, Romby P, Ebel JP, Ehresmann B. Probing the structure of RNAs in solution. *Nucleic Acids Res.* 1987;15(22):9109-28.
288. Mortimer SA, Weeks KM. A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry. *J Am Chem Soc.* avr 2007;129(14):4144-5.
289. Strobel EJ, Yu AM, Lucks JB. High-throughput determination of RNA structures. *Nat Rev Genet.* oct 2018;19(10):615-34.

## BIBLIOGRAPHIE

290. Zarudnaya Margarita, Potyahaylo Andriy, Kolomiets Iryna, Gorb Leonid. Genome sequence analysis suggests coevolution of the DIS, SD, and Psi hairpins in HIV-1 genomes. *Virus Res.* nov 2022;321:198910.
291. Sakuragi S, Yokoyama M, Shioda T, Sato H, Sakuragi J. SL1 revisited: functional analysis of the structure and conformation of HIV-1 genome RNA. *Retrovirology* [Internet]. 11 nov 2016 [cité 17 janv 2019];13. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5106843/>
292. Kappel K, Zhang K, Su Z, Watkins AM, Kladwang W, Li S, et al. Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat Methods.* juill 2020;17(7):699-707.
293. De Laat J, Gallard H. Catalytic Decomposition of Hydrogen Peroxide by Fe(III) in Homogeneous Aqueous Solution: Mechanism and Kinetic Modeling. *Environ Sci Technol.* 1 août 1999;33(16):2726-32.
294. Das R, Kudaravalli M, Jonikas M, Laederach A, Fong R, Schwans JP, et al. Structural inference of native and partially folded RNA by high-throughput contact mapping. *Proc Natl Acad Sci U S A.* 18 mars 2008;105(11):4144-9.
295. Seetin MG, Kladwang W, Bida JP, Das R. Massively Parallel RNA Chemical Mapping with a Reduced Bias MAP-Seq Protocol. In: Waldsich C, éditeur. *RNA Folding* [Internet]. Totowa, NJ: Humana Press; 2014 [cité 16 oct 2022]. p. 95-117. (Methods in Molecular Biology; vol. 1086). Disponible sur: [http://link.springer.com/10.1007/978-1-62703-667-2\\_6](http://link.springer.com/10.1007/978-1-62703-667-2_6)
296. Watkins AM, Rangan R, Das R. FARFAR2: Improved De Novo Rosetta Prediction of Complex Global RNA Folds. *Structure.* août 2020;28(8):963-976.e6.
297. Lyskov S, Chou FC, Conchúir SÓ, Der BS, Drew K, Kuroda D, et al. Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE). Uversky VN, éditeur. *PLoS ONE.* 22 mai 2013;8(5):e63906.
298. Zhuang F, Fuchs RT, Robb GB. Small RNA Expression Profiling by High-Throughput Sequencing: Implications of Enzymatic Manipulation. *J Nucleic Acids.* 2012;2012:1-15.
299. van Dijk EL, Jaszczyszyn Y, Thermes C. Library preparation methods for next-generation sequencing: Tone down the bias. *Exp Cell Res.* mars 2014;322(1):12-20.
300. Mortimer SA, Trapnell C, Aviran S, Pachter L, Lucks JB. SHAPE-Seq: High-Throughput RNA Structure Analysis. *Curr Protoc Chem Biol.* déc 2012;4(4):275-97.
301. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc.* août 2012;7(8):1534-50.
302. Padilla R, Sousa R. A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Res.* 15 déc 2002;30(24):138e-138.
303. Milisavljevič N, Perlíková P, Pohl R, Hocek M. Enzymatic synthesis of base-modified RNA by T7 RNA polymerase. A systematic study and comparison of 5-substituted pyrimidine and 7-substituted 7-deazapurine nucleoside triphosphates as substrates. *Org Biomol Chem.* 2018;16(32):5800-7.
304. Liu Q, Ke Y, Kan Y, Tang X, Li X, He Y, et al. Compatibility and Fidelity of Mirror-Image Thymidine in Transcription Events by T7 RNA Polymerase. *Mol Ther - Nucleic Acids.* sept 2020;21:604-13.
305. Enroth CH, Fehler AO, Poulsen LD, Vinther J. Excess primer degradation by Exo I improves the preparation of 3' cDNA ligation-based sequencing libraries.

## BIBLIOGRAPHIE

- BioTechniques. sept 2019;67(3):110-6.
306. Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, et al. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc Natl Acad Sci*. 5 juill 2011;108(27):11063-8.
307. Hou Z, Jiang P, Swanson SA, Elwell AL, Nguyen BKS, Bolin JM, et al. A cost-effective RNA sequencing protocol for large-scale gene expression studies. *Sci Rep*. août 2015;5(1):9570.
308. Poulsen LD, Vinther J. RNA-Seq for Bacterial Gene Expression. *Curr Protoc Nucleic Acid Chem*. juin 2018;73(1):e55.
309. Ciesiołka J. Metal Ion - Induced Cleavages in Probing of RNA Structure. In: Barciszewski J, Clark BFC, éditeurs. *RNA Biochemistry and Biotechnology* [Internet]. Dordrecht: Springer Netherlands; 1999 [cité 16 oct 2022]. p. 111-21. Disponible sur: [http://link.springer.com/10.1007/978-94-011-4485-8\\_7](http://link.springer.com/10.1007/978-94-011-4485-8_7)
310. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. avr 2012;9(4):357-9.
311. Ding P, Kharytonchuk S, Kuo N, Cannistraci E, Flores H, Chaudhary R, et al. 5'-Cap sequestration is an essential determinant of HIV-1 genome packaging. *Proc Natl Acad Sci*. 14 sept 2021;118(37):e2112475118.
312. Keane SC, Van V, Frank HM, Sciandra CA, McCowin S, Santos J, et al. NMR detection of intermolecular interaction sites in the dimeric 5'-leader of the HIV-1 genome. *Proc Natl Acad Sci U S A*. 15 nov 2016;113(46):13033-8.
313. Levin JG, Mitra M, Mascarenhas A, Musier-Forsyth K. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol*. nov 2010;7(6):754-74.
314. Yu B, Pettitt BM, Iwahara J. Dynamics of Ionic Interactions at Protein–Nucleic Acid Interfaces. *Acc Chem Res*. 15 sept 2020;53(9):1802-10.
315. Alfadhli A, Barklis E. The roles of lipids and nucleic acids in HIV-1 assembly. *Front Microbiol* [Internet]. 28 mai 2014 [cité 18 oct 2022];5. Disponible sur: <http://journal.frontiersin.org/article/10.3389/fmicb.2014.00253/abstract>
316. Darlix JL, de Rocquigny H, Mauffret O, Mély Y. Retrospective on the all-in-one retroviral nucleocapsid protein. *Virus Res*. nov 2014;193:2-15.
317. Matsuda Z, Iga M, Miyauchi K, Komano J, Morishita K, Okayama A, et al. In Vitro Translation to Study HIV Protease Activity. In: Grandi G, éditeur. *In Vitro Transcription and Translation Protocols* [Internet]. Totowa, NJ: Humana Press; 2007 [cité 16 oct 2022]. p. 135-49. Disponible sur: [http://link.springer.com/10.1007/978-1-59745-388-2\\_7](http://link.springer.com/10.1007/978-1-59745-388-2_7)
318. Hanne J, Göttfert F, Schimer J, Anders-Össwein M, Konvalinka J, Engelhardt J, et al. Stimulated Emission Depletion Nanoscopy Reveals Time-Course of Human Immunodeficiency Virus Proteolytic Maturation. *ACS Nano*. 27 sept 2016;10(9):8215-22.
319. Brule F, Marquet R, Rong L, Wainberg MA, Roques BP, Le Grice SFJ, et al. Structural and functional properties of the HIV-1 RNA 3' Lys primer complex annealed by the nucleocapsid protein: Comparison with the heat-annealed complex. *RNA*. janv 2002;8(1):8-15.
320. Kalman RE. A New Approach to Linear Filtering and Prediction Problems. *J Basic Eng*. 1 mars 1960;82(1):35-45.
321. Vaziri S, Koehl P, Aviran S. Extracting information from RNA SHAPE data: Kalman filtering approach. Barash D, éditeur. *PLOS ONE*. 21 nov 2018;13(11):e0207029.

## BIBLIOGRAPHIE

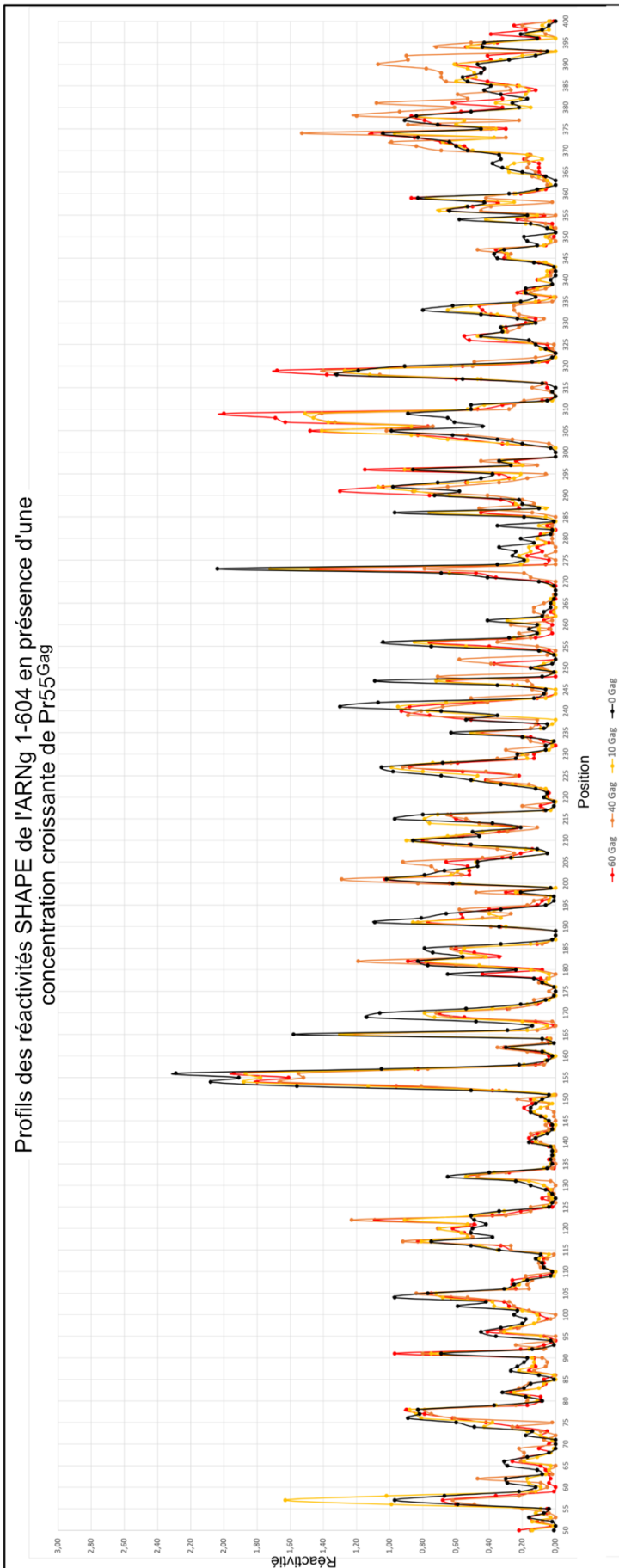
322. Altman RB. A probabilistic approach to determining biological structure: integrating uncertain data sources. *Int J Hum-Comput Stud.* juin 1995;42(6):593-616.
323. Chameettachal A, Vivet-Boudou V, Pitchai FNN, Pillai VN, Ali LM, Krishnan A, et al. A purine loop and the primer binding site are critical for the selective encapsidation of mouse mammary tumor virus genomic RNA by Pr77Gag. *Nucleic Acids Res.* 7 mai 2021;49(8):4668-88.
324. Pearce R, Omenn GS, Zhang Y. *De Novo* RNA Tertiary Structure Prediction at Atomic Resolution Using Geometric Potentials from Deep Learning [Internet]. *Bioinformatics*; 2022 mai [cité 18 oct 2022]. Disponible sur: <http://biorxiv.org/lookup/doi/10.1101/2022.05.15.491755>





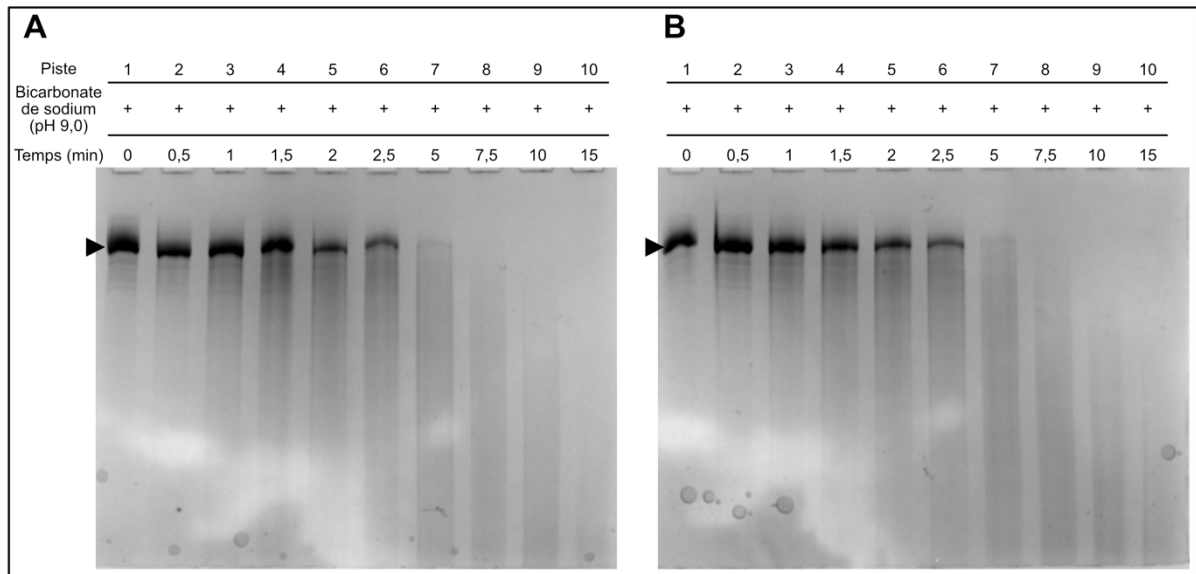


# Annexes



**Annexe 1** : Profils des réactivités SHAPE associées à l'ARNg 1-604 en présence d'une concentration croissante de Pr55<sup>Gag</sup>. Le graphique correspond à trois expériences indépendantes (n = 3) réalisées à l'aide de 2 amorces marquées 5' Vic, en présence de 0, 10, 40 et 60 Gag par ARNg. Les données obtenues en présence de 5 et 20 Gag ont été exclues de l'analyse à cause de la variabilité entre les réplicas (au moins un coefficient de corrélation inférieur à 0,65).

## ANNEXES



**Annexe 2** : Duplicata d'expériences de clivage de l'ARNg 1-604 (triangle noir) par traitement au bicarbonate de sodium à 90°C. Pour chaque condition, 0,8 µg d'ARNg 1-604 ont été mis en présence de 50 mM de bicarbonate de sodium et incubés à 90°C entre 0 et 15 min. Après 5 min d'incubation à 90°C, la quasi-totalité de l'ARNg 1-604 est clivé au moins une fois.

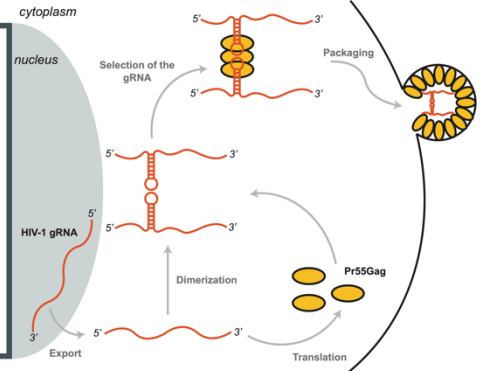
# The 3D structure of the HIV-1 genomic RNA 5' untranslated region during the selection and packaging events



Gilmer O., Mailler E., Paillart J.-C., Marquet R. & Vivet-Boudou V.  
 Architecture et Réactivité de l'ARN - UPR 9002 CNRS, Université de Strasbourg, IBMC, Strasbourg, France

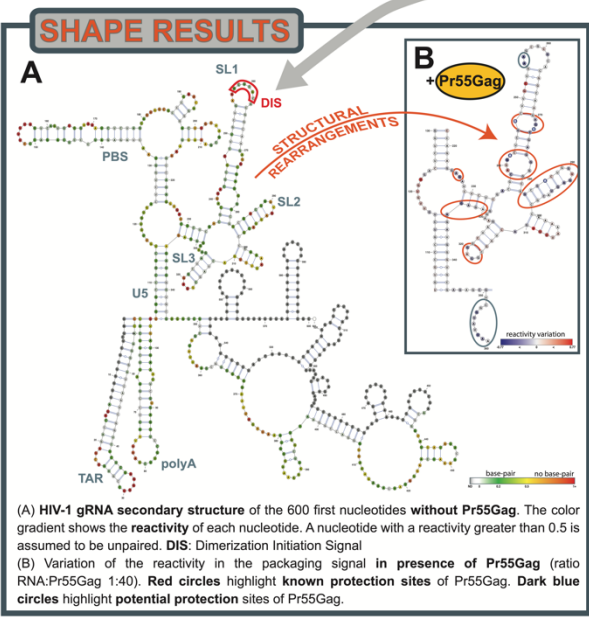
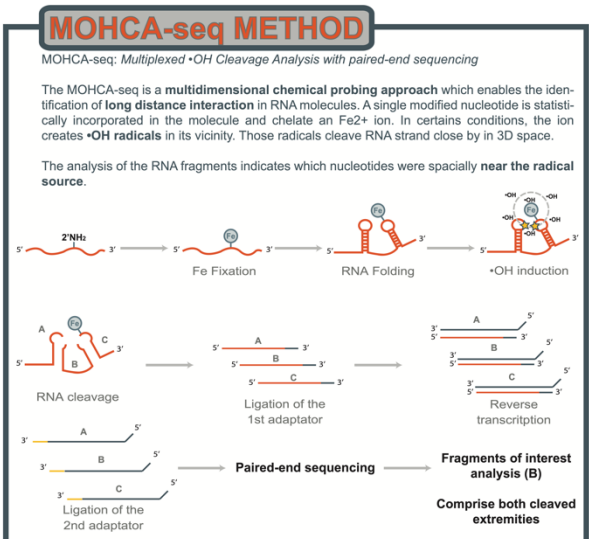
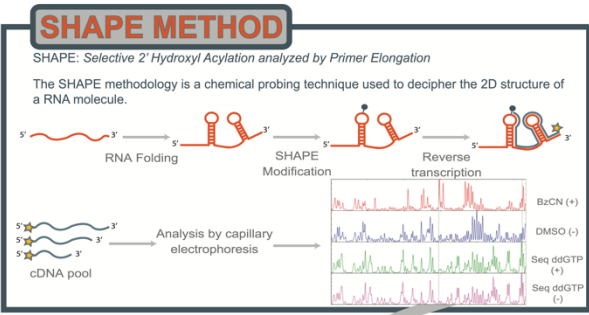


- BACKGROUND**
- Human Immunodeficiency Virus type-1 (HIV-1): **dimeric** genomic RNA (gRNA) of ~9.2 Kb
  - The 600 first nucleotides carry **many regulatory elements** involved in most steps of the viral cycle (dimerization, selection, packaging...)
  - HIV-1 gRNA is required for the synthesis of the **viral protein precursor Pr55Gag**
  - HIV-1 gRNA is specifically selected in the cytoplasm among all cellular RNAs by **Pr55Gag** which has a **chaperon activity**
  - **Selection and packaging** are two intrinsically related events



**QUESTIONS**

Does Pr55Gag trigger a structural rearrangement of HIV-1 gRNA?  
 Is this rearrangement involved in the regulation of the selection and packaging steps?



**CONCLUSIONS**

- Pr55Gag is able to bind HIV-1 gRNA on multiple sites and induces structural rearrangements on the 600 first nucleotides
- Long Distance Interaction in HIV-1 gRNA will be highlighted using the MOHCA-seq approach

**REFERENCES**

Bernacchi *et al.*, (2017). HIV-1 Pr55Gag binds genomic and spliced RNAs with different affinity and stoichiometry. *RNA Biology*. 14:1, 90-103.


Cheng *et al.*, (2015). Consistent global structures of complex RNA states through multidimensional chemical mapping. *eLife* 4 : e07600.

Abd El-Wahab *et al.*, (2014). Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat. Commun.* 5 : 4304.


Wilkinson *et al.*, (2008). High-throughput SHAPE analysis reveals structures in HIV-1 Genomic RNA strongly conserved across distinct biological states. *PLoS Biol* 6(4): e96.


Annexe 3 : Poster présenté aux Journées de l'École Doctorale de Strasbourg en 2019.

# VIH - 1 : vers un modèle de structure 3D de l'ARN génomique



Centre de biologie moléculaire et cellulaire



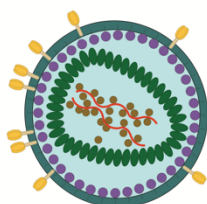


Université de Strasbourg

Gilmer O., Peterl S., Mailler E., Paillart J.-C., Marquet R. & Vivet-Boudou V.  
Architecture et Réactivité de l'ARN - UPR 9002 CNRS, Université de Strasbourg, IBMC, Strasbourg, France

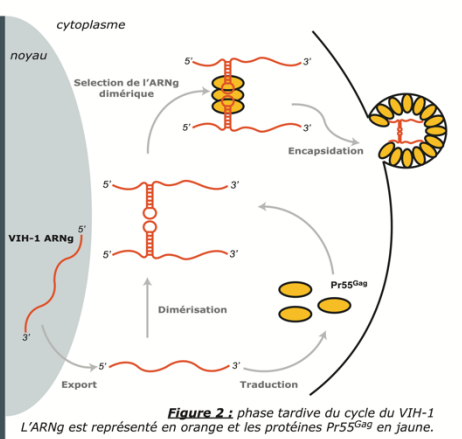
### INTRODUCTION

- Virus de l'Immunodéficience Humaine de type 1 (VIH-1) : rétrovirus humain au génome à **ARN dimérique** de 9,2 kb
- ARN : biomolécule constituée d'acides ribonucléiques et impliquée, entre autre, dans la synthèse protéique. Sa **structure** est importante à sa **fonction**



- Plusieurs **éléments de régulation** impliqués dans le cycle viral sont situés en 5' de l'ARN génomique (ARNg)
- l'ARNg sert de matrice à la synthèse de la protéine **Pr55<sup>Gag</sup>**, nécessaire aux étapes de **sélection** et d'**encapsidation**
- La **sélection** et l'**encapsidation** sont des événements intrinsèquement liés

**Figure 1 :** Particule mature du VIH-1.



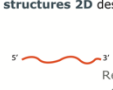
## OBJECTIFS

**La protéine Pr55<sup>Gag</sup> induit-elle un réarrangement structural de l'ARNg du VIH-1? Ce réarrangement est-il impliqué dans les mécanismes de sélection et d'encapsidation?**


### METHODE

L'approche SHAPE est une technique de cartographie chimique en solution qui détermine la réactivité de chaque nucléotide. Cette réactivité, exprimée en unité arbitraire, désigne le caractère apparié ou non de chacun des nucléotides d'un ARN donné. Lorsque la réactivité est basse (<0,5), le nucléotide est apparié. Les informations recueillies servent à modéliser des structures 2D des molécules d'ARN.


Replètement de l'ARN



Modification SHAPE



Transcription inverse

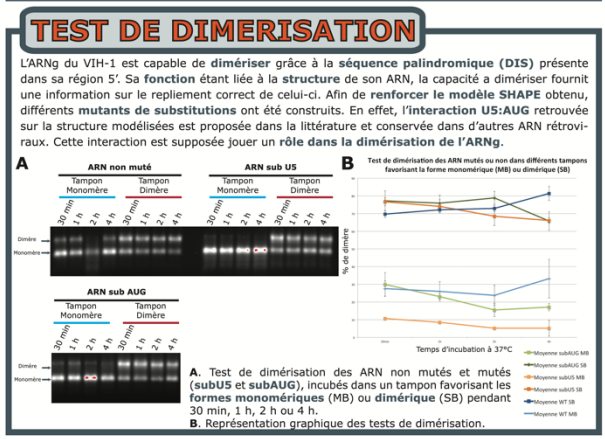
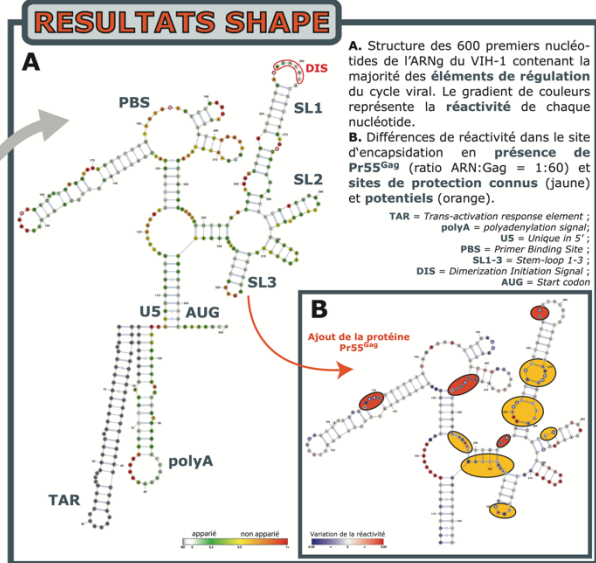


Librairie d'ADNc → Analyse par électrophorèse capillaire → NMA (+), DMSO (-), Seq ddGTP (+), Seq ddGTP (-)

L'expérience est réalisée en présence et en absence de protéines Pr55<sup>Gag</sup> pour élucider l'impact de la protéine Pr55<sup>Gag</sup> sur la structure de l'ARN.

Pr55<sup>Gag</sup> protège de la modification chimique, ce qui induit une diminution de la réactivité. Cependant, le nucléotide n'est pas apparié.

SHAPE : Selective 2'Hydroxyl Acylation analysed by Primer Extension



### CONCLUSIONS ET PERSPECTIVES

- La méthodologie SHAPE a permis de mettre en évidence les sites de fixation de Pr55<sup>Gag</sup> sur la région 5' de l'ARNg du VIH-1, ainsi que les réarrangements structuraux induits par cette fixation.
- Les tests de dimérisation laissent supposer que l'interaction U5:AUG, si elle existe, est différente de celle décrite dans la littérature, tout comme son impact sur la dimérisation de l'ARNg.
- Ces résultats seront vérifiés en appliquant l'approche de MOHCA-seq, une méthode de cartographie chimique multidimensionnelle.

### REFERENCES

- Bernacchi et al., (2017). HIV-1 Pr55Gag binds genomic and spliced RNAs with different affinity and stoichiometry. RNA Biology. 14:1, 90-103.
- Cheng et al., (2015). Consistent global structures of complex RNA states through multidimensional chemical mapping. eLife 4 : e07600.
- Abd El-Wahab et al., (2014). Specific recognition of the HIV-1 genomic RNA by the Gag precursor. Nat. Commun. 5 : 4304.
- Wilkinson et al., (2008). High-throughput SHAPE analysis reveals structures in HIV-1 Genomic RNA strongly conserved across distinct biological states. PLoS Biol 6(4): e166.

P103

# Structures de l'ARN génomique du VIH-1 et réarrangements induits par Gag

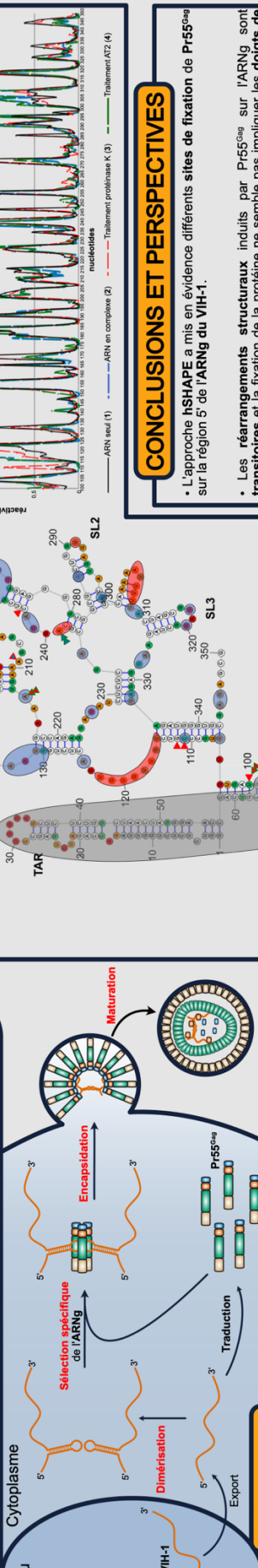


Orian Gilmer<sup>1</sup>, Elodie Mailler<sup>1</sup>, Carine Tisné<sup>2</sup>, Johnson Mak<sup>3</sup>, Jean-Christophe Paillart<sup>1</sup>, Redmond Smyth<sup>4</sup>, Roland Marquet<sup>1</sup> & Valérie Vivet-Boudou<sup>1</sup> CIFS  
 1. Architecture et Réactivité de l'ARN - CNRS, Université de Strasbourg, IBMC, Strasbourg, France 2. Institut de biologie physico-chimique (IBPC) - CNRS, Paris, France 3. Institute for Glycomics, Griffith University, Brisbane, Queensland, Australie 4. Helmholtz Institute for RNA-based Infection Research, Würzburg, Allemagne  
 contact : o.gilmer@unistra.fr

## Le précurseur protéique Pr55<sup>Gag</sup> induit-il des réarrangements structuraux de l'ARN génomique du VIH-1 ?

### INTRODUCTION

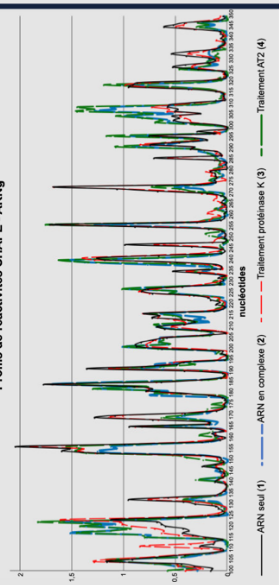
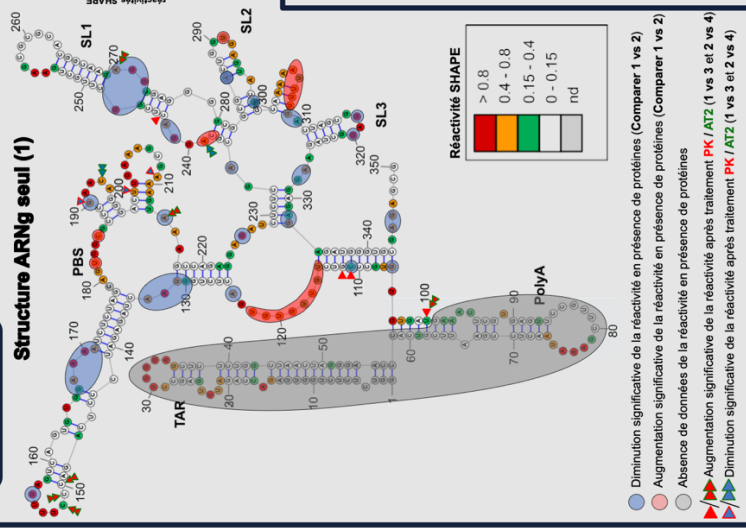
- Virus de l'immunodéficience Humaine (type 1) : génome à ARN de polarité positive dimérique
- Les 600 premiers nucléotides de la séquence contiennent de nombreux signaux impliqués dans différentes étapes du cycle réplcatif
- L'ARN génomique (ARNg) permet la **synthèse** des précurseurs protéiques viraux Pr55<sup>Gag</sup> et Pr160<sup>Gag-Pol</sup>
- L'ARNg est spécifiquement sélectionné sous forme de dimère par Pr55<sup>Gag</sup> pour être encapsidé



### METHODE

- Quatre conditions expérimentales :
1. ARN seul : Modification
  2. Empreinte des protéines : Formation du complexe - Modification - Protéinase K
  3. Activité chaperon d'ARN : Formation du complexe - Protéinase K - Modification
  4. Implication des doigts de zinc du domaine NC : Formation du complexe - ATZ - Modification - Protéinase K
- Détection des modifications par reverse transcription SUD
- ATZ = sélecteur de zinc

### RESULTATS



### CONCLUSIONS ET PERSPECTIVES

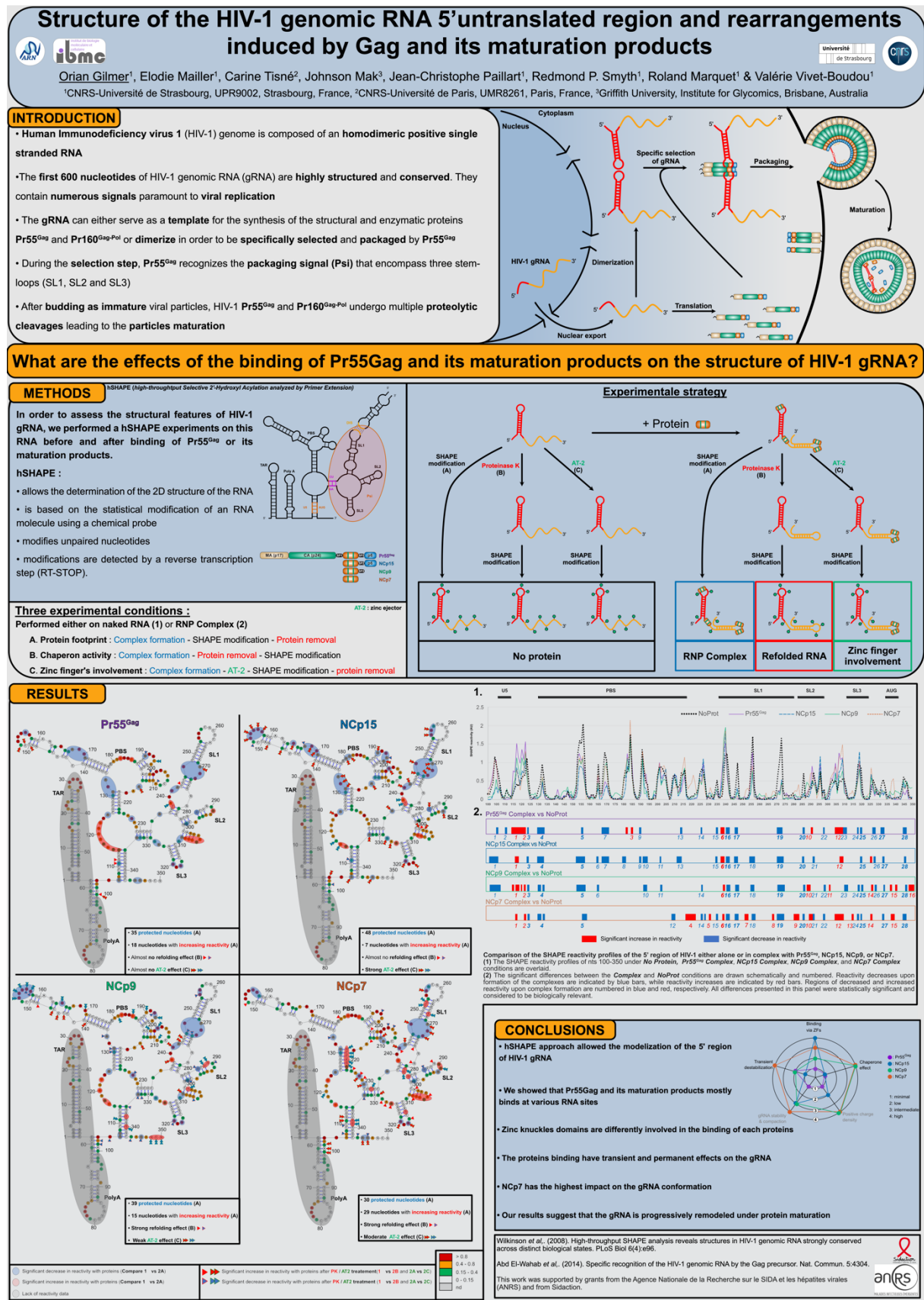
- L'approche iSHAPE a mis en évidence différents sites de fixation de Pr55<sup>Gag</sup> sur la région 5' de l'ARNg du VIH-1.
- Les réarrangements structuraux induits par Pr55<sup>Gag</sup> sur l'ARNg sont transitoires et la fixation de la protéine ne semble pas impliquer les doigts de zinc présents dans le domaine de la nucléocapside.
- Pr55<sup>Gag</sup> permet la sélection spécifique de l'ARNg du VIH-1 sans induire de changements permanents de sa structure.
- Les effets des protéines de la cascade de maturation de Pr55<sup>Gag</sup> sur la structure de l'ARNg sont actuellement étudiés pour vérifier leur implication dans la maturation génomique

Wilkinson et al., (2008). High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS Biol 6(4):e96.  
 Abd El-Wahab et al., (2014). Specific recognition of the HIV-1 genomic RNA by the Gag precursor. Nat. Commun. 5:4304.  
 Ce travail est soutenu par l'Agence Nationale de Recherche sur le SIDA et les hépatites virales (ANRS) et SIDACTION.



Annexe 5 : Poster présenté aux Journées Francophones de Virologie en 2021.





Annexe 6 : Poster présenté aux Journées de l'École Doctorale de Strasbourg et au congrès de Cold Spring Harbor Laboratory Retrovirus en 2021.

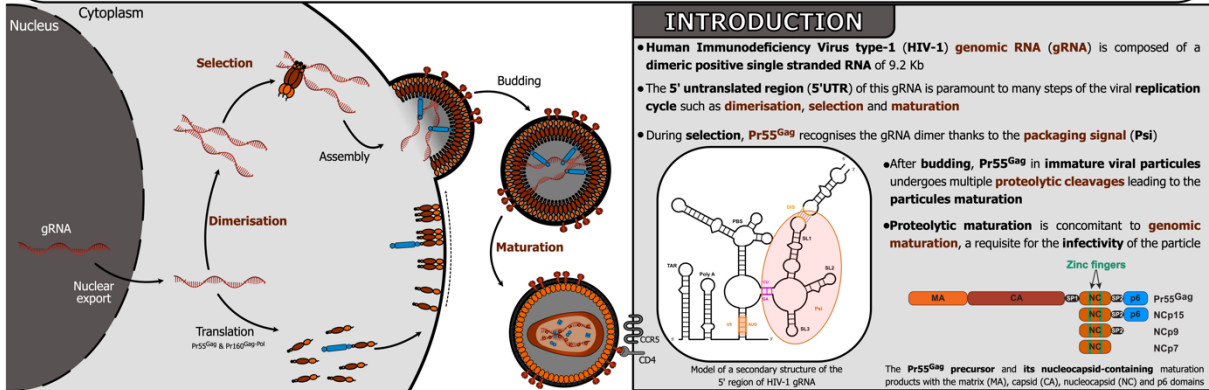
# Toward the 3D structure of the HIV-1 genomic RNA 5' untranslated region during replication late phase events



Gilmer O.<sup>1</sup>, Mailler E.<sup>1</sup>, Paillart J.-C.<sup>1</sup>, Tisné C.<sup>2</sup>, Smyth RP.<sup>3</sup>, Marquet R.<sup>1</sup> & Vivet-Boudou V.<sup>1</sup>

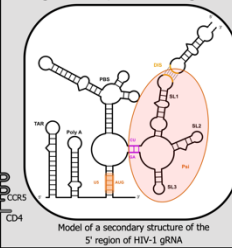


<sup>1</sup>CNRS-Université de Strasbourg, UPR9002, Strasbourg, France, <sup>2</sup>CNRS-Université de Paris, UMR8261, Paris, France, <sup>3</sup>Helmholtz Institute for RNA-based Infection Research (HIRI), Würzburg, Germany

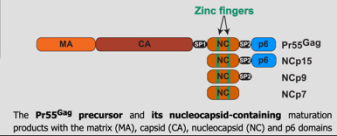


## INTRODUCTION

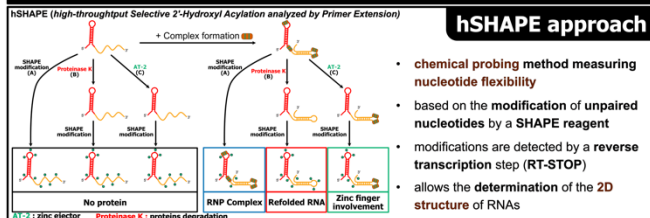
- Human Immunodeficiency Virus type-1 (HIV-1) genomic RNA (gRNA) is composed of a dimeric positive single stranded RNA of 9.2 Kb
- The 5' untranslated region (5'UTR) of this gRNA is paramount to many steps of the viral replication cycle such as **dimerisation**, **selection** and **maturation**
- During **selection**, Pr55<sup>Gag</sup> recognises the gRNA dimer thanks to the **packaging signal (Psi)**



- After budding, Pr55<sup>Gag</sup> in immature viral particles undergoes multiple **proteolytic cleavages** leading to the **particles maturation**
- **Proteolytic maturation** is concomitant to **genomic maturation**, a requisite for the **infectivity** of the particle

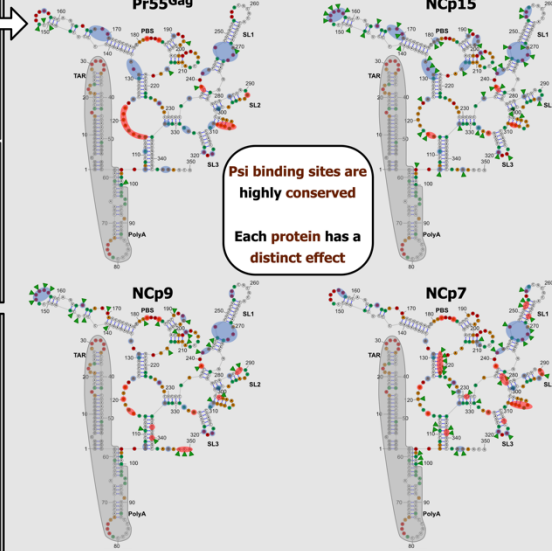


What are the effects of the binding of Pr55<sup>Gag</sup> and its nucleocapsid-containing maturation products on HIV-1 gRNA structure? What is the involvement of long-range interactions in the structure of HIV-1 gRNA and in the late steps of the replication cycle?

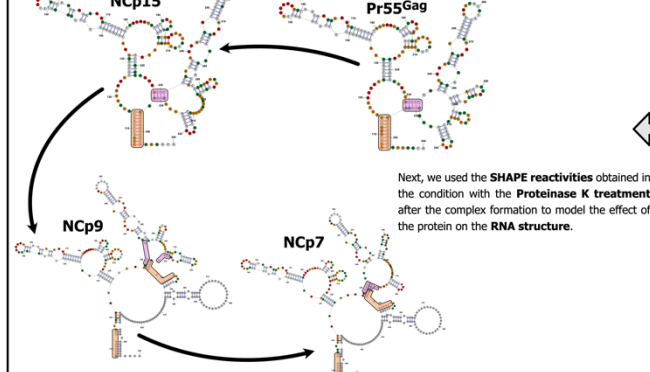


## hSHAPE results: proteins effects

Here, we used the reactivities obtained for the naked RNA to model the secondary structure of the 5' region of HIV-1 gRNA and map the effects of each NC-containing protein.



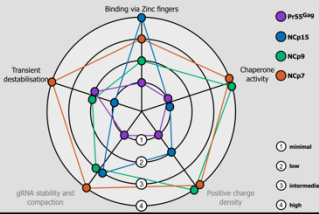
## hSHAPE results: refolding



The structural rearrangements induced by protein binding are more pronounced with NCp9 and NCp7, indicating a more important chaperone activity of these proteins. A reorganisation of the long-range interactions is also observed with the destabilisation of the initial US:AUG and CU:GA interactions.

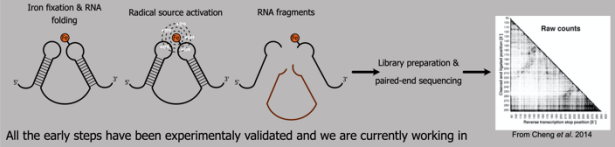
## CONCLUSIONS

- hSHAPE approach allowed the modelisation of the 5' region of HIV-1 gRNA
- Pr55<sup>Gag</sup> and its NC-containing maturation products mostly binds distinct RNA sites
- Zinc Fingers domains are differently involved in the binding of each proteins
- The binding of proteins induces transient and permanent effects on the gRNA
- NCp7 has the highest impact on the gRNA conformation
- Our results suggest that the gRNA is progressively remodeled under protein maturation



## PERSPECTIVES

In order to investigate the long-range interactions folding HIV-1 gRNA 5' region, we optimised a multidimensional chemical mapping approach called MOHCA-seq (Multiplexed •OH Cleavage Analysis with paired-end sequencing). This method should provide us information about HIV-1 gRNA 3D structure.



- All the early steps have been experimentally validated and we are currently working in resolving issues regarding the library preparation.
1. Wilkinson et al., (2008). High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* 6(4):e166.
  2. Abd El-Wahab et al., (2014). Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat. Commun.* 5:4304.
  3. Cheng et al., (2015). Consistent global structures of complex RNA states through multidimensional chemical mapping. *Elife*. 2015 Jul 21; 4:e10095.
  4. Baraschi et al., (2017). HIV-1 Pr55Gag binds genomic and spliced RNAs with different affinity and stoichiometry. *RNA Biology*. 14(1), 90-103.
  5. Gilmer et al., (2022). Structural maturation of the HIV-1 RNA 5' untranslated region by Pr55Gag and its maturation products. *RNA Biol*. 2022 Jan 19(1):191-205.
- This work was supported by grants from the Agence Nationale de la Recherche sur le SIDA et les Hépatites Virales (ANRS) and from Sidaction.



# Étude de la structure de la région 5' terminale de l'ARN génomique du virus de l'immunodéficience humaine de type 1 et des réarrangements induits par Gag et ses produits de maturation

## Résumé

La région 5' terminale de l'ARN génomique (ARNg) du virus de l'immunodéficience humaine de type 1 (VIH-1) régule de nombreuses étapes du cycle viral via les structures complexes qu'elle adopte. Lors des étapes tardives, un dimère d'ARNg est spécifiquement sélectionné par des protéines Pr55<sup>Gag</sup> pour être encapsidé dans les particules virales. Ces particules subissent une maturation protéolytique et génomique qui permet l'acquisition de l'infectivité virale. Cependant, les structures tridimensionnelles adoptées par l'ARNg et le déroulement de la maturation au niveau génomique restent majoritairement inexplorées à ce jour.

Lors de ma thèse, j'ai étudié l'impact de différentes protéines, impliquées dans l'étape de maturation et contenant le domaine de nucléocapside, sur la structure de la région 5' de l'ARNg. Cette étude a été réalisée par une cartographie chimique hSHAPE, *in vitro* et elle m'a permis d'identifier des réarrangements structuraux induits par la fixation des protéines. J'ai également développé une approche de cartographie chimique multidimensionnelle, le MOHCA-seq, afin de mettre en évidence les repliements à longue distance structurant l'ARNg du VIH-1. Cette approche a requis de nombreuses optimisations qui m'ont finalement permis de l'appliquer à l'étude de la région 5' terminale de l'ARNg du VIH-1.

Mots-clefs : VIH-1 ; ARNg ; Pr55<sup>Gag</sup> ; cartographie chimique ; structure ARN ; interaction ARN-protéine

## Abstract

The 5' terminal region of the human immunodeficiency virus type 1 (HIV-1) genomic RNA (gRNA) regulates many steps of the viral cycle via the complex structures it adopts. During the late stages, a gRNA dimer is specifically selected by Pr55<sup>Gag</sup> proteins to be packaged into nascent viral particles. These particles undergo proteolytic and genomic maturation that allows the acquisition of viral infectivity. However, the 3D structures adopted by the gRNA and the maturation process at the genomic level remain largely unexplored to date.

During my thesis, I was interested in characterizing the impact of various nucleocapsid containing proteins involved in the maturation step, on the structure of the 5' region of the gRNA. This study was carried out by a chemical mapping approach of hSHAPE, *in vitro*, and it allowed me to identify several structural rearrangements induced by protein binding. I also developed a multidimensional chemical mapping approach, MOHCA-seq, in order to identify long distance interactions structuring HIV-1 gRNA. This approach required many optimizations that finally allowed me to apply it to the study of the 5' terminal region of HIV-1 gRNA.

Keywords: HIV-1 ; gRNA ; Pr55<sup>Gag</sup> ; chemical probing ; RNA structure ; RNA-protein interaction