



**HAL**  
open science

# Etude biochimique de l'interaction ARN-protéine lors de l'infection virale par le VIH-1 et de la réponse antivirale cellulaire

Camille Louvat

## ► To cite this version:

Camille Louvat. Etude biochimique de l'interaction ARN-protéine lors de l'infection virale par le VIH-1 et de la réponse antivirale cellulaire. Biologie cellulaire. Université Claude Bernard - Lyon I, 2023. Français. NNT : 2023LYO10087 . tel-04514864

**HAL Id: tel-04514864**

**<https://theses.hal.science/tel-04514864v1>**

Submitted on 21 Mar 2024

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

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



# THESE de DOCTORAT DE L'UNIVERSITE CLAUDE BERNARD LYON 1

Ecole Doctorale N°205  
Ecole Doctorale Interdisciplinaire Sciences Santé

**Discipline** : Aspects moléculaires et cellulaires de la biologie

Soutenue publiquement/à huis clos le 20/06/2023, par :  
**Camille Raïcha LOUVAT**

---

## Étude biochimique de l'interaction ARN–protéine lors de l'infection virale par le VIH-1 et de la réponse antivirale cellulaire

---

Devant le jury composé de :

COUREUX, Pierre-Damien	PR, Université Lyon 1, UMR 5086, MMSB	Président
IMBERT, Isabelle	PR, Aix-Marseille Université, UMR 7255, LISM	Rapporteuse
PARISSI, Vincent	DR, CNRS Bordeaux, UMR 5234, MFP	Rapporteur
TISNÉ, Carine	DR, CNRS Paris, UMR 8261, EGM	Examinatrice
GOUET, Patrice	PR, Université de Lyon 1, UMR 5086, MMSB	Directeur de thèse
FIORINI, Francesca	CR, CNRS Lyon, UMR 5086, MMSB	Co-directrice de thèse



## Remerciements

Comme dit le proverbe : « Seul on va plus vite, ensemble on va plus loin », ainsi je souhaite remercier un certain nombre de personnes.

Tout d'abord, je remercie les membres de mon jury – Isabelle Imbert, Vincent Parissi, Pierre-Damien Coureux et Carine Tisé – qui ont accepté d'évaluer mes travaux de thèse. Je tiens également à remercier les membres de mon comité de suivi individuel, Cyril Bourgeois et Stéphanie Ravaud, qui m'ont accompagné sur le plan scientifique et professionnel au cours de mes trois années de thèse. Merci à mes professeurs de l'Université Claude Bernard Lyon 1 qui m'ont donné envie de découvrir le métier d'enseignant-chercheur.

Je remercie mes collaborateurs et les membres de la *Protein Science Facility* pour les équipements mis à disposition, l'aide apportée dans les réflexions scientifiques et les manips. Merci à Charlotte pour les recettes de cuisine en supplément, ça fait toujours plaisir !)

Je souhaite remercier tous mes collègues de l'équipe pédagogique de biochimie à l'Université Claude Bernard Lyon 1 pour leurs conseils et leur bienveillance. Je remercie également tous les doctorants ACE et ATER avec qui j'ai partagé les repas du midi, des astuces et de beaux moments. C'était une super expérience le monitorat et je suis super contente de l'avoir vécue à vos côtés !

Un très très grand merci à Francesca, qui m'a encadré pendant ma thèse et qui m'a appris énormément de choses. Merci Francesca pour tout ce que tu m'as apporté d'un point de vue scientifique, les idées de génie, le soutien quand ISG20 nous fait des misères, les grigris porte-bonheur pour les manips, les gâteaux pour garder la pêche au labo et tous les bons moments passés à rigoler. Merci aussi pour ton écoute et ta positivité. Je suis vraiment contente d'avoir pu apprendre et m'épanouir à tes côtés (et de pouvoir encore continuer à le faire pour quelques mois). Et comme tu dis : « dans une thèse, ce qui est important ce n'est pas la ligne d'arrivée, mais le trajet ». Merci !

Je souhaite remercier Patrice de m'avoir accueilli dans son équipe et de m'avoir encadré pendant ma thèse. Merci Patrice pour tout le temps que tu m'as accordé, ton soutien et les conseils précieux que tu m'as donnés. Et surtout un grand merci pour m'avoir accueilli la première fois dans ton équipe il y a six ans et m'avoir fait découvrir ce que c'était que la recherche en laboratoire.

Je remercie très chaleureusement Cecilia. Merci Ceci pour ton aide durant ces trois ans et demi. Merci pour tes avis, ton soutien tout au long de mon stage de M2 et ma thèse, tous les mots rassurants et encourageants que tu m'as offerts ; merci pour les repas qu'on a partagés et tous ces moments géniaux en dehors du labo. Grâce à tes enseignements, ta petite padawan est devenue une Jedi. Le côté obscur de la force n'a pas triomphé cette fois encore !

*Anche se non bevo ancora il caffè, ho amato tutti i dolci italiani che mi hai fatto conoscere. Grazie mille per tutti i bei momenti Francesca e Cecilia!*

Un grand merci à mes prédécesseurs Mathieu et Léa. Grâce à vous deux, j'ai pu écouter des chansons de Disney en faisant des purifs, citer des répliques du Seigneur des Anneaux, chanter des comptines le vendredi après-midi, et pleins d'autres folies ! Merci :) Je remercie aussi Johan, Chloé, Lysa, Mohammed, Yacine et Rémi pour tous les moments partagés au labo et en-dehors. Merci à Christophe, Frédéric, Sébastien, Lionel, Xavier et Nushin pour les discussions et les échanges sur les manips, autour d'un repas ou dans le couloir. Merci aussi pour tous vos conseils et notamment à Sébastien pour les discussions autour des enseignements. Merci à vous tous pour les découvertes insolites et les bons moments.

Je remercie les différentes équipes et les membres de l'IBCP pour les sourires et les discussions dans les couloirs. Merci à tous les doctorants pour les moments passés ensemble, j'en garde de super souvenirs !! Un grand merci aux doctorants représentants de l'EDISS et de la « *Team Orga* » qui ont été à mes côtés ces trois dernières années. Je vous souhaite plein de courage et de belles découvertes ;)

Merci à Anna et Fanny pour tous ces moments vécus ensemble et qui m'ont fait grandir pendant cette thèse. Un merci immense à ma Cynthia, je ne pourrais pas faire une liste assez exhaustive haha :) Merci pour la force que tu m'as donné pour le concours (« *Eye of the Tiger* » résonne encore dans mes oreilles) et tout au long de la thèse.

Merci à ma Naf, ma collègue de vie, pour son soutien infini, tous ses conseils, son aide et sa présence.

« *I can't carry it for you, but I can carry you* ». Merci mille fois (et encore beaucoup plus) à ma famille. Vous êtes mon tout (ma pierre de pouvoir), je vous aime !

Ces trois années de thèse étaient géniales et c'était encore plus « extra top » à vos côtés, merci du fond du cœur !!

« *It might seem unfair to reward a person for having so much pleasure over the years* »

Barbara McClintock

## Résumé

Durant ma thèse, j'ai travaillé sur l'étude moléculaire et structurale de complexes protéines:ARN impliqués dans le cycle de vie du virus de l'immunodéficience humaine de type-1 (VIH-1). Mon but était de mieux comprendre leurs mécanismes d'interaction, afin d'établir de nouvelles stratégies antirétrovirales sur le long terme.

**PARTIE I :** L'intégrase (IN) du VIH-1 est une protéine multifonctionnelle qui intervient à différentes étapes du cycle rétroviral. Plusieurs études ont montré l'importance du domaine C-terminal (CTD) de l'IN dans la transcription provirale et la morphogenèse du VIH-1. Il a été montré que le CTD liait fortement l'ARN TAR du génome viral, qui est impliqué avec la protéine Tat dans l'élongation de la transcription provirale. Au cours de ma thèse, j'ai utilisé des approches de biologie moléculaire, biochimie et biophysique, afin d'étudier les interactions entre le CTD, TAR et Tat. Ces travaux ont conduit à la publication d'un article montrant l'importance des changements de conformations induits par la liaison du CTD à TAR, car ils permettent d'améliorer la liaison de Tat. Mes résultats montrent également une interaction du CTD avec le complexe cycline T1:TAR. Ce dernier résultat suggère que l'IN pourrait intervenir dans la transcription provirale en absence de Tat.

**PARTIE II :** ISG20 (« *interferon-stimulated gene 20 kDa protein* ») est un facteur de restriction induit par l'interféron lors d'une réponse antivirale. Il s'agit d'une exonucléase 3'→5' de la famille des DEDD exonucléases. Les mécanismes d'action et de spécificité d'ISG20 vis-à-vis de ses substrats sont mal connus. J'ai mis en évidence *in vitro* que le résidu catalytique D94 d'ISG20 est critique pour la liaison à l'ARN et que l'enzyme est distributive. Une étude récente a montré que les 2'O-méthylations de l'ARN cellulaire permettaient de le protéger contre ISG20. Ce mode de protection est détourné par certains virus, comme le VIH-1. J'ai étudié et mis en évidence *in vitro* deux autres modes de protection des ARN contre l'activité d'ISG20 : l'addition d'une extrémité 3'-polyadénylée qui permet la liaison de certaines protéines et la structure de l'ARN. Ces résultats permettent de mieux expliquer la protection des ARN cellulaires et de certains ARN viraux à l'activité antivirale d'ISG20.

### Mots-clés

Interaction ARN-protéine ; intégrase du VIH-1 ; partenaires moléculaires de l'intégrase ; transcription provirale ; cibles thérapeutiques ; ISG20 ; modifications post-transcriptionnelles de l'ARN ; réponse antivirale.

## Abstract

During my thesis, I worked on the molecular and structural study of protein:RNA complexes involved in the life cycle of human immunodeficiency virus type-1 (HIV-1). My aim was to better understand their interaction mechanisms, in order to establish new long-term antiretroviral strategies.

**PART I:** HIV-1 integrase (IN) is a multifunctional protein that is involved at different stages of the retroviral cycle. Several studies have shown the importance of the C-terminal domain (CTD) of IN in HIV-1 proviral transcription and morphogenesis. The CTD has been shown to strongly bind the TAR RNA of the viral genome which is involved with the Tat protein in the elongation step of proviral transcription. During my thesis, I used molecular biology, biochemistry, and biophysical approaches to study the interactions between CTD, TAR and Tat. This work led to the publication of a paper showing the importance of conformational changes induced by the binding of CTD to TAR, as they enhance the binding of Tat. My results also show an interaction of CTD with the cyclin T1:TAR complex. This latter result suggests that IN may be involved in proviral transcription in the absence of Tat.

**PART II:** ISG20 (interferon-stimulated gene 20 kDa protein) is a restriction factor induced by interferon during an antiviral response. It is a 3'→5' exonuclease of the DEDD exonuclease family. The mechanisms of action and specificity of ISG20 towards its substrates are poorly understood. I have shown *in vitro* that the catalytic residue D94 of ISG20 is critical for RNA binding and that the enzyme is distributive. A recent study showed that 2'O-methylations of cellular RNA protect it from ISG20. This mode of protection is hijacked by some viruses, such as HIV-1. I studied and demonstrated two other modes of RNA protection against ISG20 activity *in vitro*: the addition of a 3'-polyadenylated end that allows the binding of certain proteins and the structure of the RNA. These results help explain the protection of cellular RNA and some viral RNA to the antiviral activity of ISG20.

## Keywords

RNA-protein interaction; HIV-1 integrase; molecular partners of integrase; proviral transcription; therapeutic targets; ISG20; post-transcriptional RNA modifications; antiviral response.

## L'intitulé et l'adresse de l'unité ou du laboratoire où la thèse a été préparée

Institut de Biologie et Chimie des Protéines – Unité Microbiologie Moléculaire et Biochimie Structurale  
UMR 5086, tutelles : UCBL1, CNRS

7 passage du Vercors, 69367 Lyon 7

# Table des matières

Abréviations et unités utilisées.....	9
Chapitre 1 : Introduction .....	11
I. Le virus de l'immunodéficience humaine .....	11
I.1. Historique et épidémiologie .....	11
I.2. Variants génétiques et origines du VIH.....	12
I.3. Transmission du virus .....	13
I.4. Pathologie.....	13
I.5. Diagnostic.....	15
II. La réponse immunitaire contre le VIH-1.....	15
II.1. Généralités sur la réponse immunitaire antivirale .....	15
II.2. Les facteurs de restriction du VIH-1 .....	18
III. Rétrovirus .....	24
III.1. Structure de la particule rétrovirale du VIH-1 .....	25
III.2. Génome rétroviral du VIH-1 .....	26
III.3. Le cycle répliatif du VIH-1.....	28
III.4. Traitement contre le VIH.....	41
IV. Intégrase .....	42
IV.1. Structure de l'intégrase.....	42
IV.2. Modifications post-traductionnelles et rôle multifonctionnel de l'IN du VIH-1 .....	52
IV.3. Les inhibiteurs de l'intégrase.....	56
V. ISG20 .....	58
Le projet doctoral – les objectifs de l'étude .....	59
PARTIE I : Etude fonctionnelle et structurale de l'interaction entre l'IN du VIH-1 et l'ARNg viral ..	61
Chapitre 2 : Résultats .....	61
VI. Partie fonctionnelle.....	61
VI.1. Etude des mécanismes d'interaction entre IN-CTD, Tat et l'ARN TAR .....	61
VI.1. Etude de l'interaction entre IN-CTD, le complexe P-TEFb et l'ARN TAR .....	64
VII. Partie structurale.....	70
VII.1. Purification de l'ARN chimérique minoTAR .....	71
VII.2. Optimisation des conditions de formation du complexe IN-CTD:ARN minoTAR .....	73
VII.3. Analyse de l'état oligomérique du complexe IN-CTD:ARN minoTAR.....	74
VII.4. Essais de cristallogénèse sur le complexe IN-CTD:ARN minoTAR .....	75
Chapitre 3 : Discussion et perspectives (première partie) .....	76
PARTIE II : Caractérisation biochimique de la liaison et de l'activité enzymatique du facteur de restriction viral ISG20 vis-à-vis d'un substrat ARN .....	78
Chapitre 4 : Résultats .....	78
VIII. Préparation de l'ARN.....	78

IX.	Caractérisation de l'activité enzymatique d'ISG20.....	79
IX.1.	Purification et stockage de la protéine recombinante his-ISG20 <sub>wt</sub> .....	79
IX.2.	Etude de la processivité d'ISG20 .....	80
IX.3.	Etude du mutant catalytique D94 .....	81
IX.4.	Etude de la liaison d'ISG20 à des substrats ARN <sub>db</sub> et ARN <sub>sb</sub> .....	84
X.	Mécanismes de protection des ARNm cellulaires et/ou d'échappement viral .....	85
X.1.	La protection par les modifications épitranscriptomiques.....	85
X.2.	La protection par une extrémité 3' polyadénylée .....	88
X.3.	La protection par la structure.....	91
Chapitre 5 : Discussion et perspectives (seconde partie) .....		93
Chapitre 6 : Conclusion générale .....		95
Chapitre 7 : Matériel et Méthodes .....		96
XI.	Techniques de Biologie Moléculaire.....	96
XI.1.	Plasmides.....	96
XI.2.	Transformation des bactéries.....	98
XI.3.	Extractions et purifications des acides nucléiques .....	98
XI.4.	La quantification des acides nucléiques .....	99
XI.5.	Séparation analytique et préparative des ARN.....	99
XI.6.	Conception des constructions plasmidiques.....	100
XII.	Etudes fonctionnelles .....	102
XII.1.	Surproduction et purification des protéines de fusion.....	102
XII.2.	Electrophorèse en gel de polyacrylamide en condition dénaturante pour les protéines 106	
XII.3.	Production d'ARN.....	107
XII.4.	Marquage radioactif de l'ARN catalysé par la T4 polynucléotide kinase.....	110
XII.5.	Complexes ribonucléoprotéiques .....	110
XIII.	Etudes structurales.....	113
XIII.1.	Analyse de la monodispersité des échantillons par diffusion dynamique de la lumière 113	
XIII.2.	Cristallographie aux rayons X.....	114
Références bibliographiques .....		118
Contributions scientifiques en lien avec les travaux de thèse .....		135
Annexes.....		136

## Abréviations et unités utilisées

7SK snRNP, <i>7SK small nuclear ribonucleoprotein</i>	dGTP, déoxyguanosine triphosphate
aa, acide aminé	dGTPase, déoxyguanosine triphosphate
A260, absorbance à 260 nm	triphosphohydrolase
A280, absorbance à 280 nm	DIS, <i>dimerization initiation site</i>
ADN, acide désoxyribonucléique	DISF, <i>DRB sensitivity-inducing factor</i>
ADNc, acide désoxyribonucléique complémentaire	dNTP, désoxyribonucléotide triphosphate
ADNdb, acide désoxyribonucléique double brin	DLS, <i>dynamic light scattering</i> , diffusion dynamique
ADNsb, acide désoxyribonucléique simple brin	de la lumière
AFF1, <i>AF4/FMR2 family member 1</i>	DMSO, diméthylsulfoxyde
AFF4, <i>AF4/FMR2 family member 4</i>	DNase, désoxyribonucléase
ALLINI, <i>allosteric integrase inhibitors</i>	DTT, dithiothréitol
AlphaScreen, <i>Amplified Luminescent Proximity Homogeneous Assay</i>	$\epsilon$ , coefficient d'absorption molaire
ALV, <i>Avian leukosis virus</i> , virus de la leucose aviaire	<i>E. coli</i> , <i>Escherichia coli</i>
APOBEC3, <i>apolipoprotein B mRNA-editing enzyme catalytic polypeptide 3</i>	EDTA, <i>ethylenediaminetetraacetic acid</i>
APS, <i>ammonium persulfate</i> , persulfate d'ammonium	Env, enveloppe
ARM, <i>arginine-rich motif</i>	ESCRT, <i>endosomal sorting complex required for transport</i>
ARN, acide ribonucléique	ESRF, <i>European Synchrotron Radiation Facility</i>
ARNdb, acide ribonucléique double brin	FDA, <i>Food and Drug Administration</i>
ARNg, acide ribonucléique génomique	<i>gag</i> , <i>group-specific antigen</i>
ARNm, acide ribonucléique messenger	GCN2, <i>general control non-derepressible 2</i>
ARNsb, acide ribonucléique simple brin	GCN5, <i>general control non-depressible 5</i>
ARNt, acide ribonucléique de transfert	gp41, glycoprotéine de 41 kDa
ATP, adénosine triphosphate	gp120, glycoprotéine de 120 kDa
AZT, azidothymidine	GPI, <i>glycosyl phosphatidylinositol</i>
BAF, <i>barrier to autointegration factor</i> , protéine barrière à l'auto-intégration	GST, glutathion S-transférase
BET, bromure d'éthidium	HAART, <i>highly active antiretroviral therapy</i>
BLI, <i>bio-layer interferometry</i> , l'interférométrie de bio-couche	HAT, histone acétyltransférase
BSA, <i>bovine serum albumin</i> , albumine de sérum bovin	HCl, acide chlorhydrique
<i>Bulge</i> , renflement	HDGF, <i>hepatoma-derived growth factor</i>
C-terminale, carboxy-terminale	HDGF2, <i>hepatoma-derived growth factor like 2</i>
CA, capsid	HEPES, acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
CARD, <i>caspase activation and recruitment domain</i>	HEXIM1/2, <i>hexamethylene bis-acetamide inducible 1 or 2</i>
CC, <i>coiled coil</i>	HMGA1, <i>high-mobility group protein A1</i>
CCD, <i>catalytic core domain</i>	HRM, <i>histidine-rich motif</i>
CCR5, <i>chemokine receptor type 5</i>	HTLV-1, <i>Primate T-lymphotropic virus type-1</i>
CD4, <i>cluster of differentiation 4</i>	IBD, <i>integrase-binding domain</i>
CDC, <i>cleaved donor complex</i>	ICTV, <i>International Committee on Taxonomy of Viruses</i>
CDK, <i>cyclin-dependent kinase</i> , kinase dépendante d'une cycline	IFITM1, <i>IFN-induced transmembrane protein 1</i>
CIAP, <i>calf intestinal alkaline phosphatase</i> , phosphatase alcaline intestinale du veau	IFN, interféron
CIC, <i>conserved intasome core</i>	IFNAR, <i>IFN-<math>\alpha/\beta</math> receptor</i>
CIRI, Centre International de Recherche en Infectiologie	IMAC, <i>immobilized metal ion affinity chromatography</i> , chromatographie d'affinité sur ions métalliques immobilisés
CLIP-Seq, <i>Crosss-Linking ImmunoPrecipitation Sequencing</i>	IN, intégrase
CNRS, centre nationale de la recherche scientifique	INI1, <i>integrase interactor 1</i>
CPSF6, <i>cleavage and polyadenylation specificity factor subunit 6</i>	INSTI, <i>integrase strand transfer inhibitors</i>
CT, extrémité C-terminale du CTD	IPTG, isopropyl- $\beta$ -D-1-thiogalactopyranoside
CTD, <i>carboxy-terminal domain</i> , domaine C-terminal	ISG, <i>IFN-stimulated gene</i>
Cullin2/VHL, <i>cullin2-based von Hippel-Lindau</i>	ISG, <i>IFN-stimulated gene protein</i>
CXCR4, <i>C-X-C motif chemokine receptor type 4</i>	ISG20, <i>IFN-stimulated gene 20 kDa protein</i>
Da, dalton	ISRE, <i>IFN-stimulated response element</i>
	JAK-STAT, <i>Janus kinase – Signal transducers and activators of transcription</i>
	JNK, <i>c-Jun N-terminal kinase</i>
	kb, kilobases
	$K_D$ , constante de dissociation à l'équilibre
	$k_{on}$ , constante de vitesse d'association

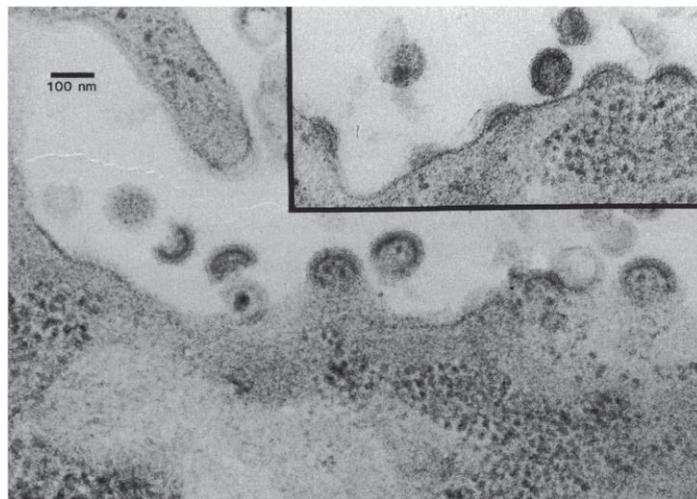
$k_{off}$ , constante de vitesse de dissociation  
 LARP7, *La ribonucleoprotein domain family member 7*  
 LB, *lysogeny broth*, bouillon lysogène  
 LEDGF, *lens epithelium-derived growth factor*  
 LRR, *leucine-rich repeats*  
 LTR, *long terminal repeat*  
 MA, matrice  
 MAPK, *mitogen-activated protein kinases*  
 MePCE, *methylphosphate capping enzyme*  
 MLV, *Murine leukemia virus*  
 MMTV, *Mouse mammary tumor virus*  
 Mn, manganèse  
 MVV, *Visna-maedi virus*  
 MX1 et 2, *myxovirus resistance 1 and 2*  
 NC, nucléocapside  
 NEB, New England BioLabs  
 NED, *N-terminal extension domain*  
 Nef, *negative regulatory factor*  
 NELF, *negative elongation factor*  
 NES, *nuclear export signal*  
 NF- $\kappa$ B, *nuclear factor- $\kappa$ B*  
 Ni, nickel  
 NK, *natural killer*  
 NLR, *NOD-like receptor*  
 NLS, *nuclear localization signal*  
 NPC, *nuclear pore complex*, complexe du pore nucléaire  
 nt, nucléotide  
 NTA, acide nitrilotriacétique  
 NTD, *amino-terminal domain*  
 N-terminale, amino-terminale  
 $\Psi$ , *psi/packaging signal sequence*  
 p7, protéine de 7 kDa  
 p17, protéine de 17 kDa  
 p24, protéine de 24 kDa  
 p51, protéine de 51 kDa  
 p66, protéine de 66 kDa  
 PAF1, *polymerase-associated factor complex 1*  
 PAGE, *polyacrylamide gel electrophoresis*, électrophorèse sur gel de polyacrylamide  
 PAMP, *pathogen associated molecular pattern*  
 pb, paire de bases  
 PBS, *primer binding site*  
 PCI, phénol : chloroforme : alcool isoamylique  
 PCR, *polymerase chain reaction*, réaction en chaîne par polymérase  
 PDB, *protein data bank*  
 PDI, *polydispersity index*, coefficient de polydispersité  
 PEG, *polyethylene glycol*  
 PFV, *prototype foamy virus*  
 PIC, *pre-integration complex*, complexe de pré-intégration  
 Pin1, *peptidyl prolyl-isomerase*  
 polyA, polyadénylée  
 PPT, *polypurine tract*  
 PR, protéase  
 PRR, *pattern recognition receptor*  
 R, *repeat element*  
 $R_H$ , rayon hydrodynamique  
 RE, réticulum endoplasmique  
 Rev, *regulator of virion expression protein*  
 RIG-I, *retinoic acid inducible gene 1*  
 RING, *really interesting new gene*  
 RLR, *RIG-like receptor*  
 RMN, résonance magnétique nucléaire  
 RNAPII, *RNA polymerase II*, ARN polymérase II  
 RNase, ribonucléase  
 RNP, ribonucléoprotéine  
 RRE, *Rev response element*  
 RSV, *Rous sarcoma virus*  
 RT, *reverse transcriptase*, transcriptase inverse  
 SAMHD1, *sterile alpha motif domain- and histidine-aspartic (HD) domain-containing protein 1*  
 SD, *splice donor*  
 SDN1, *small RNA degrading nuclease 1*  
 SDS, *sodium dodecyl sulfate*, dodécylsulfate de sodium  
 SDS-PAGE, *sodium dodecyl sulfate – polyacrylamide gel electrophoresis*  
 SEC, *size-exclusion chromatography*, chromatographie d'exclusion stérique  
 SEC, *super elongation complex*  
 SH3, *Src homology 3 domain*  
 SIDA, syndrome de l'immunodéficience humaine  
 SLFN11, Schlafen 11  
 SOLEIL, Source Optimisée de Lumière d'Énergie Intermédiaire du LURE  
 SSC, *stable synaptic complex*  
 STC, *strand transfer complex*  
 SU, *surface glycoprotein*  
 SWI/SNF, *SWItch/Sucrose Non-Fermentable*  
 TAE, Tris Acétate EDTA  
 TAR, *trans-activation response element*  
 Tat, protéine trans-activatrice de la transcription  
 TBE, Tris Borate EDTA  
 TCC, *target capture complex*  
 TEMED, *N,N,N',N'-tetramethylethane-1,2-diamine*  
 TEV, *tobacco etch virus*, virus de la gravure du tabac  
 TLR, *Toll-like receptor*  
 TM, *transmembrane glycoprotein*  
 TRAF, *TRIF-related adaptator molecule*  
 Tris, *tris(hydroxymethyl)aminomethane*  
 TRIF, *TIR-domain containing adaptater inducing interferon- $\beta$*   
 TRIM, *tripartite motif*  
 TRM, *Tat-TAR recognition motif*  
 TRN-SR2, transportine-SR2  
 U, *unique element*  
 UTR, *un-translated terminal region*, région terminale non traduite  
 UV, ultraviolet  
 VC, volume colonne  
 VIH, virus de l'immunodéficience humaine  
 VIH-1, virus de l'immunodéficience humaine de type-1  
 VIH-2, virus de l'immunodéficience humaine de type-2  
 Vif, *viral infectivity factor*  
 VIS, virus de l'immunodéficience simienne  
 Vpr, *viral protein R*  
 Vpu, *viral protein U*  
 Vpx, *viral protein X*  
 VSV, virus de la stomatite vésiculaire  
 wt, *wild-type*, sauvage  
 Zn, zinc

# Chapitre 1 : Introduction

## I. Le virus de l'immunodéficience humaine

### I.1. Historique et épidémiologie

Les premiers cas de patients atteints du syndrome de l'immunodéficience acquise (SIDA) ont été reportés en 1981 aux Etats-Unis (ressource : <https://www.hiv.gov>). Ces patients immunodéprimés étaient sensibles à des infections opportunistes ordinairement bénignes. Deux ans plus tard, en 1983, l'agent causal du SIDA est identifié et isolé par Luc Montagnier et Françoise Barré-Sinoussi à l'Institut Pasteur, sous le nom du virus de l'immunodéficience humaine (VIH) (**figure 1**) (Barré-Sinoussi *et al.*, 1983). Ces travaux leur vaudront quinze ans plus tard le prix Nobel de médecine.



**Figure 1.** Image de microscopie électronique réalisée à partir de sections fines de lymphocytes de cordon ombilical produisant des virus. L'encart montre plusieurs stades de bourgeonnement des particules virales à la surface cellulaire. (Barré-Sinoussi *et al.*, 1983)

Les premiers tests de diagnostic de l'infection par le VIH sont mis au point à l'Institut Pasteur en 1985. Il s'agit de tests sérologiques permettant de détecter dans le sang d'un patient la présence d'anticorps spécifiques dirigés contre le virus. Le génome viral est séquencé la même année (Wain-Hobson *et al.*, 1985) et le VIH de type 2 (VIH-2) est isolé (Sanchez-Pescador *et al.*, 1985).

Le premier médicament contre le VIH est mis au point en 1987. Il s'agit de l'azidothymidine (AZT), un inhibiteur d'une enzyme spécifique du VIH, la transcriptase inverse (RT), qui rétrotranscrit l'ARN génomique (ARNg) viral en ADN double brin (ADNdb) avant son intégration dans le génome de la cellule hôte (voir **chapitre III.3.2**). Des avancées dans le traitement contre le VIH ont permis la conception des premières trithérapies, avec la combinaison de trois molécules ciblant différentes étapes du cycle viral pour augmenter la puissance du traitement (Shafer et Vuitton, 1999).

Aujourd'hui, le VIH est toujours une menace avec 38 millions de personnes infectées et environ 40 millions de décès depuis sa découverte (ressource : <https://www.unaids.org>). Un nouveau variant plus virulent du sous-type B du VIH de type 1 (VIH-1) a été découvert récemment (Wymant *et al.*, 2022).

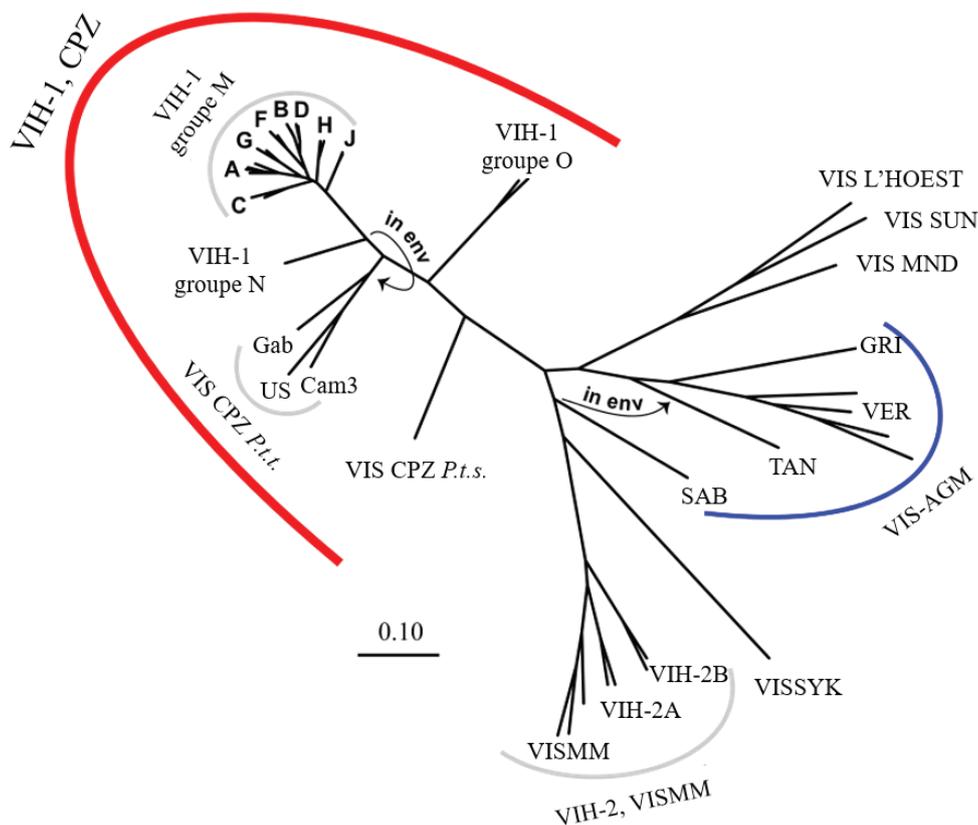
## I.2. Variants génétiques et origines du VIH

Le VIH est un virus présentant une très grande variabilité génétique lui permettant de détourner les défenses immunitaires de la cellule ainsi que les thérapies médicamenteuses et vaccinales. Cette variabilité est due à trois mécanismes : 1) les erreurs due à l'activité de la RT qui introduit environ  $10^{-4}$  –  $10^{-5}$  mutations par base par cycle de réplication ; 2) le taux élevé de réplication virale qui permet de produire environ  $10^{10}$  virions par jour dans la première phase de l'infection ; 3) les processus de recombinaison génétique entre différents VIH au sein d'une même personne infectée (Rezende et Prasad, 2004 ; Fanales-Belasio *et al.*, 2010 ; Coffin, 1995).

Le VIH comporte deux types, le VIH-1 et le VIH-2, qui dérivent du virus de l'immunodéficience simienne (VIS) qui infecte le singe (**figure 2**). Plus particulièrement, le VIH-1 dérive du VIS infectant les chimpanzés et les gorilles d'Afrique centrale de l'Ouest, tandis que le VIH-2 dérive du VIS infectant le mangabey d'Afrique de l'Ouest et possède le gène *vpx* qui est absent dans le VIH-1 (King *et al.*, 2011).

Plusieurs variants du VIH-1 ont été identifiés : le groupe M (« *major group* »), le groupe O (« *outlier group* »), le groupe N (« *non-M, non-O group* ») et le groupe P. Le groupe M comprend neuf sous-types ou clades (de A à D, de F à H, J et K), il est largement distribué sur l'ensemble des continents et correspond à 99% des infections par le VIH-1. Les trois autres groupes sont endémiques de l'Afrique centrale. Le groupe O, moins fréquent, correspond à 1% des infections par le VIH-1 et est limité au Cameroun et au Gabon. Les groupes N et P correspondent respectivement à 14 et 2 cas au Cameroun (Mourez *et al.*, 2013).

Le VIH-2 est moins pathogène et est endémique de l'Afrique de l'Ouest et de l'Europe, avec quelques cas sporadiques en Inde et aux Etats-Unis (Campbell-Yesufu et Gandhi, 2011 ; Goujon *et al.*, 2009). Il est responsable de 1% des infections par le VIH. Neuf groupes du VIH-2 ont été identifiés (de A à I), A et B étant les deux groupes majoritaires. Ainsi, le VIH-1 est responsable de la pandémie mondiale du SIDA.



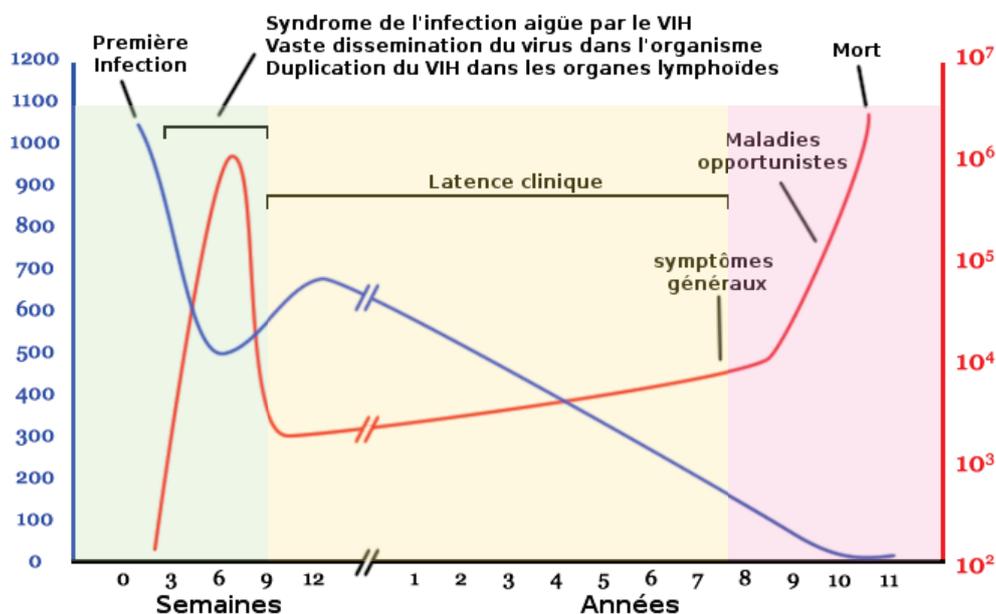
**Figure 2.** Arbre phylogénétique du VIH et du VIS. VISagm SAB : « African green monkey *Sabaeus subspecies* » ; VIScpz P.t.t. : « chimpanzee subspecies *Pan troglodytes troglodytes* » ; VIS P.t.s. : « *Pan troglodytes shweinfuthii* » ; TAN : « *Tantalus* » ; VER : « *Vervet* » ; GRI : « *Grivet* » ; SAB : « *Sabaeus* » ; SIVmm : « *Sooty Mangaby* » ; VISSyk : « *Skyes' monkeys* » ; SIV MND : « *Mandril* ». Source : Kuiken et al., 1999.

### I.3. Transmission du virus

Le VIH peut se transmettre par voie sanguine via des blessures, des injections ou des transfusions ; par voie sexuelle via le sperme ou les sécrétions vaginales ; de la mère à l'enfant avant la naissance par le placenta, lors de l'accouchement ou par le lait maternel lors de l'allaitement. Le risque de transmission de la mère à l'enfant peut être réduit à moins de 1% grâce aux nouveaux traitements antirétroviraux qui réduisent la charge virale de la personne infectée (ressource : <https://www.usaid.gov>).

### I.4. Pathologie

La charge virale évolue au cours du temps chez une personne séropositive sans suivi médical. Ainsi, trois phases de la pathogénie liée au VIH peuvent être décrites selon la charge virale et le nombre de lymphocytes T CD4<sup>+</sup> présent dans le plasma (**figure 3**) : la primo-infection ou phase aiguë (en vert), la phase asymptomatique (en jaune) et le dernier stade qui correspond au SIDA (en rose). Le nombre de lymphocytes T CD4<sup>+</sup> dans le plasma est un marqueur de l'avancement de la pathologie liée au VIH, car le virus les infecte après reconnaissance de leurs glycoprotéines de surface CD4 (voir **chapitre III.3.1**). Le VIH peut infecter d'autres cellules exprimant ces clusters de différenciation, tels les macrophages et certaines cellules dendritiques.



**Figure 3.** Evolution de la charge virale au cours du temps. La courbe bleue représente le nombre de lymphocytes T CD4<sup>+</sup> par mm<sup>3</sup> de plasma ; la courbe rouge représente le nombre de copies de l'ARN viral par mL de plasma. Source : Wikipédia.

#### I.4.1. La primo-infection ou phase aiguë

La phase de primo-infection ou phase aiguë est caractérisée par une augmentation brutale suivie d'une baisse importante de la charge virale et une diminution importante du nombre de lymphocytes T CD4<sup>+</sup>, qui réaugmente par la suite (**figure 3**, partie verte). En effet, la charge virale est très élevée et le nombre de lymphocytes T CD4<sup>+</sup> qui sont ciblés par le virus diminue brutalement. Puis, le système immunitaire réagit, ce qui permet de diminuer la charge virale et entraîne une augmentation du nombre de lymphocytes T CD4<sup>+</sup>. C'est à ce moment que des anticorps contre le VIH sont produits par le système immunitaire et qu'ils peuvent être détectés par les tests de dépistage. Il s'agit de l'étape de séroconversion avec le passage d'une personne séronégative à séropositive. Ainsi, dans les premiers temps de cette phase, les personnes infectées sont très contagieuses car leur charge virale est importante.

Lors de cette phase le nombre de lymphocytes T CD4<sup>+</sup> par mm<sup>3</sup> de plasma reste supérieur ou égal à 500/mm<sup>3</sup> et la personne infectée est encore en bonne santé avec un système immunitaire fonctionnel. En effet, un taux normal de lymphocytes T CD4<sup>+</sup> se situe entre 600 et 1 200/mm<sup>3</sup>.

La phase de primo-infection peut s'étendre sur quelques semaines ou quelques mois. Les personnes infectées peuvent être asymptomatiques ou présenter des symptômes comme ceux liés au virus de la grippe (fatigue, forte fièvre, douleurs musculaires, maux de tête et diarrhée) ou encore d'autres symptômes plus spécifiques (vomissements, éruption cutanée, perte de poids, maux de gorge, gonflement des ganglions lymphatiques, ...).

#### I.4.2. La phase asymptomatique

La phase asymptomatique correspond à l'augmentation progressive de la charge virale et à une diminution lente et régulière du nombre de lymphocytes T CD4<sup>+</sup> (**figure 3**, partie jaune). Lors de cette phase le taux de lymphocytes T CD4<sup>+</sup> est inférieur à 500/mm<sup>3</sup>, ce qui entraîne une perte de fonction du système immunitaire.

La phase asymptomatique peut s'étendre sur plusieurs années. Les traitements antirétroviraux actuellement sur le marché permettent d'allonger cette phase et d'empêcher l'évolution de la maladie vers le stade SIDA. En effet, un traitement antirétroviral efficace permet de diminuer drastiquement la charge virale et d'interrompre le déclin du nombre de lymphocytes T CD4<sup>+</sup>.

#### I.4.3. Le SIDA

La dernière phase correspond au SIDA. Elle est caractérisée par une augmentation brutale de la charge virale suite à la perte de fonctionnalité du système immunitaire avec un taux de lymphocytes T CD4<sup>+</sup> quasiment nul (**figure 3**, partie rose). Lorsque le nombre de lymphocytes T CD4<sup>+</sup> est inférieur à 200/mm<sup>3</sup>, le risque de développer des maladies opportunistes est très élevé. Ces maladies opportunistes peuvent être causées par des infections secondaires bactériennes, fongiques, parasitaires, virales ou encore liées à des cancers.

Cette phase conduit généralement au décès de la personne séropositive au bout de 1 à 2 ans sans traitement antirétroviral.

#### I.4.4. Exception chez certains individus

Quelques personnes infectées résistent spontanément à l'infection (moins de 1%) et présentent une charge virale inférieure au seuil de détection. Ces personnes possèdent en effet certains gènes spécifiques qui permettent une réponse très forte contre le VIH par les lymphocytes T CD8<sup>+</sup>.

D'autres personnes sont même résistantes à l'infection grâce à une mutation dans le gène codant pour le corécepteur CCR-5 (« *chemokine receptor type 5* ») qui est présent à la surface des lymphocytes T CD4<sup>+</sup> (voir **chapitre III.3.1**).

#### I.5. Diagnostic

Comme décrit précédemment, différents stades sont observés au cours de la pathologie causée par le VIH. Différents marqueurs biologiques et symptômes cliniques peuvent être utilisés pour détecter l'infection par le VIH et mettre en place un traitement médical adapté au patient. Ces marqueurs sont l'ARN viral/la charge virale, l'antigène p24 qui est la protéine de capsid rétrovirale (voir **chapitre III.1**), les anticorps contre le VIH et les lymphocytes T CD4<sup>+</sup> (Parekh *et al.*, 2019). Ces marqueurs sont utilisés pour déterminer le statut séropositif/séronégatif d'une personne (anticorps), le stade de l'infection (anticorps, antigène p24, charge virale), suivre l'avancement de la pathologie (lymphocytes T CD4<sup>+</sup>) ou encore identifier/suivre l'efficacité ou l'échec d'un traitement (charge virale).

Le diagnostic peut être réalisé grâce à un test sanguin de quatrième génération, qui combine la détection des anticorps contre le VIH et la détection de l'antigène p24 (Huynh et Kahwaji, 2022). L'antigène p24 est détectable 14 jours après exposition au virus. Il est possible de réaliser un test ARN contre le VIH pour une détection avant 14 jours.

D'autres méthodes de dépistage peuvent être utilisées sous formes de tests rapides, afin de détecter uniquement les anticorps ; comme un test rapide d'orientation diagnostique à partir d'une goutte de sang ; ou un autotest à partir d'une goutte de sang ou de la salive. Ces tests permettent de détecter une infection par le VIH de plus de trois semaines.

## II. La réponse immunitaire contre le VIH-1

### II.1. Généralités sur la réponse immunitaire antivirale

Le système immunitaire des vertébrés à mâchoire comprend l'immunité innée et l'immunité adaptative qui sont étroitement imbriquées (Janeway *et al.*, 2001 ; Abbas *et al.*, 2016 ; Reece *et al.*, 2010). L'immunité innée permet une réponse rapide, tandis que l'immunité adaptative est plus lente mais plus spécifique et permet la mise en place d'une mémoire immunitaire.

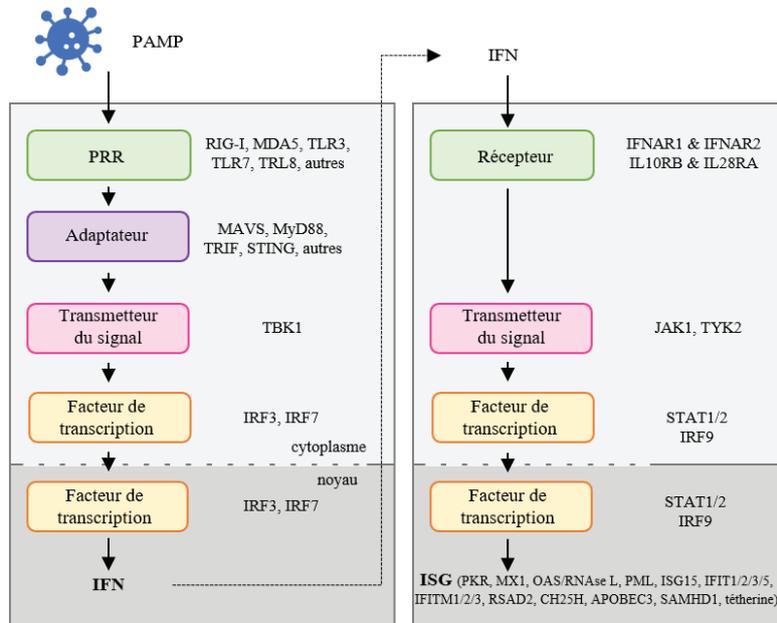
Lors d'une infection par un pathogène, les cellules du système immunitaire inné (mastocytes-macrophages, cellules dendritiques, cellules « *natural killer* » (NK)) le détectent grâce à leurs récepteurs de motifs moléculaires (PRR, « *pattern recognition receptor* ») (Schoggins, 2014 ; Yin *et al.*, 2020). Ces récepteurs, externes ou internes, reconnaissent des motifs moléculaires associés aux différents types de pathogènes (PAMP, « *pathogen associated molecular pattern* »), comme les lipopolysaccharides bactériens ou l'ARN double brin (ARNdb) de certains virus. Les PAMP sont des séquences conservées indispensables au fonctionnement biologique du pathogène. Lorsqu'il y a une interaction entre un PAMP et un PRR, un signal est transduit à la cellule et permet de l'activer (Janeway *et al.*, 2001 ; Abbas *et al.*, 2016 ; Reece *et al.*, 2010). Cette activation se traduit par la transcription de nouveaux gènes, qui entraînent l'acquisition de nouvelles fonctions comme la sécrétion de médiateurs pro-inflammatoires, de chimiokines ; des modifications des molécules d'adhésion ; la sécrétion de molécules de costimulation qui permettent d'orienter une réponse adaptative ; et la sécrétion des cytokines

pléiotropiques : interférons (IFN) de type I (IFN- $\alpha$  et IFN- $\beta$ ) qui confèrent de nouvelles fonctions effectrices aux cellules émettrices (sécrétion autocrine) et environnantes (sécrétion paracrine).

Les trois familles principales de PRR sont les suivantes (Janeway *et al.*, 2001 ; Reece *et al.*, 2010) :

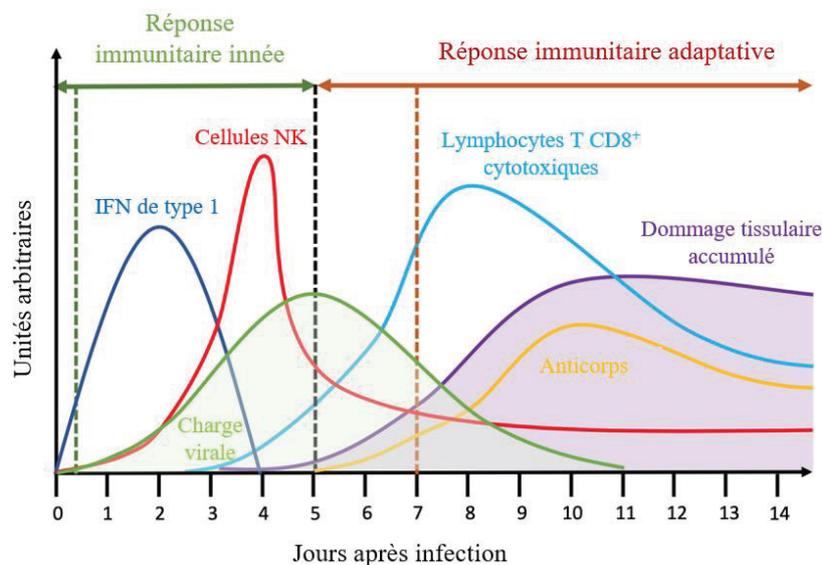
- Les TLR (« *Toll-like receptor* ») sont une grande famille de PRR très étudiée qui contient dix membres (TLR1 à TLR10). Ces récepteurs sont présents à la surface de la membrane plasmique cellulaire ou endosomale. MyD88 permet la transduction du signal de tous les TLR sauf les TLR3 et TLR4, pour lesquels la voie TRIF (« *TIR-domain containing adaptater inducing interferon- $\beta$*  ») / TRAF (« *TRIF-related adaptator molecule* ») est utilisée. Le TLR3 permet de détecter les ARNdb viraux, tandis que les TLR 7 et TLR8 permettent de détecter les ARN simple brin (ARNsb) viraux.
- Les NLR (« *NOD-like receptor* ») sont une famille contenant une vingtaine de membres. Ils sont connectés à deux voies principales de transduction du signal : la voie NF- $\kappa$ B (« *nuclear factor- $\kappa$ B* ») / MAPK (« *mitogen-activated protein kinases* ») et la voie de l'inflammasome. Les NLR sont des récepteurs cytosoliques comprenant un domaine LRR (« *leucine-rich repeats* ») qui permet la reconnaissance du PAMP et un domaine NACHT qui permet l'oligomérisation.
- Les RLR (« *RIG-like receptor* ») sont des récepteurs cytoplasmiques qui permettent la reconnaissance des ARNdb. Les RLR contiennent un ou deux domaines CARD (« *caspase activation and recruitment domain* ») qui permettent la transduction du signal en aval du récepteur, un domaine hélicase permettant l'ouverture des ARNdb et un domaine de liaison à l'ARN. Ces récepteurs sont connectés à deux voies de transduction du signal : la voie TRAF-6 et la voie NF- $\kappa$ B.

Lors d'une infection virale, ces trois familles de PRR permettent de détecter le virus avec spécifiquement dans le cas d'un virus à génome ARN : TLR3, TLR7, TLR8, RIG-I (« *retinoic acid inducible gene 1* »), MDA5 (**figure 4**). Les protéines adaptatrices initient la cascade de signalisation qui conduit à l'activation de facteurs de transcription. Les cellules activées sécrètent des médiateurs solubles, notamment les IFN de type I. Les IFN de type I sont reconnus par les récepteurs IFNAR (« *IFN- $\alpha/\beta$  receptor* »), puis une transduction du signal par la voie JAK-STAT (« *Janus kinase – Signal transducers and activators of transcription* ») permet d'activer l'expression de gènes spécifiques appelés ISG (« *IFN-stimulated gene* »). La production des protéines codées par les ISG aboutit à la production de facteurs de restriction, susceptibles de bloquer l'infection virale (voir **chapitre suivant**). Les protéines codées par les ISG les plus connues sont la protéine kinase R (PKR), le système 2',5'-oligoadénylate synthase (OAS) / ribonucléase L (RNAse L) (Schoggins, 2014 ; Kristiansen *et al.*, 2011) et les protéines MX1 et 2 (« *myxovirus resistance 1 and 2* ») qui inhibent plusieurs étapes du cycle réplcatif du VIH-1 suivant l'entrée du virus dans la cellule infectée (Goujon *et al.*, 2013 ; Liu *et al.*, 2013 ; Kane *et al.*, 2013 ; Doyle *et al.*, 2015).



**Figure 4.** Schéma récapitulant la réponse antivirale innée par le système immunitaire. Sources : Schoggins, 2014 ; Yin et al., 2020.

Pour ce qui concerne la réponse adaptative, les cellules infectées peuvent présenter à leur surface des fragments peptidiques viraux (antigènes exogènes) aux lymphocytes T CD8<sup>+</sup> cytotoxiques pour les activer (**figures 5 et 6**). Cette réponse cytotoxique cellulaire complète celle des NK pour éliminer les cellules infectées. Dans un second temps, une réponse humorale due aux lymphocytes B permet d'éliminer les virus circulants par production d'anticorps neutralisant (**figures 5 et 6**). Les lymphocytes T CD4<sup>+</sup> sont, pour leur part, activés par des cellules présentatrices d'antigènes exogènes comme les macrophages. Ils orchestrent l'amplification des réponses adaptatives cellulaire et humorale ainsi que la production de lymphocytes mémoires (**figure 6**) qui permettent une réponse plus forte et plus rapide lors d'une seconde infection.



**Figure 5.** Graphique représentant la production de médiateurs et d'effecteurs de la réponse immunitaire au cours du temps. Source : Sego et al., 2020.

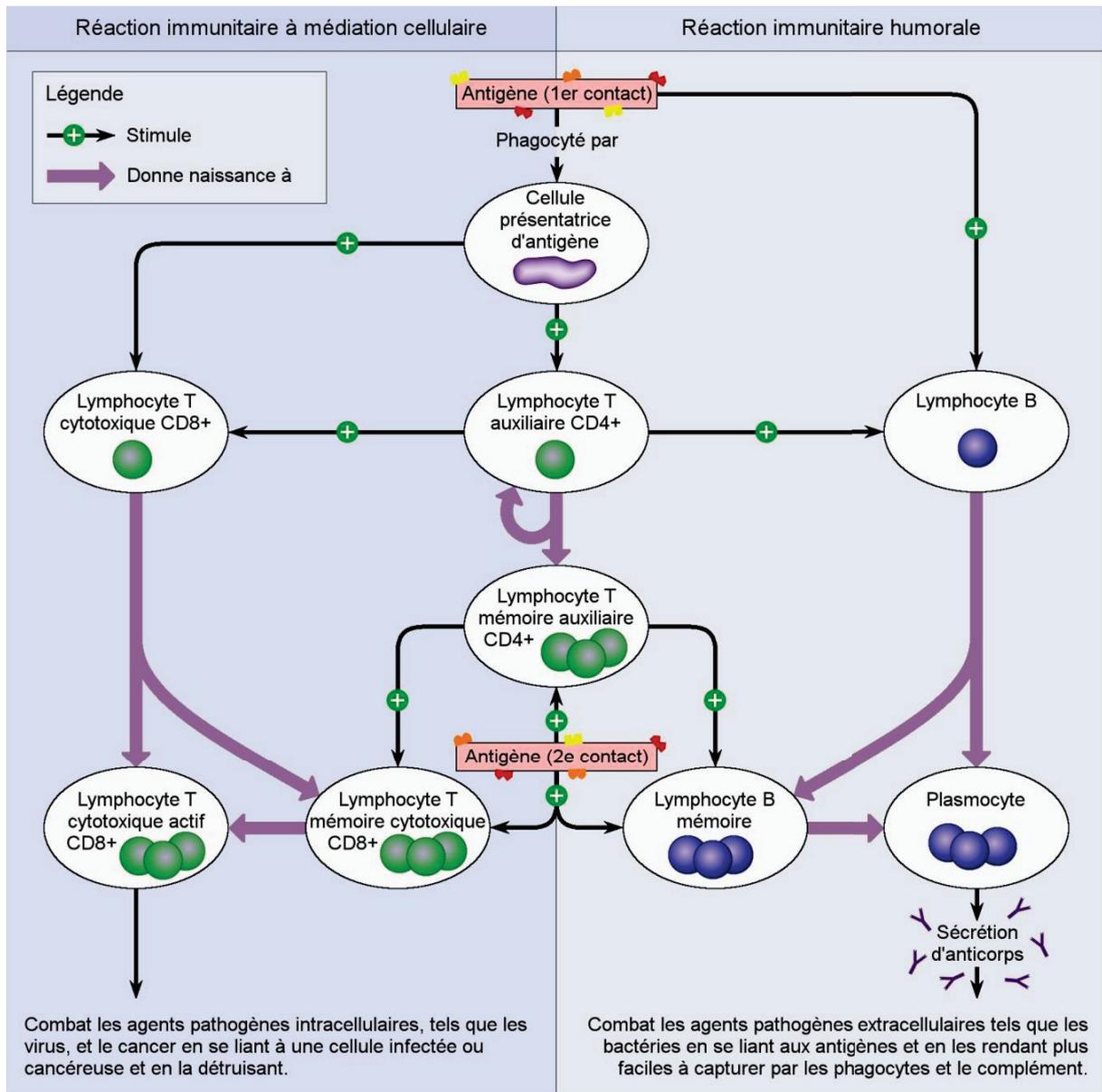


Figure 6. Schéma récapitulatif de l'immunité adaptative. Source : Reece et al., 2010.

## II.2. Les facteurs de restriction du VIH-1

Les facteurs de restriction sont des protéines cellulaires limitant la réplication virale (Malim et Bieniasz, 2012 ; Harris *et al.*, 2012). Ils peuvent conduire le virus à évoluer et à développer un mécanisme d'échappement.

Les virus ne possédant pas ce mécanisme d'échappement ne peuvent se répliquer que dans certains types cellulaires (les cellules permissives) et pas dans d'autres (les cellules non-permissives), ce qui permet de mettre en évidence les mécanismes d'échappement viraux (Harris *et al.*, 2012). Les interactions entre facteur de restriction et mécanisme d'échappement viral entraînent une signature d'évolution rapide dans la séquence du gène codant le facteur de restriction. L'expression de facteurs de restriction est fortement liée à la réponse immunitaire innée, elle est souvent fortement induite par l'IFN (Harris *et al.*, 2012).

Dans le cas du VIH-1, les principaux facteurs de restriction sont les protéines de la famille APOBEC3 (« apolipoprotein B mRNA-editing enzyme catalytic polypeptide 3 »), de la famille TRIM (« tripartite motif »), SAMHD1 (« sterile alpha motif domain- and histidine-aspartic (HD) domain-containing

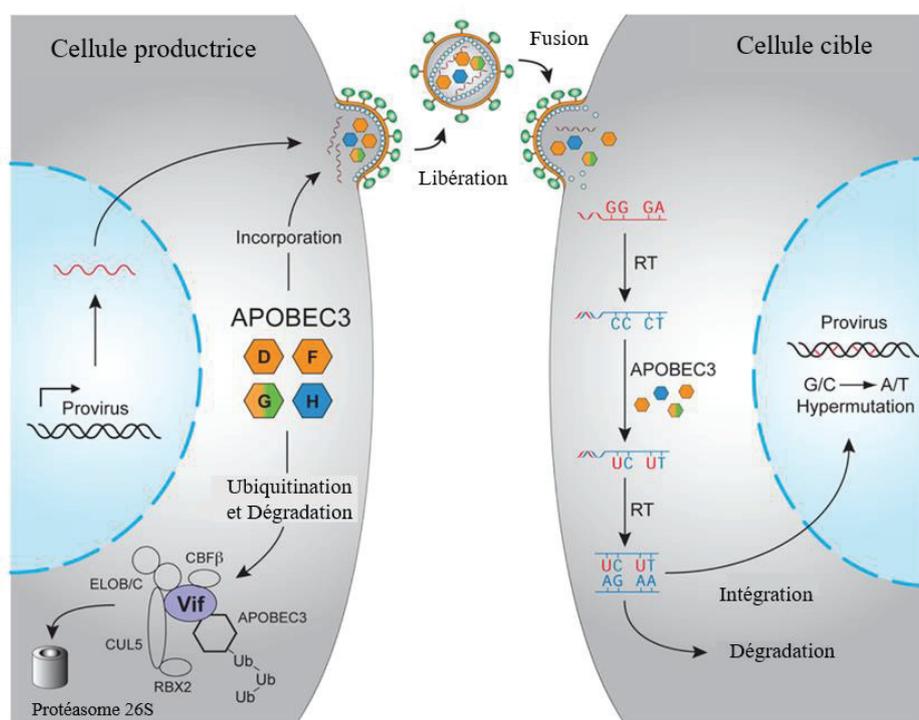
*protein 1* ») et la tétherine. Les mécanismes d'échappement du VIH-1 reposent sur ses protéines accessoires (voir **chapitre III.2**) avec : Vif (« *viral infectivity factor* ») comme antagoniste des protéines APOBEC3 ; Vpx (« *viral protein X* ») pour SAMHD1 ; Vpu (« *viral protein U* ») et Nef (« *negative regulatory factor* ») pour la tétherine. Les mécanismes d'échappement à TRIM sont peu compris actuellement (Koepke *et al.*, 2021).

Le mode d'action de ces facteurs de restriction est détaillé par la suite en indiquant les étapes du cycle viral impactées. Pour le détail sur les étapes du cycle viral voir **chapitre III.3**.

### II.2.1. La famille APOBEC3

Cette famille contient sept membres : APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G et APOBEC3H. Les protéines APOBEC3D, APOBEC3F, APOBEC3G et APOBEC3H inhibent la réplication du VIH (Harris *et al.*, 2012).

Ces protéines contiennent un ou deux domaines cytosine désaminase qui catalysent la désamination de bases azotées C en U dans le premier brin négatif d'ADN complémentaire (ADNc) lors de la transcription inverse (**figure 7**, à droite) (Malim et Bieniasz, 2012). Ces sites désaminés auront une base complémentaire A au lieu de G lors de la synthèse du second brin d'ADN, résultant en un ADNdb hypermuté susceptible d'être dégradé ou de conduire à la synthèse de protéines virales non-fonctionnelles (Levy, 2009 ; Huthoff et Malim, 2005 ; Li *et al.*, 2023 ; Liddament *et al.*, 2004 ; Sato *et al.*, 2014 ; Shaban *et al.*, 2018). Cependant l'action de ces protéines est bloquée par l'interaction directe avec la protéine virale Vif, qui empêche leur incorporation dans les particules virales. De plus, Vif détourne le complexe ligases Cullin-RING et le cofacteur transcriptionnel CBF $\beta$ , entraînant l'ubiquitination et la dégradation par le protéasome de ces quatre protéines de la famille APOBEC3 (**figure 7**, à gauche).

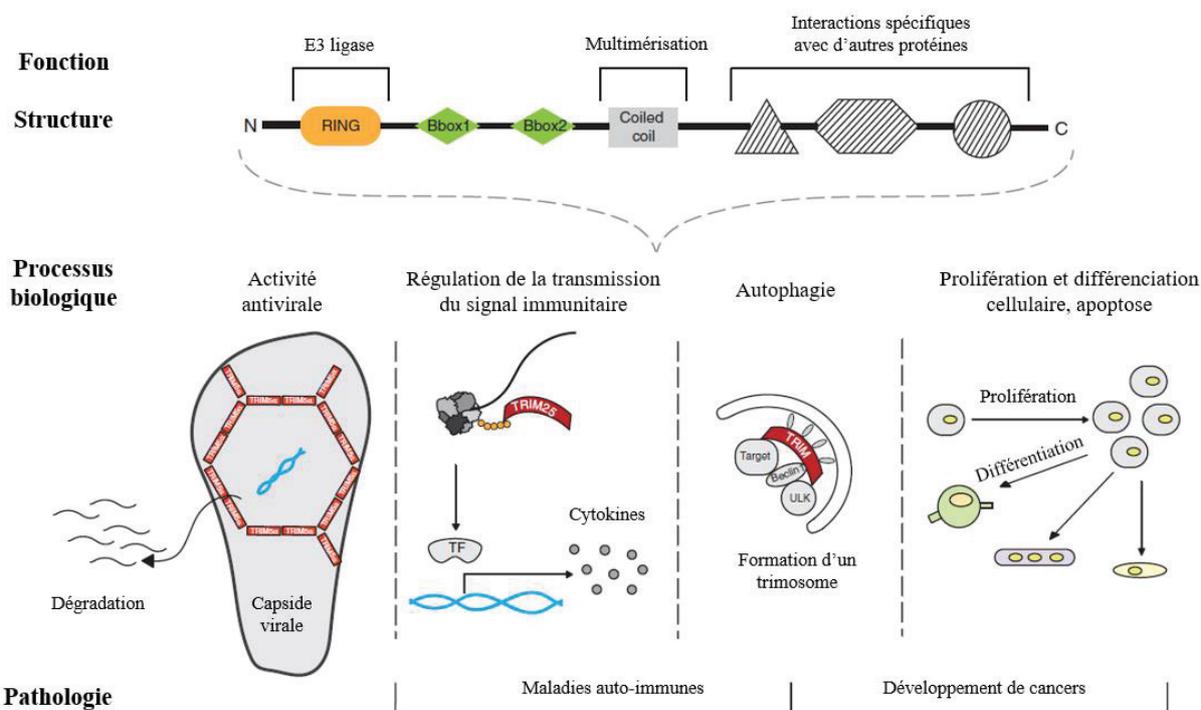


**Figure 7.** Mécanisme de restriction du VIH-1 par les protéines APOBEC3 et mécanisme d'échappement viral par la protéine rétrovirale Vif. Source : Harris *et al.*, 2012.

## II.2.2. La famille TRIM

La famille TRIM constitue une vaste famille de protéines qui regroupe environ quatre-vingts protéines. Il s'agit de protéines multimériques caractérisées par un motif commun RBCC qui est organisé comme ceci de l'extrémité N-terminale à l'extrémité C-terminale : un domaine RING qui est un domaine E3 ligase en doigt à zinc, suivi d'un ou de deux autres domaines en doigt à zinc nommés « *Bbox* » et un domaine « *coiled coil* » qui est un domaine d'oligomérisation (Nisole *et al.*, 2005 ; Vunjak et Versteeg, 2019) (**figure 8**). Les protéines de la famille TRIM diffèrent par leur domaine C-terminal qui leur confère une activité enzymatique ou leur permet de réaliser des interactions spécifiques avec des partenaires/cibles protéiques ou nucléiques (Gent *et al.*, 2018).

Les fonctions cellulaires des protéines TRIM sont multiples, elles ont des rôles dans plusieurs processus biologiques tels que la prolifération, l'apoptose, la différenciation, la transcription et la réparation de l'ADN. Elles jouent également un rôle dans la défense cellulaire contre les virus en tant que facteurs de restriction viral. Plusieurs protéines TRIM interviennent et inhibent des étapes du cycle réplcatif de nombreux virus, notamment le VIH-1. De plus, les protéines TRIM ont un rôle antiviral additionnel car elles permettent de moduler la réponse IFN antivirale.



**Figure 8.** Représentation de l'organisation des domaines des protéines TRIM, ainsi que les différentes fonctions et pathologies associées. Source : Vunjak et Versteeg, 2019.

Parmi les différentes protéines TRIM étudiées pour leur implication dans l'inhibition de l'infection par le VIH-1, TRIM5 $\alpha$  est la plus étudiée. TRIM5 $\alpha$  inhibe les étapes suivant l'entrée du VIH-1 dans la cellule infectée en interagissant avec la capsid virale (Ganser-Pornillos *et al.*, 2010 ; Li *et al.*, 2016 ; Yu *et al.*, 2020) (**figure 8**). Trois mécanismes d'inhibition sont possibles : (1) TRIM5 $\alpha$  cause une décapsidation prématurée de la capsid virale ; (2) TRIM5 $\alpha$  recrute l'ubiquitine qui conduit à une décapsidation non productive du virus et une dégradation par le protéasome ; (3) TRIM5 $\alpha$  conduit à la dégradation de la capsid virale par des enzymes lysosomales dans la voie de l'autophagie (Stremlau *et al.*, 2004 ; Levy, 2009 ; Ganser-Pornillos et Pornillos, 2019 ; Jimenez-Guardenõ *et al.*, 2019). Dans les trois mécanismes, l'étape de transcription inverse est bloquée.

D'autres protéines TRIM inhibent le VIH-1, par exemple :

- TRIM11 interagit avec la capsid virale et la nucléocapside, entraînant l'ouverture de la capsid et la libération de l'ARNg prématurément, inhibant les étapes suivantes du cycle répliatif (van Tol *et al.*, 2017). Le mode d'action de TRIM11 est différent de celui de TRIM5 $\alpha$ , car il n'induit pas une dégradation médiée par le protéasome ou des enzymes lysosomales.
- TRIM19 est transloquée du noyau dans le cytoplasme lors d'une infection par le VIH-1, elle se lie à Daxx (van Tol *et al.*, 2017). Cette interaction empêche la dégradation de Daxx par le protéasome et permet d'interférer avec l'étape de transcription inverse par la suite de façon Daxx dépendante (Dutrieux *et al.*, 2015).
- TRIM22 empêche la liaison du facteur de transcription Sp1 aux séquences LTR (« *long terminal repeat* ») du VIH-1, ce qui inhibe la transcription provirale (Tissot et Mechti, 1995 ; van Tol *et al.*, 2017 ; Kajaste-Rudnitski *et al.*, 2011). De plus, TRIM22 perturbe le transport de la polyprotéine Gag à la membrane plasmique (Gent *et al.*, 2018).

D'autres protéines TRIM sont impliquées dans l'inhibition du VIH-1 : TRIM15 (Uchil *et al.*, 2008), TRIM26 (Uchil *et al.*, 2008), TRIM28 (Ma *et al.*, 2019), TRIM31 (Uchil *et al.*, 2008), TRIM32 (Uchil *et al.*, 2008), TRIM37 (Tabah *et al.*, 2014) et TRIM69 (Song *et al.*, 2022). Le mécanisme d'action moléculaire n'est pas entièrement élucidé pour certaines protéines TRIM, et l'activité antivirale contre le VIH-1 est spécifique à un type cellulaire dans certains cas (Song *et al.*, 2022).

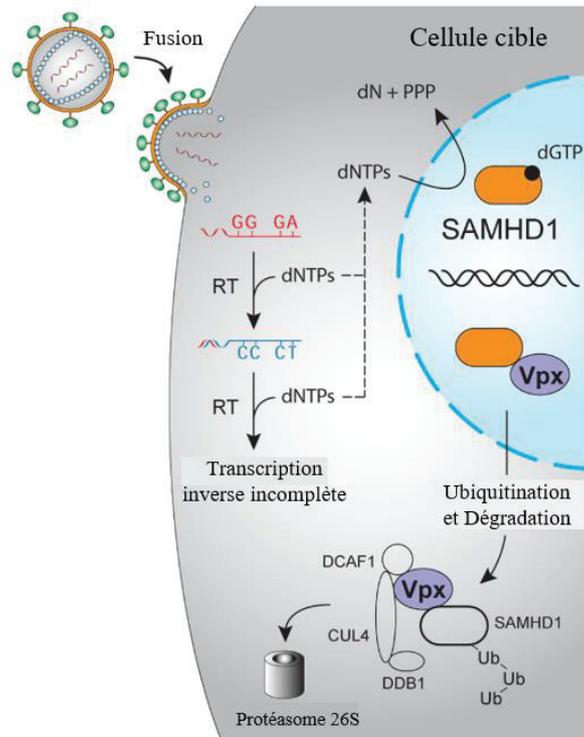
### II.2.3. SAMHD1

SAMHD1 contient un domaine HD qui lui confère une activité désoxyguanosine triphosphate triphosphohydrolase (dGTPase) (Goldstone *et al.*, 2011 ; Harris *et al.*, 2012). Ce domaine HD permet également des interactions avec des partenaires protéiques, une multimérisation et la liaison d'acides nucléiques (White *et al.*, 2012 ; Tüngler *et al.*, 2013).

SAMHD1 est un facteur de restriction du VIH qui inhibe la synthèse de l'ADNc lors de l'étape de transcription inverse (Hrecka *et al.*, 2011) (**figure 9**). Deux mécanismes d'inhibition médiée par SAMHD1 ont été décrits : (1) une inhibition grâce à l'activité nucléase de SAMHD1 qui entraîne une diminution des désoxyribonucléotides triphosphates (dNTP), en particulier les dGTP, et donc limite la synthèse d'ADNc (Lahouassa *et al.*, 2013 ; Goldstone *et al.*, 2011) ; (2) une inhibition grâce à l'activité exonucléase 3' vers 5' de SAMHD1 qui dégrade directement l'ARNg (Beloglazova *et al.*, 2013 ; Ryoo *et al.*, 2014 ; Choi *et al.*, 2015).

Une hypothèse serait que lorsque la quantité de dGTP cellulaire est faible, les protéines SAMHD1 sont présentes sous forme de monomères ou de dimères et possèdent uniquement une activité RNase permettant la dégradation de l'ARNg (Ryoo *et al.*, 2014 ; Ji *et al.*, 2013 ; Yan *et al.*, 2013). Lorsque la quantité de dGTP cellulaire est élevée, les protéines SAMHD1 multimérisent sous forme tétramérique et ont une activité dGTPase permettant une dégradation des dGTP et une inhibition de la synthèse d'ADNc.

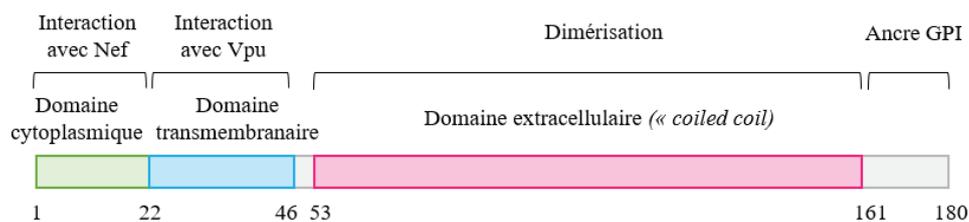
La protéine accessoire Vpx du VIH-2 et de VIS induit la dégradation de SAMHD1 par le protéasome via le détournement de l'ubiquitine E3 ligase CRL4 (Laguet *et al.*, 2011 ; Hrecka *et al.*, 2011) (**figure 9**).



**Figure 9.** Mécanisme de restriction du VIH par SAMHD1 présentant une activité dGTPase et mécanisme d'échappement viral par la protéine rétrovirale Vpx. Source : Harris et al., 2012.

#### II.2.4. Téthérine

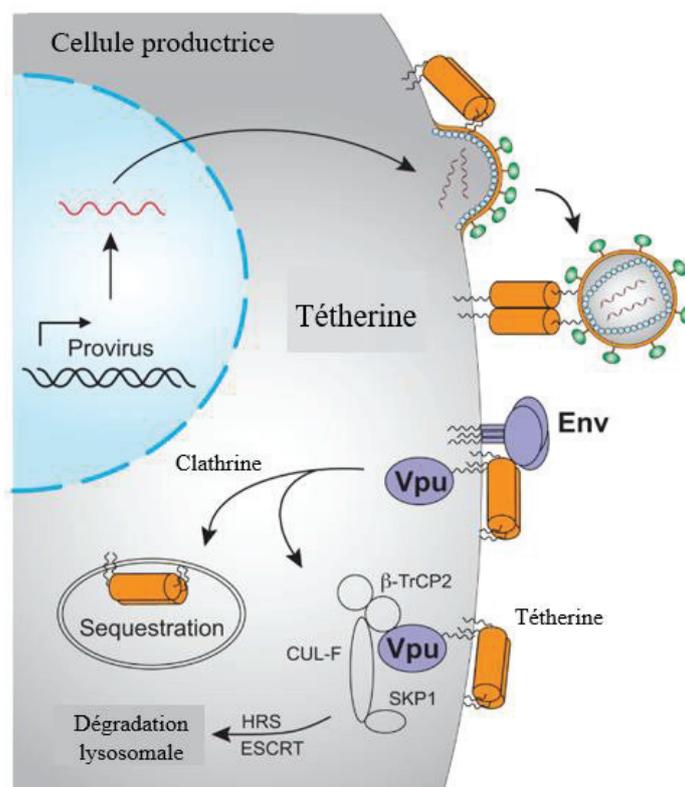
La téthérine est une protéine induite par l'IFN et donc de type ISG (« *IFN-stimulated gene protein* »). Elle est composée d'un domaine N-terminal cytoplasmique, suivi d'un domaine transmembranaire, d'un domaine de dimérisation extracellulaire et d'une ancre GPI (« *glycosyl phosphatidylinositol* ») en C-terminal (Neil et al., 2008) (**figure 10**). La téthérine est présente au niveau de la membrane plasmique et au niveau de la membrane endosomale (Dubé et al., 2011).



**Figure 10.** Représentation de l'organisation des domaines de la téthérine. Source : Malim et Bieniasz, 2012.

La téthérine permet l'attachement des particules virales enveloppées à la surface de la cellule infectée en se liant à la fois à la membrane plasmique de la cellule infectée et à l'enveloppe de la particule virale à la surface cellulaire (Perez-Caballero et al., 2009 ; Harris et al., 2012) (**figure 11**). Dans le cas du VIH-1, la téthérine agit sous la forme d'un homodimère, avec ses domaines N-terminaux localisés dans le cytoplasme qui interagissent avec la machinerie d'endocytose cellulaire. La particule virale attachée par la téthérine est ensuite internalisée par une voie dépendante de la clathrine et dégradée dans les lysosomes (Miyakawa et al., 2009 ; Lau et al., 2011).

Lors d'une infection par le VIH-1, Vpu interagit avec la tétherine et empêche son transport aux sites d'assemblage des particules virales au niveau de la membrane plasmique. Vpu induit la dégradation endosomale de la tétherine.



**Figure 11.** Mécanisme de restriction du VIH-1 par la tétherine et mécanisme d'échappement viral par la protéine rétrovirale Vpu. Source : Harris *et al.*, 2012.

Lors d'une infection par le VIH-2, en absence de Vpu, certaines protéines de l'enveloppe virale (gp41) permettent l'internalisation et la séquestration de la tétherine par une voie dépendante de la clathrine, ce qui permet de libérer les virus à la surface cellulaire (Hauser *et al.*, 2010 ; Le Tortorec et Neil, 2009 ; Lau *et al.*, 2011). Chez le VIS, la protéine accessoire Nef permet d'inhiber la tétherine par une voie dépendante de la clathrine également (Zhang *et al.*, 2009 ; Zhang *et al.*, 2011).

## II.2.5. Identification de nouveaux facteurs de restriction du VIH-1

D'autres facteurs de restriction viraux sont connus. Il s'agit de protéines codées par des *ISG*, tels que : IFITM1 (« *IFN-induced transmembrane protein 1* »), ITFM2 et ITFM3 qui inhibent la fusion et la transmission du virus (Compton *et al.*, 2014) ; Schlafen 11 (SLFN11) qui inhibe la traduction de l'ARN messager (ARNm) viral (Li *et al.*, 2012) ; ISG15 (« *IFN-stimulated gene 15 kDa protein* ») qui inhibe l'assemblage et le bourgeonnement de la particule virale (Pincetic *et al.*, 2010 ; Okumura *et al.*, 2005).

De nouveaux facteurs de restriction du VIH-1 ont été découverts récemment, il s'agit de protéines codées par des *ISG* qui impactent la transcription et la traduction virale, comme SAMD9L (Legrand *et al.*, sur bioRxiv) et ISG20 (« *IFN-stimulated gene 20 kDa protein* ») (Espert *et al.*, 2005 ; Deymier *et al.*, 2022 ; El Kazzi *et al.*, 2022).

Durant ma thèse j'ai travaillé sur ISG20, une exonucléase ciblant les ARN viraux. Cette protéine est présentée en détail dans le **chapitre V**.

### III. Rétrovirus

Le VIH fait partie de la famille des *Retroviridae* aussi communément appelée rétrovirus. Cette famille comprend des virus infectant les invertébrés et les vertébrés (Strauss et Strauss, 2008). Les rétrovirus infectant les mammifères et les oiseaux sont les plus étudiés, comme le VIH-1 pour l'homme et le virus de la leucose aviaire (ALV) pour les poulets principalement. La classification taxonomique des virus est définie par l'« *International Committee on Taxonomy of Viruses* » (ICTV), qui utilise une méthode assez semblable à la classification du vivant et par David Baltimore, lauréat du prix Nobel de médecine en 1975, en se basant sur le type d'acide nucléique du virus et son mode d'expression (ressources : <https://ictv.global/> ; <https://viralzone.expasy.org/>).

Selon la classification ICTV mise à jour en mars 2021, la taxonomie du VIH-1 est la suivante :

**Règne :** *Riboviria*

**Royaume :** *Pararnavirae*

**Phylum :** *Arteriviricota*

**Classe :** *Revtraviricetes*

**Ordre :** *Ortervirales*

**Famille :** *Retroviridae*

**Sous-famille :** *Orthoretrovirinae*

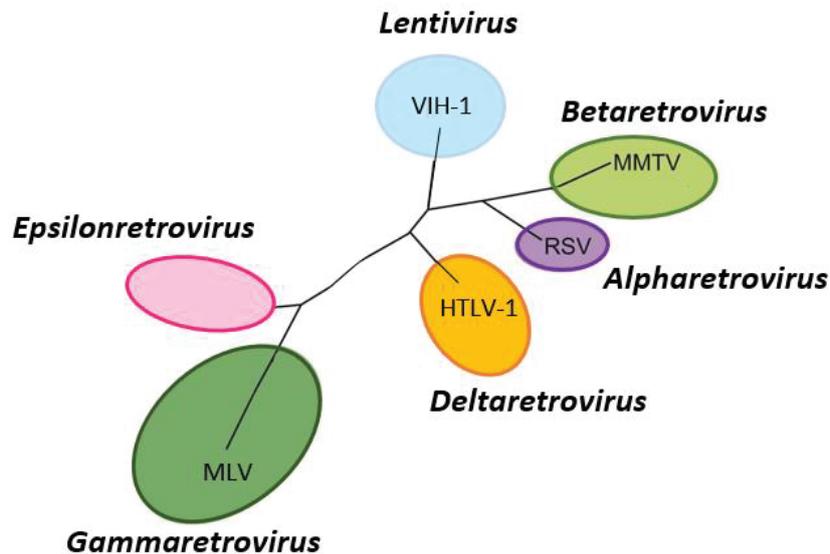
**Genre :** *Lentivirus*

**Espèce :** *VIH-1*

La famille des rétrovirus comprend deux sous-familles dans la classification ICTV : *Orthoretrovirinae* et *Spumaretrovirinae* (ressource : <https://ictv.global/>).

La sous-famille *Orthoretrovirinae* comprend 6 genres et 49 espèces (**figure 12**) :

- Le genre *Alpharetrovirus* qui comprend 9 espèces, dont ALV et RSV (« *Rous sarcoma virus* »).
- Le genre *Betaretrovirus* qui comprend 5 espèces, dont MMTV (« *Mouse mammary tumor virus* »).
- Le genre *Deltaretrovirus* qui comprend 4 espèces, dont HTLV-1 (« *Primate T-lymphotropic virus type-1* »).
- Le genre *Epsilonretrovirus* qui comprend 3 espèces.
- Le genre *Gammaretrovirus* qui comprend 18 espèces, dont MLV (« *Murine leukemia virus* »).
- Le genre *Lentivirus* qui contient 10 espèces dont le VIH-1.



**Figure 12.** Analyse phylogénétique des régions conservées du gène *pol* communs à tous les rétrovirus (ressource : <https://ictv.global/>).

La sous-famille des *Spumaretrovirinae* est moins grande et comprend 5 genres et 19 espèces, dont l'espèce *Eastern chimpanzee simian foamy virus*, du genre *Simiispumavirus*, qui est connue sous le nom de PFV (« *prototype foamy virus* ») (Khan *et al.*, 2018).

Selon la classification de Baltimore, le VIH-1 fait partie du groupe VI contenant les virus avec un génome à ARNs positif, qui est rétrotranscrit lors du cycle réplcatif.

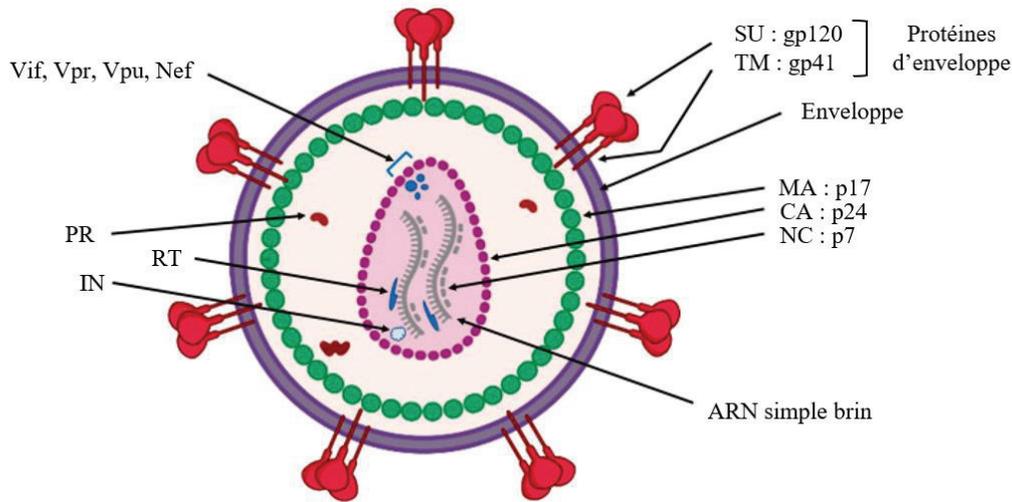
### III.1. Structure de la particule rétrovirale du VIH-1

La particule virale du VIH-1 est enveloppée et mesure entre 80 et 100 nm de diamètre (King *et al.*, 2011). Elle contient une enveloppe, une matrice (MA), une capsid (CA), une nucléocapsid (NC), des enzymes virales, des protéines accessoires et un génome ARN (**figure 13**).

L'enveloppe virale provient de la membrane de la cellule hôte. Elle est constituée de glycolipides et de glycoprotéines cellulaires. Plusieurs spicules sont présents à la surface de l'enveloppe virale. Ils sont composés de deux types de glycoprotéines virales : les glycoprotéines de surface (SU) qui sont des glycoprotéines de 120 kDa (gp120), et les glycoprotéines transmembranaires (TM) qui sont des glycoprotéines de 41 kDa (gp41) dans le cas du VIH-1.

Le cœur de la particule virale est constitué de protéines de matrice (p17) qui renferment une capsid conique composée de protéines p24. La capsid contient les protéines de nucléocapsid (p7) en complexe avec deux molécules d'ARN.

Trois types d'enzymes virales sont présentes dans la particule virale : l'intégrase (IN), la RT et la protéase (PR). De plus, des protéines accessoires (Vif, Vpr (« *viral protein R* »), Vpu) et régulatrices (Nef) sont présentes dans la particule du VIH-1.



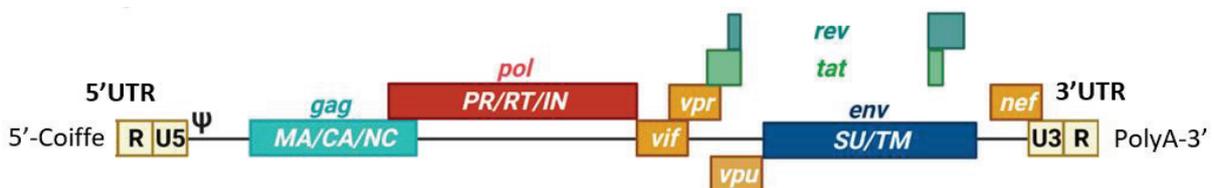
**Figure 13.** Représentation d'une particule rétrovirale mature du VIH-1. Sources : van Heuvel et al., 2022 ; Rogez, 2022.

### III.2. Génome rétroviral du VIH-1

Le génome du VIH-1 est diploïde et compte deux copies d'ARNsb de polarité positive. Chaque monomère d'ARN est polyadénylée (polyA) à l'extrémité 3' et possède une coiffe à l'extrémité 5' (King et al., 2011). Les deux copies d'ARN sont liées au niveau de leurs extrémités 5'. La taille d'une copie d'ARN varie entre 7 et 13 kb. Dans le cas du VIH-1, un monomère d'ARN a une taille de 9,7 kb et est organisé comme montré sur la **figure 14** (Sertznig et al., 2018).

Trois gènes sont communs à tous les rétrovirus : *gag* (« *group-specific antigen* »), *pol* et *env*. Le gène viral *gag* code les trois protéines structurales : MA, CA et NC. Le gène viral *pol* code les enzymes : PR, RT et IN. Le gène viral *env* code les protéines d'enveloppe : SU et TM.

Les rétrovirus complexes, comme ceux du genre *Lentivirus*, possèdent des protéines accessoires (Vif, Vpr, Vpu) et des protéines régulatrices (la protéine « trans-activatrice de la transcription » appelée Tat, Rev (« *regulator of virion expression protein* »), Nef) qui jouent un rôle dans l'expression des gènes, la réplication virale et l'inhibition des facteurs de restriction cellulaires. Ces protéines additionnelles confèrent une virulence supplémentaire aux rétrovirus complexes en comparaison aux rétrovirus simples, ce qui conduit souvent à la lyse des cellules infectées à l'issue de la réplication virale (King et al., 2011 ; Strauss et Strauss, 2008).

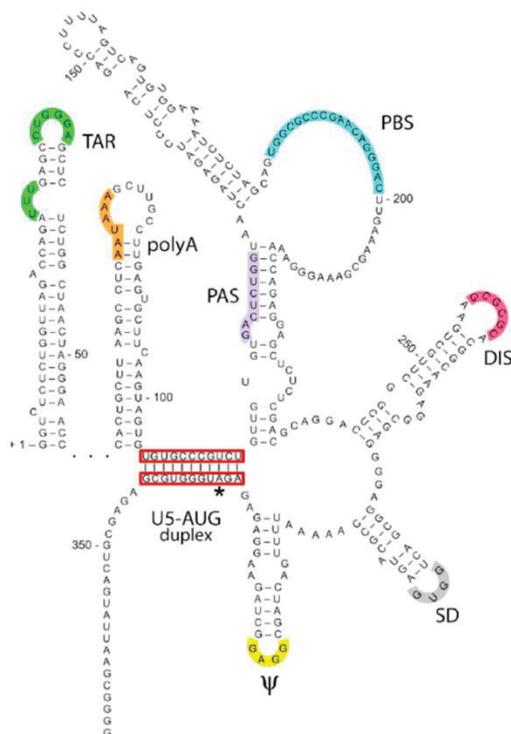


**Figure 14.** Représentation du génome du VIH-1 avec les régions non traduites appelées « un-translated terminal region » (UTR) aux extrémités 5' et 3' et les régions codantes pour des protéines (*gag*, *pol*, *env*, *vif*, *vpr*, *vpu*, *tat*, *rev* et *nef*). Source : van Heuvel et al., 2022.

Le 5' UTR contient une séquence R (« *repeat element* ») et une séquence U5 (« *unique element* »). Le 3' UTR contient une séquence U3 et une séquence R. Selon les virus, la séquence R a une taille qui varie

entre 15 et 230 nucléotides, la séquence U5 entre 70 et 220 nucléotides et la séquence U3 entre 230 et 1 200 nucléotides de long (Strauss et Strauss, 2008).

Le 5'UTR de l'ARNg du VIH-1 est fortement structuré et présente des structures secondaires particulières en tige-boucle (**figure 15**). Il s'agit de la région de l'ARNg du VIH-1 la plus conservée (Paillart *et al.*, 2004). Cette région est centrale pour la régulation de la réplication virale (Brigham *et al.*, 2019). Elle contient six éléments ARN principaux : TAR, PolyA, PBS, DIS, SD et Ψ.



**Figure 15.** Représentation des éléments structurés dans le 5'UTR du VIH-1. Source : van Heuvel *et al.*, 2022.

TAR (« *transactivation response* ») est le premier élément du 5'UTR, il contient 59 nucléotides (Chavali *et al.*, 2019). TAR joue un rôle dans la transcription de l'ADN proviral (l'ADN du VIH-1 intégré dans celui de la cellule hôte) en interagissant avec la protéine rétrovirale Tat, ce qui permet de détourner de manière productive le mécanisme d'élongation de la transcription cellulaire (voir **chapitre III.3.4**).

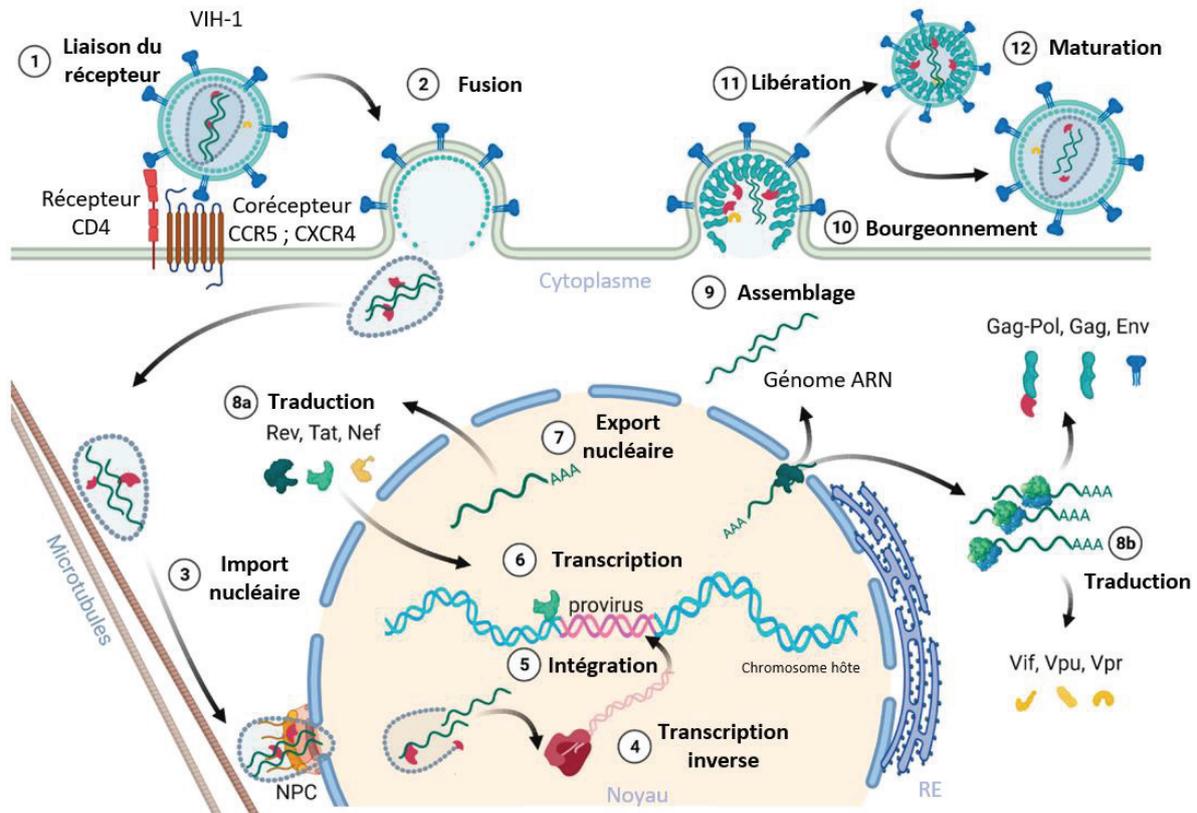
PBS (« *primer binding site* ») est un site de liaison d'un ARNt spécifique permettant d'amorcer la synthèse d'ADN lors de l'étape de transcription inverse (Strauss et Strauss, 2008). En effet, cet élément contient dans sa séquence 18 nucléotides complémentaires à l'extrémité 3' d'un ARNt spécifique. Dans le cas du VIH-1, l'ARNt est un ARNt-Lys (Strauss et Strauss, 2008).

DIS (« *dimerization initiation site* ») est l'élément qui permet l'initiation de la dimérisation dans la forme « *kissing* » ou linéaire de l'ARNg. Le 5'UTR varie entre différentes conformations : une conformation linéaire où l'ARNg est monomérique avec l'élément DIS qui est séquestré par un appariement spécifique des bases ; et une conformation « *kissing* » où l'élément DIS est exposé et le génome présente une boucle de dimérisation « *kissing* » (Brigham *et al.*, 2019).

Les éléments ARN PolyA, SD (« *splice donor* ») et Ψ (« *psi/ packaging signal sequence* ») sont des signaux permettant, respectivement, la polyadénylation, l'épissage et l'encapsidation de l'ARNg.

### III.3. Le cycle réplcatif du VIH-1

Le cycle rétroviral du VIH-1 comprend plusieurs étapes détaillées par la suite (**figure 16**). Il permet la production de particules virales matures capables d'infecter des cellules cibles.



**Figure 16.** Représentation du cycle réplcatif du VIH-1. NPC, complexe du pore nucléaire (« nuclear pore complex ») ; CCR5 ; CXCR4 (« C-X-C motif chemokine receptor type 4 ») ; RE, réticulum endoplasmique. Source : van Heuvel et al., 2022.

#### III.3.1. La reconnaissance et l'entrée dans la cellule

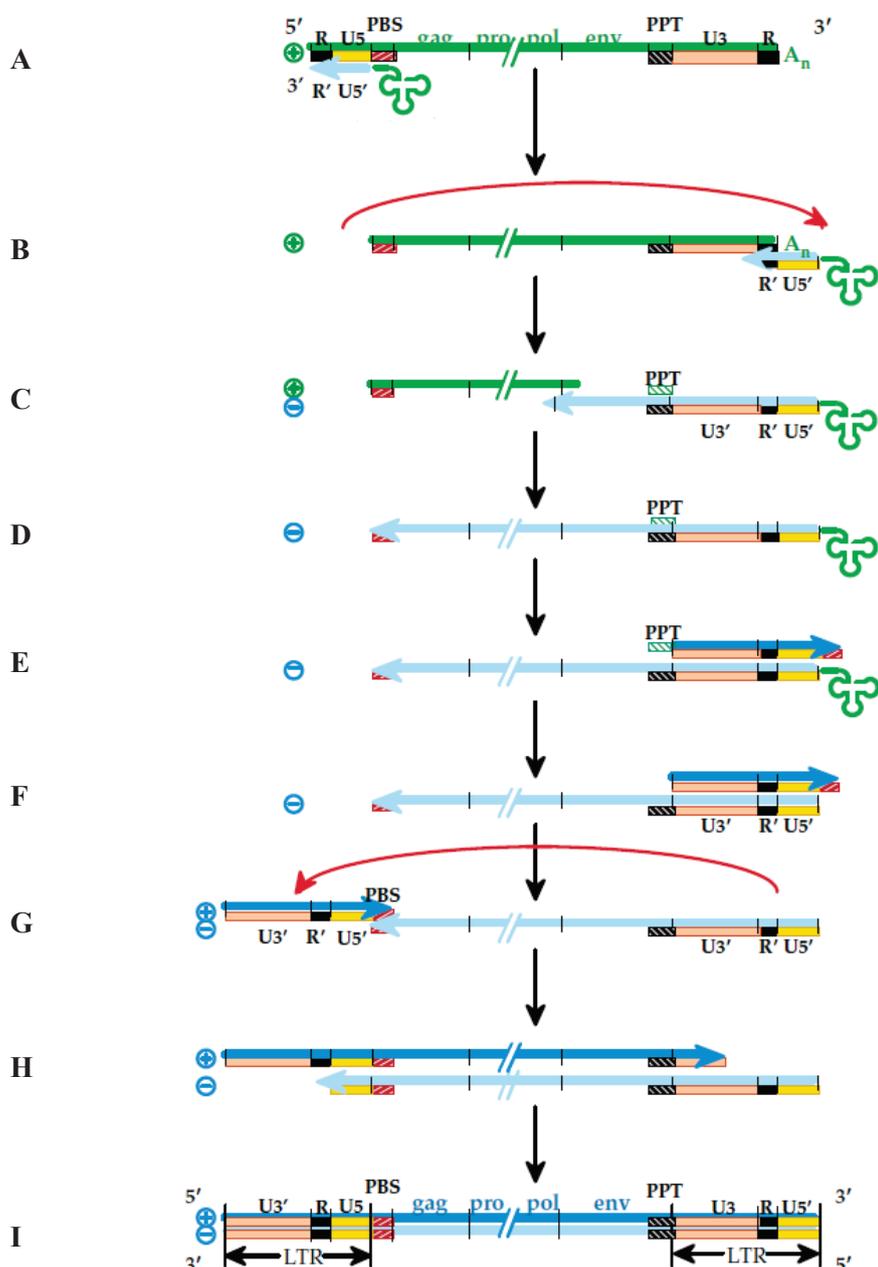
Chaque rétrovirus a un tropisme cellulaire spécifique. Dans le cas du VIH-1, les cellules infectées sont les cellules qui expriment le récepteur CD4, il s'agit des lymphocytes T CD4<sup>+</sup>, des macrophages et de certaines cellules dendritiques (Crowe *et al.*, 1987 ; Romani *et al.*, 1994). Le virus pénètre dans la cellule hôte en formant un complexe entre les glycoprotéines virales SU (gp120) et les récepteurs cellulaires CD4 avec les corécepteurs CCR5 ou CXCR4. Cela permet l'attachement de la particule virale à la surface de la cellule hôte (Arnolds et Spencer, 2014). La formation de ce complexe permet d'activer les glycoprotéines virales TM (gp41) qui conduisent à la fusion directe de l'enveloppe virale avec la membrane plasmique (Engelman et Cherepanov, 2012). La capsid est ainsi libérée dans le cytoplasme.

#### III.3.2. La transcription inverse

La transcription inverse débute dans le cytoplasme au sein de la capsid rétrovirale et se termine dans le noyau. Cette étape est catalysée par la RT. La RT est active sous la forme d'un hétérodimère entre la protéine entière (appelée p66) et la protéine sans le domaine RNase H (appelée p51).

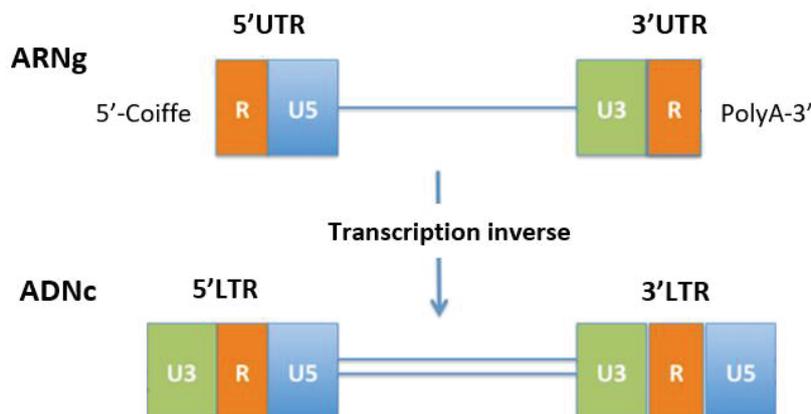
La transcription inverse débute par l'hybridation d'un ARNt spécifique au niveau de l'élément PBS dans la séquence U5 du 5'UTR de l'ARNg (**figure 17.A**) (Strauss et Strauss, 2008). A partir de cette amorce ARN, la RT prolonge le brin et produit un ADN contenant les séquences U5 et R. La synthèse du brin s'arrête lorsque l'extrémité 5' de l'ARNg est atteinte. Par la suite, ce brin nouvellement synthétisé est transféré de l'extrémité 5' de l'ARNg à l'extrémité 3' par hybridation à l'élément R du 3'UTR (**figure 17.B**). Grâce à l'activité RNase H de la RT, le brin matrice d'ARNg en duplex avec l'ADNc est dégradé au cours de la synthèse du brin d'ADNc (**figure 17.C**). La synthèse du brin d'ADN simple brin (ADNsb) de polarité négative est arrêtée lorsque l'extrémité 5' de l'ARNg est atteinte (**figure 17.D**).

La synthèse du second brin d'ADN (ADNsb de polarité positive) est réalisée à partir de l'ADNsb de polarité négative nouvellement synthétisé par la RT. Pour cela, une autre amorce ARN est utilisée, il s'agit d'une séquence ARN nommée PPT (« *polypurine tract* »). La séquence PPT s'hybride en 5' de la séquence U3 du 3'LTR sur l'ADNsb matrice et permet la synthèse d'un ADNsb (**figure 17.E**). La synthèse de l'ADN s'arrête lorsque l'extrémité 5' de l'ADN matrice est atteinte. Ensuite, l'activité RNase H de la RT entraîne la dégradation de l'ARNt et de la séquence PPT (**figure 17.F**). Par la suite, l'ADN nouvellement synthétisé est transféré à l'extrémité 3' de l'ADN matrice grâce à l'hybridation des éléments PBS présents dans les séquences U5 de l'ADN naissant et de l'ADN matrice (**figure 17.G**). Les brins d'ADN sont complétés aboutissant à la synthèse d'un ADNdb (**figure 17. H et I**).



**Figure 17.** Schéma du processus de transcription inverse. Les lignes vertes représentent l'ARN ; les lignes bleu clair représentent l'ADNs<sub>b</sub> de polarité négative ; les lignes bleu foncé représentent l'ADNs<sub>b</sub> de polarité positive. Source : Strauss et Strauss, 2008.

Ainsi, les régions non codantes dans l'ARN<sub>g</sub> et dans l'ADN<sub>c</sub> sont différentes (**figure 18**). Comme décrit précédemment, le 5' UTR contient une séquence R et une séquence U5, et le 3' UTR contient une séquence U3 et une séquence R. Tandis que l'ADN<sub>c</sub> qui est plus grand que l'ARN<sub>g</sub>, contient à la place des régions UTR, des régions LTR qui contiennent chacune une séquence R, U3 et U5.



**Figure 18.** Représentation schématique des régions non codantes dans l'ARN<sub>g</sub> du VIH-1 et dans l'ADN<sub>c</sub>. Source : Zhang et Crumpacker, 2022.

### III.3.3. L'intégration

La capsid est transportée dans le cytoplasme grâce au réseau de microtubules et entre dans le noyau via les NPC (Zila *et al.*, 2021 ; Shen *et al.*, 2021). Elle se désassemble ensuite et l'IN catalyse l'intégration de l'ADN<sub>db</sub> dans le génome de la cellule hôte (voir **chapitre IV.1.6.1**). L'insertion se fait dans des sites préférentiels ou aléatoirement selon le genre rétroviral.

Dans le cas du VIH-1, les sites d'intégration se trouvent dans des gènes fortement exprimés (Mitchell *et al.*, 2004 ; Schröder *et al.*, 2002). Cette insertion préférentielle est permise grâce à l'interaction de l'IN avec des facteurs cellulaires, comme LEDGF/p75 (« *lens epithelium-derived growth factor* ») et INI1 (« *IN interactor 1* ») dans le cas du VIH-1 (Kvaratskhelia *et al.*, 2014 ; Kalpana *et al.*, 1994). INI1 lie l'IN du VIH-1 ce qui stimule l'intégration médiée par l'IN (Kalpana *et al.*, 1994) ou peut l'inhiber en fonction de la concentration d'IN (Das *et al.*, 2009). D'autres facteurs cellulaires permettent également l'intégration provirale à des sites spécifiques comme HDGF2 (« *Hepatoma-derived growth factor like 2* ») (Passos *et al.*, 2021) et CPSF6 (« *cleavage and polyadenylation specificity factor subunit 6* ») (Achuthan *et al.*, 2019).

LEDGF/p75 est un co-activateur transcriptionnel de 60 kDa qui appartient à la famille des facteurs de croissances dérivés d'hépatome (HDGF, « *hepatoma-derived growth factor* ») (Ge *et al.*, 1998 ; Shinohara *et al.*, 2002). LEDGF/p75 est une protéine nucléaire impliquée dans la régulation génique et la réponse cellulaire aux stress. La liaison de l'IN à ce facteur cellulaire permet de stimuler les activités enzymatiques de l'IN, qui sont le clivage 3' et l'intégration concertée, détaillées dans le **chapitre IV.1.6.1** (Cherepanov *et al.*, 2003). LEDGF/p75 possède un domaine de liaison à l'IN (IBD, « *IN-binding domain* ») en C-terminal (Cherepanov *et al.*, 2004). Ce domaine d'environ 80 résidus d'acide aminé est suffisant et nécessaire pour la liaison de LEDGF/p75 à l'IN, il n'est cependant pas unique à ce facteur cellulaire et est partagé par d'autres membres de la même famille protéique que LEDGF/p75.

L'IN interagit également avec les extrémités des histones pour permettre une intégration optimale du génome viral dans la chromatine hôte (Benleulmi *et al.*, 2017 ; Mauro *et al.*, 2019). L'IN du VIH-1 lie

préférentiellement l'extrémité de l'histone H4 *in vitro* et un hétérodimère d'histone H2A-H2B (Maskell *et al.*, 2015). De plus, l'IN interagit avec différents partenaires cellulaires qui favorisent l'intégration du VIH-1, comme les protéines HMGA1 (« *high-mobility group protein A1* ») (Farnet et Bushman, 1997) et les protéines Barrière à l'Auto-intégration ou BAF (Chen et Engelman, 1998 ; Lin et Engelman, 2003).

La forme intégrée du virus est appelée provirus, elle est stable dans le temps. Le génome ADN intégré dans la chromatine de l'hôte est appelé ADN proviral.

### III.3.4. La transcription provirale

Mes travaux de thèse portent sur l'implication de l'IN dans la transcription provirale et pour cela cette partie est plus détaillée.

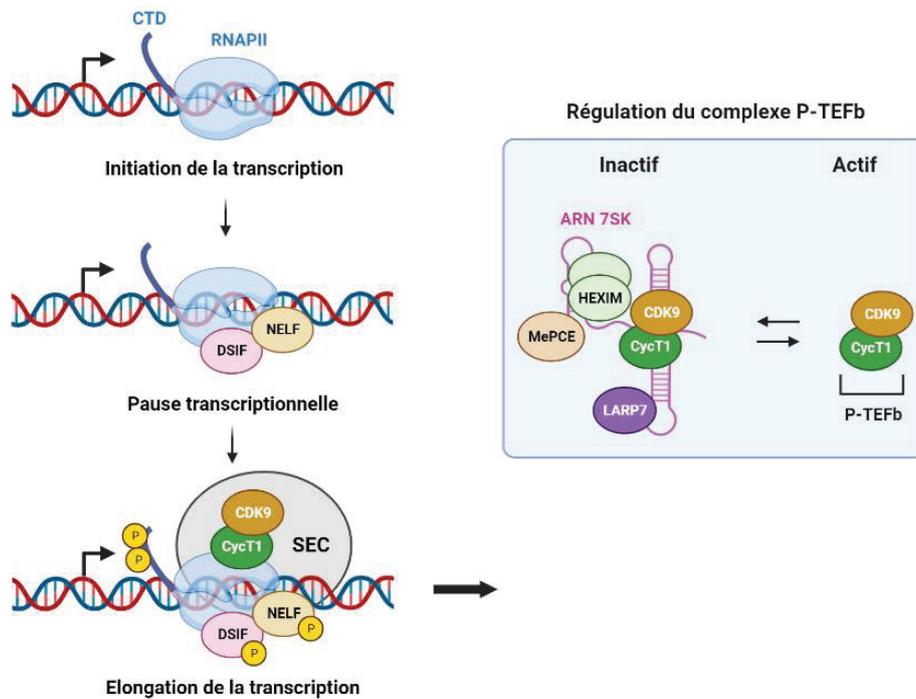
#### III.3.4.1. L'élongation de la transcription dans un contexte cellulaire sain

Dans un contexte sain (sans infection rétrovirale), la transcription cellulaire est réalisée grâce à plusieurs facteurs cellulaires qui permettent la transcription par l'ARN polymérase II (RNAPII, « *RNA polymerase II* »). Ces facteurs favorisent la bonne conformation de la RNAPII pour permettre l'initiation, puis l'élongation et enfin la terminaison de la transcription. L'étape limitante dans le cas de l'infection virale est l'élongation de la transcription, le blocage cellulaire de cette étape est contournée par le VIH-1 grâce à la protéine régulatrice Tat qui recrute le complexe cellulaire P-TEFb.

P-TEFb est un hétérodimère composé de la kinase cycline-dépendante 9 (CDK9, « *cyclin-dependent kinase 9* ») et de la cycline T1 ou 2. Ce complexe joue un rôle clé dans l'élongation de la transcription cellulaire en activant la RNAPII. En effet, après l'initiation de la transcription, la RNAPII marque une pause transcriptionnelle suite à la liaison de deux facteurs d'élongation négatifs : DISF (« *DRB sensitivity-inducing factor* ») et NELF (« *negative elongation factor* ») (Muniz *et al.*, 2012) (**figure 19**). Le recrutement de P-TEFb au niveau du super complexe d'élongation (SEC, « *super elongation complex* »), induit la phosphorylation par la CDK9 du domaine carboxy-terminal de la RNAPII sur le résidu Ser2, ainsi que la phosphorylation des facteurs d'élongation négatifs, permettant la dissociation de ces facteurs du complexe et ainsi l'activation de la RNAPII et l'élongation de la transcription.

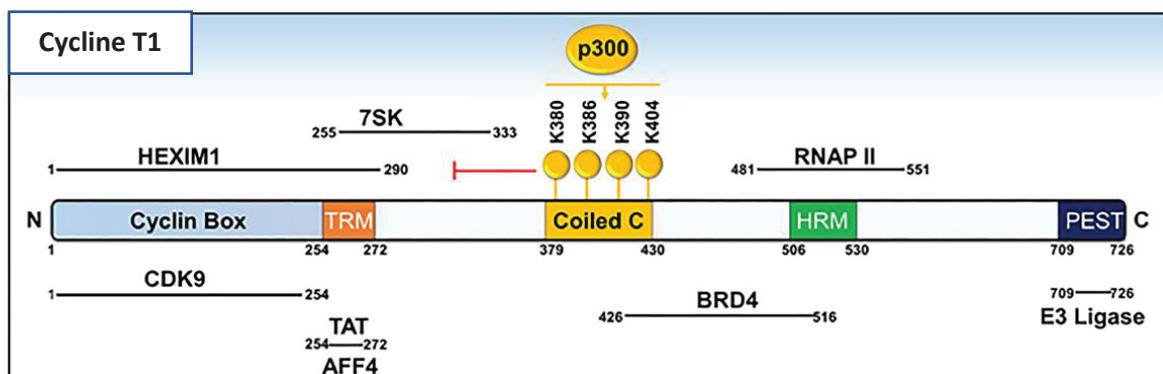
La régulation de l'élongation de la transcription est permise en régulant l'activité de P-TEFb par le biais d'un ARN nucléaire non codant, l'ARN 7SK (**figure 19**). 7SK est un ARN structuré possédant des structures en tige-boucle, il contient 331 nucléotides (Mousseau et Valente, 2018). Cet ARN en complexe avec différentes protéines forme une ribonucléoprotéine (RNP) appelée 7SK snRNP (7SK snRNP, « *7SK small nuclear ribonucleoprotein* »). Les protéines impliquées sont LARP7 (« *La ribonucleoprotein domain family member 7* »), MePCE (« *methylphosphate capping enzyme* ») et HEXIM1/2 (« *hexamethylene bis-acetamide inducible 1 or 2* ») :

- LARP7 protège l'ARN 7SK de la dégradation en le liant à l'extrémité 3'.
- MePCE stabilise l'ARN 7SK en le liant et lui ajoutant une coiffe méthyl-phosphate à l'extrémité 5'.
- HEXIM sous forme de dimère se lie à une structure en tige-boucle située à l'extrémité 5' de l'ARN 7SK, entraînant un changement conformationnel de la protéine HEXIM. Ce changement conformationnel permet l'interaction directe de HEXIM avec la cycline T1. Cette liaison conduit à la séquestration du complexe P-TEFb et donc à l'inactivation de l'activité kinase du complexe P-TEFb. La majorité des complexes P-TEFb cellulaires sont sous forme inactive. Un complexe P-TEFb peut être libéré selon les stimuli cellulaires, par exemple par la protéine Tat et l'ARN TAR dans le cas d'une infection par le VIH-1. La libération du complexe P-TEFb conduit à son activation et permet ensuite une cascade de phosphorylation médiée par la CDK9 qui aboutit à l'activation de la RNAPII et à l'élongation de la transcription.



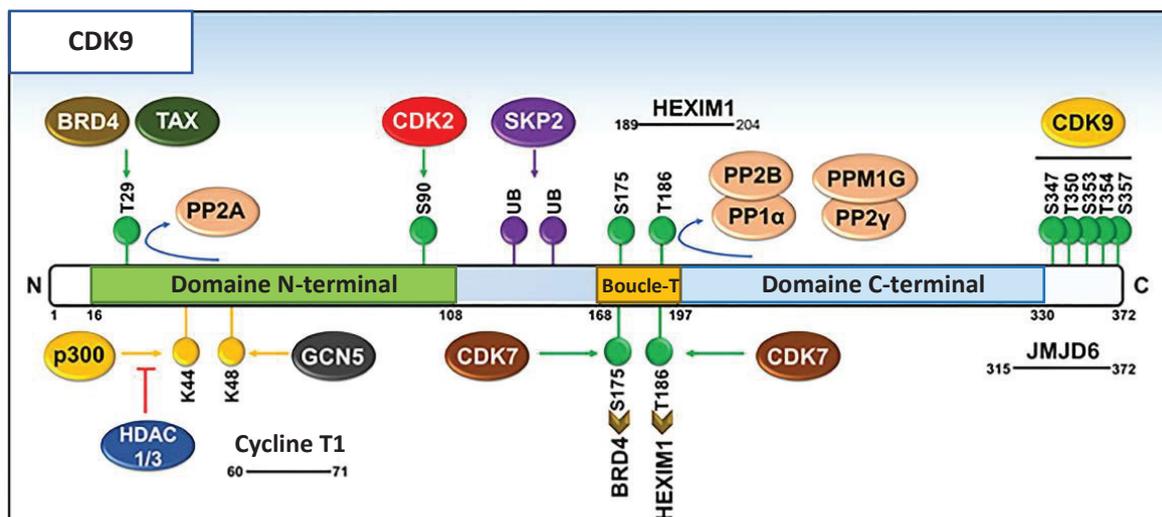
**Figure 19.** Schéma du mécanisme d'activation de l'élongation de la transcription cellulaire. Source : Muniz et al., 2012. Illustration créée avec BioRender.

La cycline T1 contient plusieurs domaines (Anshabo *et al.*, 2021) (**figure 20**) avec en premier un domaine N-terminal commun aux cyclines (aa 1-263), composé de deux boîtes cycline (« *cyclin box* »), qui participent dans la liaison avec les protéines cellulaires CDK9 et HEXIM, et à la protéine rétrovirale Tat. Ces deux boîtes cycline s'organisent en deux paquets de cinq hélices  $\alpha$ . Dans chaque paquet, certaines hélices entourent des hélices centrales de façon antiparallèle. Ensuite, il y a un domaine TRM (« *Tat-TAR recognition motif* ») (aa 254-272) qui est impliqué dans la liaison au complexe Tat:TAR et au complexe HEXIM:7SK. Puis, une région « *coiled coil* » (aa 379-430) qui est modifiée post-traductionnellement par une histone acétyltransférase, p300 ; suivie d'un domaine HRM (« *histidine-rich motif* ») (aa 506-530) riche en histidine qui participe à la liaison de la RNAPII. A l'extrémité C-terminale de la cycline T1 se trouve une séquence PEST (Pro-Glu-Ser-Thr) (aa 709-769) qui participe à la liaison avec une E3 ligase.



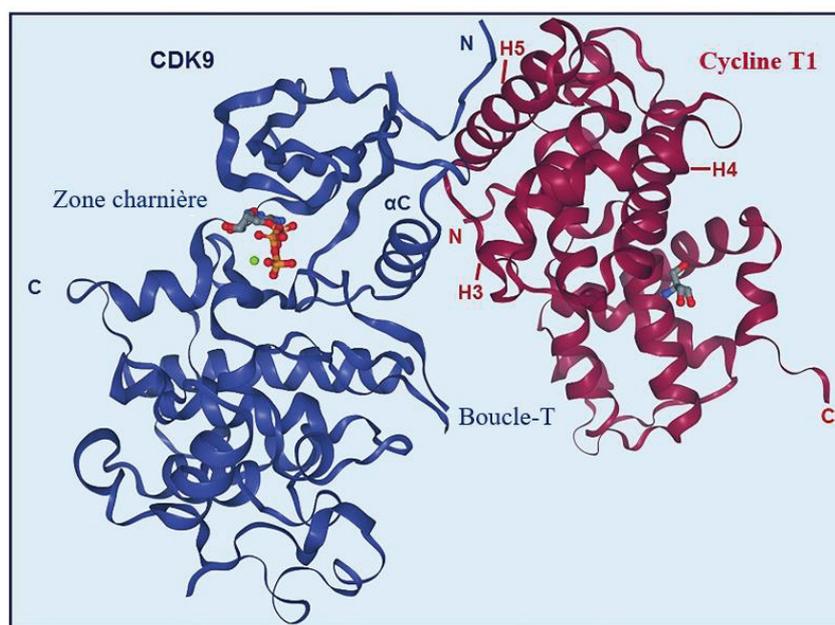
**Figure 20.** Représentation de l'organisation en domaines de la cycline T1. Source : Anshabo et al., 2021.

La CDK9 contient également plusieurs domaines (Anshabo *et al.*, 2021) (**figure 21**) dont un domaine N-terminal (aa 16-108) qui comprend cinq brins  $\beta$  et une hélice  $\alpha$ . L'interaction entre la CDK9 et la cycline T1 se fait majoritairement via cette hélice  $\alpha$ . Et un domaine C-terminal (aa 109-330) qui contient sept hélices  $\alpha$  et quatre brins  $\beta$ .



**Figure 21.** Représentation de l'organisation en domaines de la CDK9. Source : Anshabo *et al.*, 2021.

La cycline T1 et la CDK9 interagissent ensemble pour former le complexe P-TEFb via l'hélice  $\alpha$ C et le brin  $\beta$ 4 de la CDK9 et les hélices H3, H4 et H5 de la cycline T1 (**figure 22**).



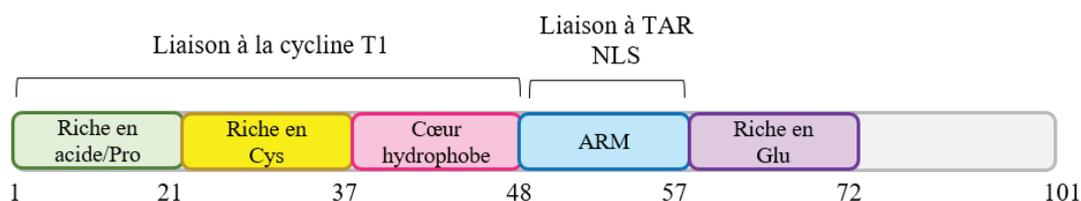
**Figure 22.** Structure du complexe CDK9 (aa 1-330):cycline T1 (aa 1-259) en présence d'ATP (Baumli *et al.*, 2008 ; code PDB : 3BLQ). Source : Anshabo *et al.*, 2021.

### III.3.4.2. La protéine Tat du VIH-1

La protéine trans-activatrice de la transcription, Tat, est une protéine intrinsèquement désordonnée de 99 à 103 résidus d'acide aminé ; toutefois une version plus courte est aussi fonctionnelle (Jeang *et al.*, 1988 ; Jeang *et al.*, 1999 ; Mousseau et Valente, 2018 ; Clark *et al.*, 2017). Les 72 premiers acides aminés sont le module minimum requis pour permettre une transcription médiée par Tat (Mousseau et Valente, 2018 ; Clark *et al.*, 2017). Ce module est composé de cinq domaines (**figure 23**) avec : en N-terminal un domaine contenant un motif riche en acide/proline (aa 1-21) ; puis un motif doigt à zinc/ domaine riche en cystéine (aa 22-37) ; un domaine central hydrophobe (aa 38-48) ; un motif basique riche en arginine (ARM, « *arginine-rich motif* ») (aa 49-57) ; enfin un motif riche en acide glutamique (aa 59-72).

Les trois premiers domaines (aa 1-48) constituent le domaine minimum d'activation. Ils permettent la liaison à la cycline T1, tandis que le domaine ARM est le domaine de liaison à l'ARN. Ce domaine contient également un signal de localisation nucléaire (NLS, « *nuclear localization signal* »).

Les quatre premiers domaines (aa 1-57) permettent la majorité des interactions de Tat avec des partenaires cellulaires. La liaison de Tat à ces partenaires permet de structurer la protéine dans des conformations spécifiques qui présentent une grande affinité selon le partenaire concerné (Tahirov *et al.*, 2010). Le domaine C-terminal de Tat n'est pas critique pour la fonction de transactivation de la transcription de la protéine, mais il joue un rôle dans la pathogenèse en interférant avec la réponse immunitaire.

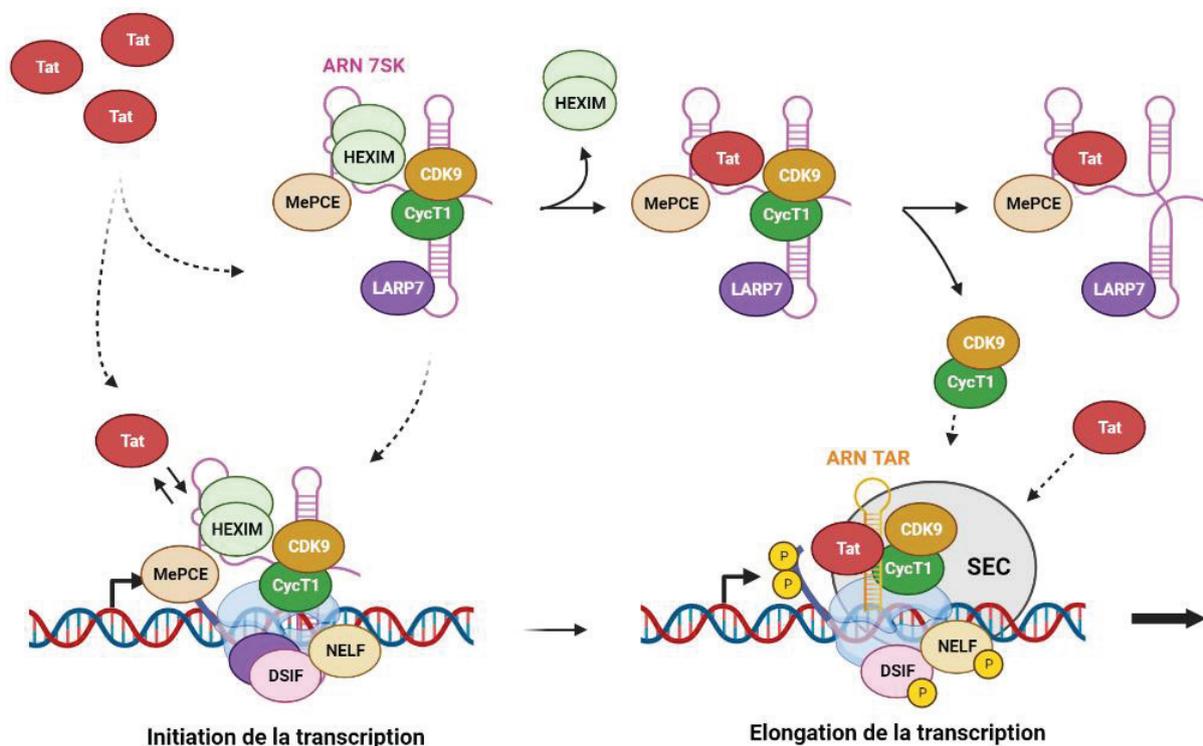


**Figure 23.** Représentation schématique de l'organisation en domaines de la protéine rétrovirale Tat. Sources : Mousseau et Valente, 2018 ; Clark *et al.*, 2017.

La transcription du provirus permet en premier de produire l'ARN TAR situé dans le 5' LTR. Puis, la RNAPII fait une pause transcriptionnelle comme détaillé dans la partie précédente. Le VIH-1 est capable de détourner le complexe P-TEFb et de permettre une élongation productive. Pour cela, le complexe P-TEFb peut être recruté au niveau du promoteur du VIH-1 par la protéine Tat (Dingwall *et al.*, 1989 ; Dingwall *et al.*, 1990 ; Churcher *et al.*, 1993) (**figure 24**). Tat et HEXIM lient la cycline T1 en N-terminal sur le même site de liaison, ce qui entraîne une compétition entre les deux protéines lors d'une infection virale. La protéine AFF1 ou AFF4 (« *AF4/FMR2 family member 1 or 4* ») fait partie du complexe SEC et favorise sous sa forme entière la liaison de la protéine Tat au complexe P-TEFb en augmentant l'affinité de Tat pour le complexe (Lu *et al.*, 2013). Ainsi, Tat entraîne la dissociation de HEXIM du complexe 7SK snRNP, ce qui conduit à la libération du complexe P-TEFb piégé au sein du complexe 7SK snRNP (Mancebo *et al.*, 1997 ; Quaresma *et al.*, 2016 ; Lu, Li *et al.*, 2013). D'autre part, la protéine Tat peut également se lier directement à l'ARN 7SK sur le même site de liaison que la protéine HEXIM. Ainsi les deux protéines entrent en compétition pour la liaison à l'ARN 7SK, ce qui conduit au remplacement de la protéine HEXIM par la protéine Tat au sein des complexes 7SK snRNP (Muniz *et al.*, 2010 ; Sedore *et al.*, 2007 ; Sobhian *et al.*, 2010). Dans ces deux cas, le complexe P-TEFb:Tat:AFF1 s'intègre au sein du complexe SEC et lie l'ARN TAR, ce qui stimule l'élongation de la transcription.

Un troisième mécanisme conduisant également à l'élongation de la transcription repose sur le recrutement du complexe 7SK snRNP avec le complexe P-TEFb inactivé et la protéine HEXIM au sein de la RNP au niveau du promoteur du VIH-1, où il s'associe au PIC (« *pre-integration complex* »)

(D'Orso et Frankel, 2010 ; Zhou et Yik, 2006). Par la suite, l'interaction de la protéine Tat avec l'ARN TAR conduit à la dissociation du complexe 7SK snRNP libérant le complexe P-TEFb. Dans ce mécanisme, il y a une compétition entre les deux ARN pour la liaison du complexe Tat:P-TEFb.



**Figure 24.** Schéma du mécanisme d'activation de l'élongation de la transcription médiée par la protéine Tat. Sources : Mousseau et Valente, 2018 ; Muniz et al., 2012. Illustration créée avec BioRender.

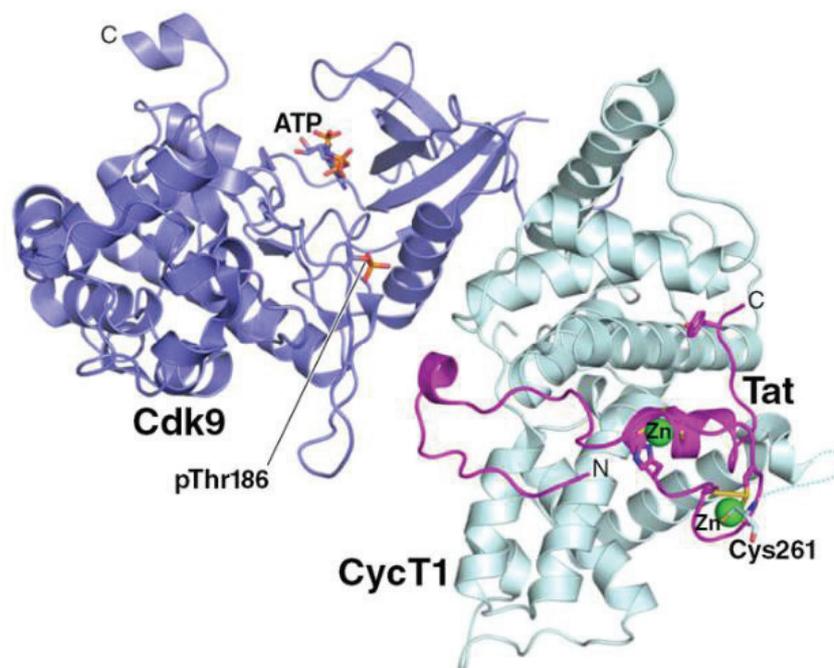
La protéine Tat recrute également d'autres facteurs cellulaires afin : d'améliorer l'activité de la RNAPII grâce à la protéine PAF1 (« *polymerase-associated factor complex 1* ») ; de permettre un remodelage de la chromatine grâce au recrutement des histones acétyltransférase et du complexe SWI/SNF (« *SWItch/Sucrose Non-Fermentable* »), ce qui favorise un environnement où la chromatine est sous forme d'euchromatine et donc une augmentation de la transcription cellulaire (Mousseau et Valente, 2018 ; Clark et al., 2017).

Le mécanisme expliquant la présence de Tat au début de la transcription provirale n'est pas encore élucidé. En effet, Tat n'est pas présente dans la particule virale lors de l'infection d'une cellule hôte par le VIH-1. L'origine de Tat pourrait s'expliquer par une fuite transcriptionnelle qui permet de produire les protéines rétrovirales en faible concentration, puis Tat permettrait d'augmenter le rendement transcriptionnel du provirus (Rampalli et al., 2003). En effet, Tat est sécrétée par les cellules infectées, conduisant à une accumulation de protéines Tat dans le milieu extracellulaire, entraînant une internalisation de Tat au sein des cellules voisines et un adressage de la protéine au noyau grâce à sa séquence NLS (Clark et al., 2017).

### III.3.4.3. Structures du complexe TAR:Tat:P-TEFb:AFF4

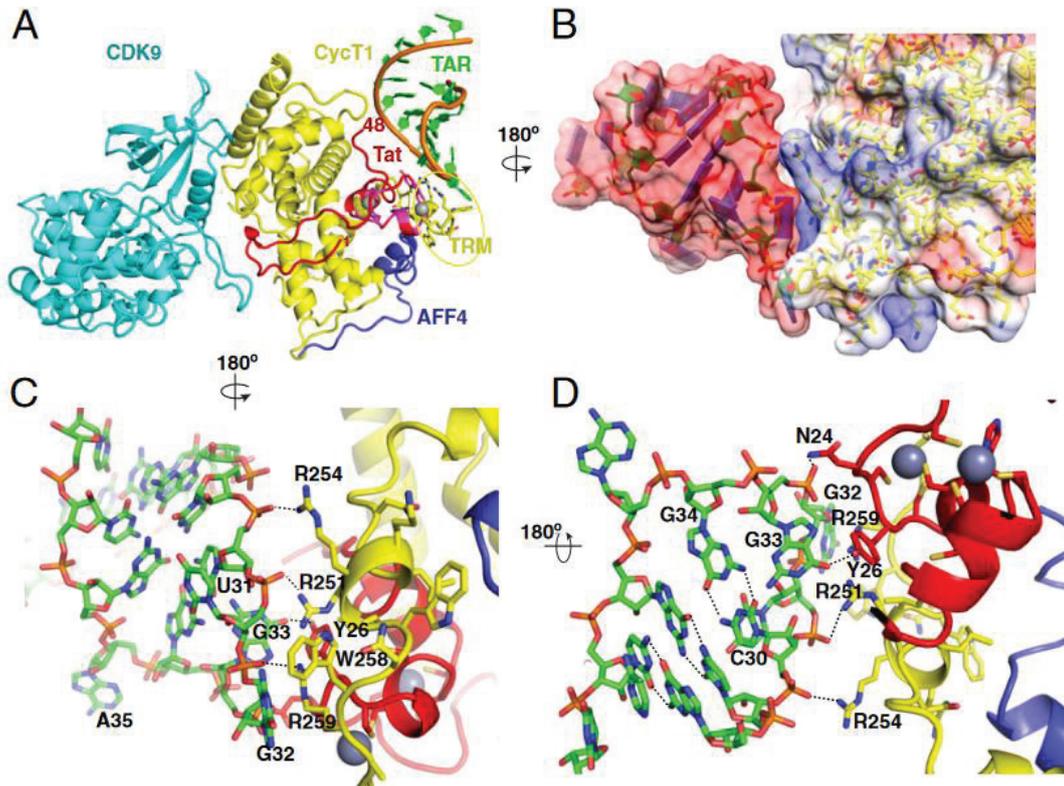
Les structures obtenues ces dix dernières années ont montré comment le domaine de liaison désordonné d'AFF4 se replie et lie le sillon hydrophobe à la surface de la cycline T1 (Schulze-Gahmen et al., 2013 ; Chou et al., 2012). AFF4 agit comme un « *scaffold* » (échafaudage) flexible qui permet le recrutement

de plusieurs facteurs d'élongation de la transcription au niveau du complexe SEC (Schulze-Gahmen *et al.*, 2013 ; Chou *et al.*, 2012). Le motif ARM de Tat lie un motif particulier de la cycline T1, le motif TRM qui comprend le résidu Cys261 et qui est critique pour l'interaction de la cycline T1 avec Tat en présence d'ions  $Zn^{2+}$  (Tahirov *et al.*, 2010) (**figure 25**). Ces deux motifs interagissent également avec une hélice d'AFF4 (Schulze-Gahmen *et al.*, 2014 ; Tahirov *et al.*, 2010 ; Gu *et al.*, 2014) (**figure 26**). Ces différents sites d'interaction sont flexibles lorsqu'ils sont libres et ils se structurent pour former une surface stable chargée positivement lors de l'interaction qui permet la liaison à l'ARN.



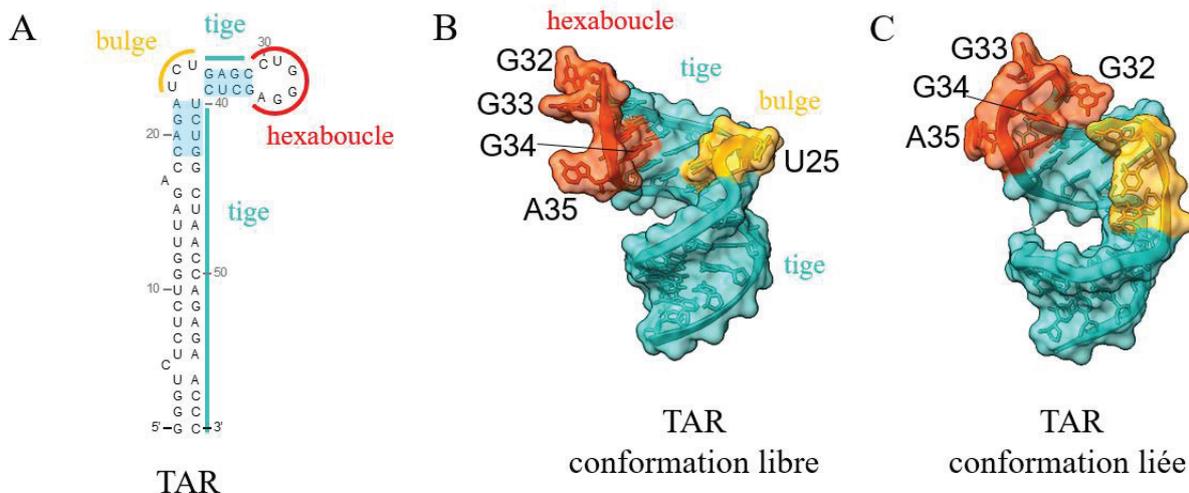
**Figure 25.** Mise en évidence des interactions entre Tat et le résidu Cys261 de la cycline T1 en présence d'ions  $Zn^{2+}$ . Représentation en rubans du complexe Tat (aa 1-86) (violet):cycline T1 (aa 1-266) (bleu pâle):CDK9 (aa 1-345) (bleu foncé) (Tahirov *et al.*, 2010 ; code PDB : 3MIA). Source : Ott *et al.*, 2011.

La structure résolue du complexe Tat:AFF4:P-TEFb lié à l'ARN TAR met en évidence les interactions mises en jeu entre la boucle de TAR, le motif TRM de la cycline T1 et la boucle de Tat coordonnant des ions  $Zn^{2+}$  (Garber *et al.*, 1998 ; Schulze-Gahmen *et al.*, 2016 ; Schulze-Gahmen et Hurley, 2018) (**figures 25 et 26**). Dans cette interaction, la boucle de TAR est stabilisée par des liaisons hydrogènes entre les nucléotides C30 et G34 et des contacts additionnels avec le nucléotide G33. Les nucléotides G32, A35 et U31 de la boucle de TAR s'orientent dans une conformation ouverte. Cette structuration de la boucle de TAR permet d'établir des contacts directs entre les bases des nucléotides G32 et G33 avec les résidus Trp258 de la cycline T1 et Tyr26 de Tat. D'autres contacts sont établis entre le squelette phosphate-ribose des nucléotides G32 et G33 avec des résidus du motif TRM de la cycline T1 (Arg251, Arg254 et Arg259).



**Figure 26.** Mise en évidence des interactions entre l'ARN TAR (nt 21-40) et le complexe Tat (aa 1-57):AFF4 (aa 32-67):P-TEFb. **A.** Représentation en rubans du complexe TAR (vert):Tat (rouge):AFF4 (bleu foncé):cycline T1 (jaune):CDK9 (turquoise). **B.** Représentation de la surface de TAR (à gauche) et de celle de la cycline T1 (à droite). **C et D.** Représentations en rubans et bâtonnets des contacts mis en place au sein du complexe. Source : Schulze-Gahmen et Hurley, 2018 ; code PDB : 6CYT.

L'ARN TAR est un ARN structuré avec une longue tige, suivi d'un « *bulge* » (renflement) de trois bases, d'une petite tige et d'une hexaboucle (**figure 27**). La séquence de TAR est fortement conservée au sein du génome viral (Chavali *et al.*, 2019). Cet ARN adopte deux conformations différentes selon s'il est libre ou lié à un partenaire, dans ce cas la protéine Tat.



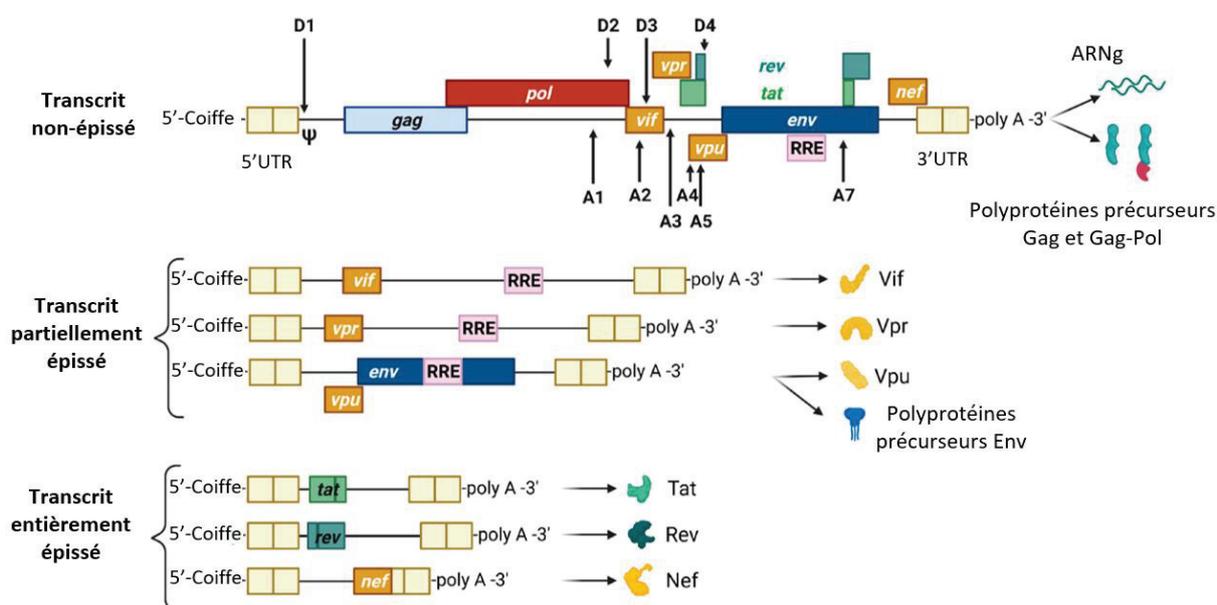
**Figure 27.** Changements de conformation de l'ARN TAR en absence et en présence de la protéine Tat (aa 40-59). **A.** La séquence de l'ARN TAR du VIH-1. **B.** Structure tridimensionnelle de TAR (nt 17-45) en conformation libre (Aboul-Ela *et al.*, 1996 ; code PDB : 1ANR). **C.** Structure tridimensionnelle de

TAR (nt 18-44) en conformation liée (Belashov *et al.*, 2018 ; code PDB : 6CMN). Sources : Chavali *et al.*, 2019 ; Rocchi *et al.*, 2022.

### III.3.4.4. Les mécanismes d'épissage, d'addition d'une coiffe et de polyadénylation

L'ADN proviral contient neuf séquences codantes qui se chevauchent (van Heuvel *et al.*, 2022) (**figure 28**). L'épissage alternatif permet de générer trois différents ARNm codant pour l'ensemble des protéines rétrovirales : un ARNg de 9,7 kb qui ne subit pas de processus d'épissage codant pour les polyprotéines précurseurs Gag et Gag-Pol ; un ARNm de 4,5 kb avec quelques introns ayant subi un épissage partiel codant pour les polyprotéines précurseurs Env et les protéines accessoires (Vif, Vpr, Vpu) ; et un ARNm entièrement épissé de 1,8 kb codant pour les protéines régulatrices (Tat, Rev, Nef) (Sertznig *et al.*, 2018 ; Kammler *et al.*, 2006).

Ces transcrits possèdent une coiffe à l'extrémité 5' et une séquence polyA à l'extrémité 3'. Ces modifications post-transcriptionnelles sont réalisées grâce à des mécanismes cellulaires qui reconnaissent des séquences régulatrices dans les LTR (Strauss et Strauss, 2008 ; Karn et Stoltzfus, 2012).



**Figure 28.** Représentation des transcrits du VIH-1 avec les sites d'épissage donneurs (D) et accepteurs (A). Les protéines codées par les ARNm sont représentées sur la droite. Source : van Heuvel *et al.*, 2022.

La protéine rétrovirale Rev permet de réguler l'épissage des ARNm et leur transport, en liant les éléments RRE (« *Rev response element* ») (King *et al.*, 2011 ; Strauss et Strauss, 2008). Un élément RRE contient 234 nucléotides et présente des structures secondaires en tige-boucle. De plus, Rev interagit avec des protéines cellulaires impliquées dans l'export nucléaire pour permettre le transport des ARNm du noyau dans le cytoplasme.

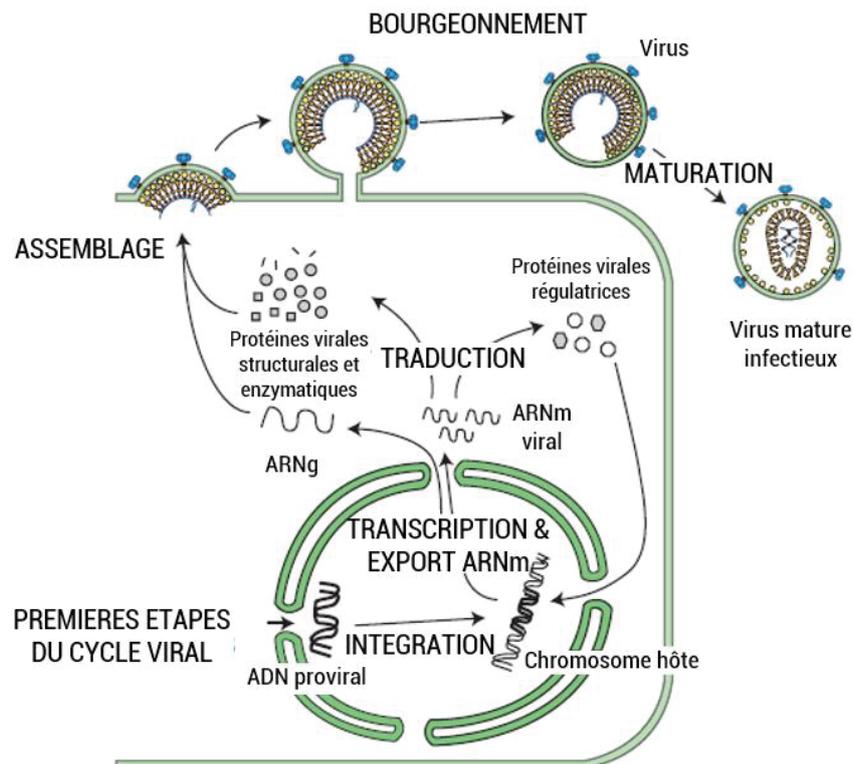
### III.3.5. La traduction et l'assemblage des particules

La traduction des ARNm codant pour les polyprotéines Env permet de produire des précurseurs des glycoprotéines qui constituent par la suite l'enveloppe de la particule virale (Strauss et Strauss, 2008). Ces précurseurs contiennent une séquence signal en N-terminal qui permet leur ciblage au réticulum endoplasmique (RE). Ils contiennent également une séquence d'ancrage à la membrane et un domaine

cytoplasmique en C-terminal. La séquence signal est clivée lors de la translocation des précurseurs dans le RE. Par la suite, les protéines sont glycosylées et transportées à la membrane plasmique où elles sont clivées par la furine en SU et TM.

Une partie des transcrits non-épissés est traduite par le mécanisme de traduction cellulaire en polyprotéines virales Gag et Gag-Pol (**figure 29**). Les deux polyprotéines sont obtenues par un phénomène de décalage du cadre de lecture ribosomale avec un rapport de production Gag:Gag-Pol de 20:1 (Shehu-Xhilaga *et al.*, 2001). Ces polyprotéines sont adressées dans les radeaux lipidiques de la membrane plasmique grâce aux myristoylations et aux acides aminés hydrophobes contenus dans les protéines MA (Bukrinskaya, 2007). Cela permet l'assemblage des particules virales au niveau de la membrane plasmique de la cellule hôte. Les polyprotéines Gag et Gag-Pol s'assemblent pour former une capsid sphérique.

Les transcrits non-épissés qui n'ont pas été traduits, qui correspondent à l'ARNg, sont incorporés dans les particules virales en cours d'assemblage. Les protéines NC au sein des polyprotéines Gag et Gag-Pol permettent de lier l'ARNg (Strauss et Strauss, 2008). Elles se lient au signal d'encapsidation sur l'ARNg, ce qui permet leur incorporation avec l'ARNg dans la particule virale. Les ARNm partiellement ou entièrement épissés ne contiennent pas la séquence  $\psi$  contrairement aux ARNg ce qui permet l'encapsidation sélective des ARNg dans les nouvelles particules virales (van Heuvel *et al.*, 2022). Chaque ARNg est recruté dans la capsid et dimérise ensuite grâce à l'élément ARN DIS du 5'UTR en conformation « *kissing* » (Brigham *et al.*, 2019).



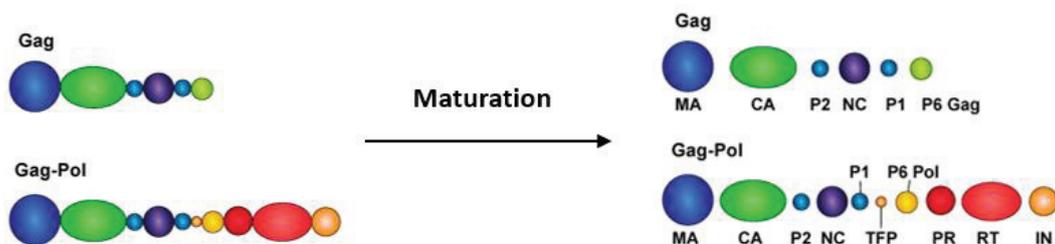
**Figure 29.** Schéma illustrant les étapes suivant l'intégration de l'ADN proviral du VIH-1. Source : Sundquist et Krausslich, 2012.

### III.3.6. Le bourgeonnement et la maturation des particules virales

Le bourgeonnement de la particule virale est réalisé grâce au détournement de la machinerie cellulaire ESCRT (« *endosomal sorting complex required for transport* ») qui permet la formation de « bourgeons » à la surface cellulaire, puis facilite la fission membranaire, entraînant la libération des particules virales (Sundquist et Krausslich, 2012 ; Morita et Sundquist, 2004 ; Hurley et Hanson, 2010).

Les polyprotéines Env sont incorporées dans les particules virales au cours du bourgeonnement grâce à leur interaction avec les protéines MA (Strauss et Strauss, 2008 ; Bukrinskaya, 2007).

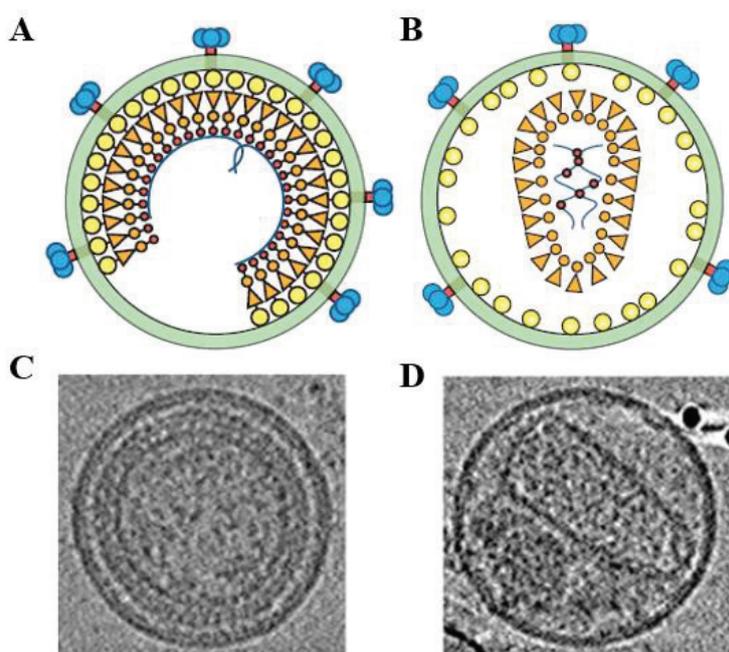
Les polyprotéines rétrovirales sont clivées grâce à la PR au sein des particules virales pendant le bourgeonnement (**figure 30**). Cette étape est requise pour que le virus soit infectieux, il s'agit de l'étape de maturation. La PR est active seulement sous la forme d'un homodimère, où chaque monomère apporte un acide aspartique dans le site actif (Strauss et Strauss, 2008). L'enzyme n'est pas active sous forme de monomère donc cela cause un délai dans le clivage des polyprotéines virales en protéines individuelles. Lorsque la concentration de polyprotéines Gag-Pol dans les particules virales est élevée, la PR dimérise et devient active (Strauss et Strauss, 2008).



**Figure 30.** Schéma illustrant l'étape de maturation des polyprotéines virales Gag et Gag-Pol en protéines individuelles catalysée par la PR. Source : Elliott et Kutluay, 2020.

Suite à l'action de la PR, la capside sphérique se réassemble sous forme conique dans le cas du VIH-1 et la particule virale s'organise comme décrit précédemment : une capside au centre de la particule virale qui protège l'ARNg en complexe avec la NC (Rossi *et al.*, 2021) (**figure 31**). La NC sert également de chaperonne pour l'ARNg (Sundquist et Krausslich, 2012). Après cette étape de maturation, la particule virale peut se répliquer dans une nouvelle cellule hôte.

La particule virale infectieuse contient deux copies de l'ARNg de polarité positive liées à des protéines de NC et les trois enzymes virales : PR, RT et IN. Tous ces éléments sont protégés par des protéines de CA, entourée d'une matrice. La particule virale est enveloppée et présente des glycoprotéines d'enveloppe (SU et TM) (Sundquist et Krausslich, 2012). SU et TM restent associées, ce qui permet à TM d'être active pour l'étape de fusion lors du prochain cycle réplcatif viral (Strauss et Strauss, 2008).



**Figure 31.** Représentations schématiques d'une particule virale immature (A) et mature (B) du VIH-1. Sections centrales d'une reconstruction d'une particule virale du VIH-1 immature (C) et mature (D) en microscopie électronique par analyse tomographique. Source : Sunquist et Krausslich, 2012.

### III.4. Traitement contre le VIH

#### III.4.1. Inhibiteurs rétroviraux

Les antirétroviraux sont utilisés afin de contrôler l'infection par le VIH. Ils permettent de réduire la charge virale à un taux non détectable et de restaurer le nombre de cellules CD4<sup>+</sup>, assurant le fonctionnement du système immunitaire du patient.

L'AZT est le premier médicament commercialisé contre le VIH-1. Il s'agit d'un inhibiteur nucléosidique de la RT (Strauss et Strauss, 2008). Lorsque la RT incorpore l'AZT dans l'ADNc en cours de synthèse, cela entraîne le blocage de la transcription inverse et de la réplication virale. Cependant des résistances sont rapidement apparues en monothérapie. Par la suite, une thérapie plus efficace a été mise au point, elle est appelée HAART (« *highly active antiretroviral therapy* »). Elle combine l'utilisation de plusieurs antirétroviraux, par exemple une trithérapie avec l'utilisation de deux analogues nucléosidiques différents et d'un inhibiteur de protéase (Shafer et Vuitton, 1999). Cette thérapie est très efficace chez les patients séropositifs et permet de réduire la charge virale en-dessous du seuil de détection (Strauss et Strauss, 2008). Cependant, l'arrêt du traitement conduit à l'augmentation de la charge virale, il s'agit donc d'un traitement à vie.

Les inhibiteurs rétroviraux ciblent différentes étapes du cycle réplcatif rétroviral et différentes protéines rétrovirales (Arts et Hazuda, 2012).

- L'entrée du virus est inhibée par le ciblage de l'interaction entre gp120 et le récepteur CD4 ; et celle entre gp41 et le corécepteur CCR5.
- La transcription inverse est inhibée par le ciblage de la RT grâce à deux classes d'inhibiteurs de la RT : les inhibiteurs nucléosidiques qui sont des analogues des substrats nucléosides natifs et les inhibiteurs non nucléosidiques qui se lient dans une poche allostérique non-catalytique de l'enzyme.
- L'intégration par le ciblage de l'IN avec les inhibiteurs du transfert de brin (INSTI, « *integrase strand transfer inhibitors* ») (voir **chapitre IV.3.1**).
- La maturation en inhibant la PR grâce à des inhibiteurs qui ciblent le site actif de la PR.

Depuis peu, une seconde génération d'inhibiteurs de l'IN est à l'étude. Il s'agit d'inhibiteurs allostériques (ALLINI, « *allosteric integrase inhibitors* ») qui entraînent la formation de VIH immatures, incapables d'infecter de nouvelles cellules (voir **chapitre IV.3.2**).

La prophylaxie pré-exposition vise à prévenir l'infection par le VIH chez des personnes séronégatives et exposées au virus. Elles suivent un traitement médicamenteux qui permet de réduire le risque de transmission du VIH par la voie sexuelle de 99% et par injection de 74% (ressource : <https://www.cdc.gov> ; Beyrer *et al.*, 2022).

#### III.4.2. Essais vaccinaux

Plusieurs essais vaccinaux ont été réalisés ces vingt dernières années. Seul le vaccin RV144 a été efficace contre le VIH-1 (ressource : <https://vih.org>). Ce vaccin permet une protection de 31% contre l'infection lors d'une stratégie vaccinale de pré-exposition. Ces résultats ont été confirmés lors d'études cliniques en phase III (Vaccari *et al.*, 2010 ; Zolla-Pazner *et al.*, 2021). Afin d'améliorer l'efficacité

vaccinale, de nouvelles stratégies vaccinales sont étudiées. Elles reposent sur l'identification des antigènes qui déclenchent une production d'anticorps neutralisants à spectre large ou sur la technologie de l'ARNm à l'instar des vaccins développés pendant la pandémie de COVID-19 (Fortner et Bucur, 2022).

Je vais maintenant détailler les deux protéines sur lesquelles j'ai travaillé pendant ma thèse, l'IN du VIH-1 et le facteur de restriction ISG20.

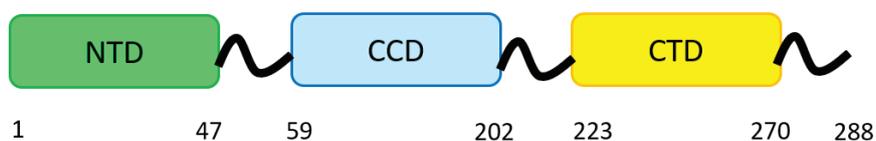
## IV. Intégrase

Comme de nombreuses protéines virales, l'IN est une protéine multifonctionnelle qui interagit avec de nombreux partenaires cellulaires et viraux à différentes étapes du cycle réplcatif viral. Elle comporte différents domaines reliés par des boucles flexibles et est susceptible de changer de conformation pour assurer ses différentes fonctions. Je me suis tout particulièrement intéressée à son rôle dans la morphogénèse et la transcription du VIH-1.

### IV.1. Structure de l'intégrase

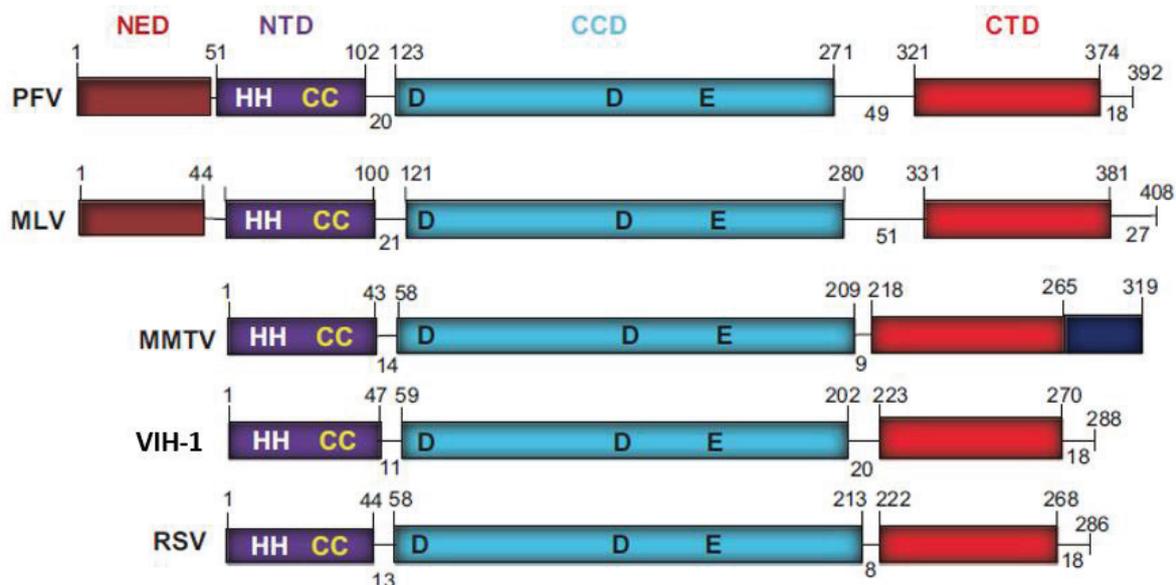
L'IN rétrovirale est une protéine modulaire de 280 à 400 résidus d'acide aminé contenant trois domaines reliés par des boucles flexibles : un domaine amino-terminal (NTD, « *N-terminal domain* »), un domaine catalytique central (CCD, « *catalytic core domain* ») et un domaine carboxy-terminal (CTD, « *C-terminal domain* »). Le CTD est suivi d'une extrémité C-terminale (CT) flexible de taille variable selon le genre rétroviral (Grandgenett et Aihara, 2018) (**figure 32**). De plus, les IN des genres simiispumavirus, gammarétrovirus et épsilon-rétrovirus possèdent un domaine additionnel en N-terminal, qui est appelé NED (« *N-terminal extension domain* ») (Passos *et al.*, 2021).

L'IN du VIH-1 contient 288 résidus d'acide aminé. Son NTD s'étend des résidus 1 à 47, son CCD de 59 à 202, son CTD de 223 à 270 et son extrémité CT de 271 à 288 (Grawenhoff et Engeman, 2017 ; Grandgenett et Aihara, 2018).



**Figure 32.** Représentation schématique de l'organisation des domaines de l'IN du VIH-1.

Les boucles flexibles inter-domaines sont de grande taille chez les simiispumavirus comme PFV et les gammarétrovirus. Elles sont plus petites chez les alpharétrovirus comme RSV, les betarétrovirus comme MMTV et les lentivirus comme VIH-1 (Grandgenett et Aihara, 2018) (**figure 33**). Ces boucles confèrent des différences d'assemblage chez les différentes IN rétrovirales (Passos *et al.*, 2021 ; voir **chapitre IV.1.6.2**).



**Figure 33.** Représentation schématique de l'organisation des domaines de l'IN chez cinq espèces caractéristiques de cinq genres de rétrovirus. Source : Grandgenett et Aihara, 2018.

La structure de l'IN a d'abord été étudiée par fragments dans les années 90, car les boucles inter-domaines rendaient la protéine entière trop flexible pour une étude par cristallographie aux rayons X (Rice *et al.*, 1996). L'IN du VIH-1 est de plus peu soluble et la plupart des résultats structuraux ont été obtenus sur des protéines mutées (Dyda *et al.*, 1994).

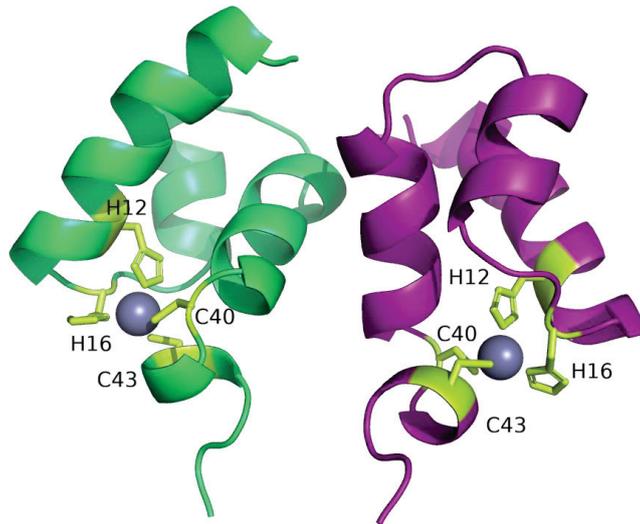
#### IV.1.1. Le domaine d'extension du NTD

Le NED est un domaine précédant le NTD. Il comprend environ 50 résidus d'acide aminé organisés en une hélice  $\alpha$  et trois brins  $\beta$  antiparallèles. Le NED participe au bon repliement du NTD et à l'intégration (Passos *et al.*, 2021).

#### IV.1.2. Le domaine amino-terminal

Le NTD comprend entre 40 et 50 résidus d'acide aminé. Il contient deux paires invariantes de résidus d'histidine-cystéine (HHCC) chélatant un ion  $Zn^{2+}$  et comporte un motif hélice-tour-hélice qui facilite la liaison à l'ADN (Andrake et Skalka, 2015 ; Zheng *et al.*, 1996).

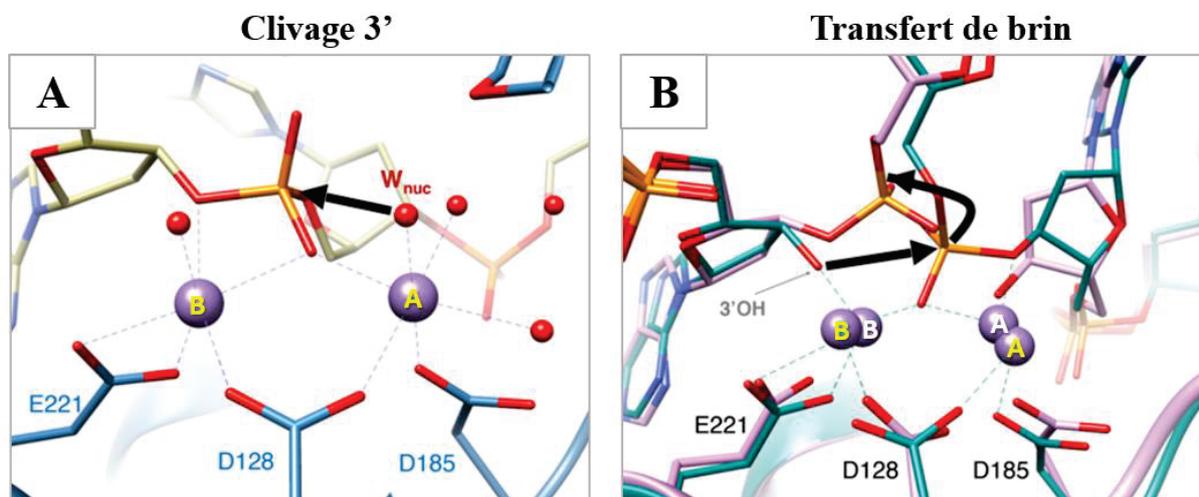
La structure du NTD du VIH-1 a été résolue en solution par RMN sous forme dimérique (Cai *et al.*, 1997) (**figure 34**). Chaque monomère contient quatre hélices, qui permettent de coordonner un atome de zinc grâce aux résidus suivants : His12, His16, Cys40 et Cys43.



**Figure 34.** Structure de l'interface dimérique du NTD de l'IN du VIH-1 obtenue par RMN en solution (Cai *et al.*, 1997 ; code PDB : 1WJC). La structure de chaque monomère est représentée en rubans (vert et violet). Les chaînes latérales des résidus d'acide aminé du motif HHCC sont représentées en bâtonnets. Les ions  $Zn^{2+}$  sont représentés par des sphères grises. Figure réalisée avec PyMOL.

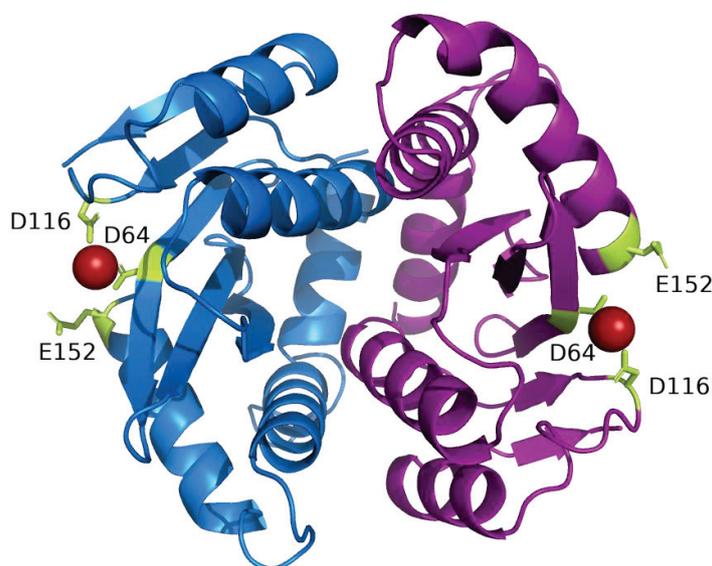
#### IV.1.3. Le domaine catalytique central

Le CCD contient environ 150 résidus d'acide aminé. Il est caractérisé par une triade catalytique de résidus invariants qui constituent un motif DD(35)E, les deux derniers résidus étant séparés de 35 résidus (Andrake et Skalka, 2015 ; Dyda *et al.*, 1994). Cette triade coordonne deux cations métalliques nécessaires pour les réactions de clivage 3' puis de transfert de brin lors de l'intégration (voir **chapitre III.3.3**). Les cations métalliques permettent de sélectionner et positionner les groupements réactifs (Yang *et al.*, 2006 ; Passos *et al.*, 2021) (**figure 35**). En effet, le CCD permet grâce à un cation métallique (cation A sur la **figure 35**) d'activer une molécule d'eau pour permettre une attaque nucléophile lors du clivage 3'. L'état de transition est stabilisé par le second cation métallique (cation B sur la **figure 35**) (Nowotny et Yang, 2006). Par la suite, les rôles des deux cations métalliques s'inversent lors de la réaction de transfert de brin avec l'activation de l'attaque nucléophile par le second cation métallique (Hare *et al.*, 2012). Le  $Mg^{2+}$  est considéré comme étant le cofacteur physiologique, au vu de son abondance dans la cellule, même s'il a été montré *in vitro*, que la réaction peut être catalysée en présence de  $Mn^{2+}$  (Bushman et Craigie, 1991).



**Figure 35. A.** Mécanismes impliquant la triade catalytique du CCD de l'IN du VIH-1 en présence d'ions  $Mn^{2+}$  lors de la réaction de clivage 3' (Hare *et al.*, 2012 ; code PDB : 4E7I). L'ADN viral est représenté en jaune. Les résidus d'acide aminé de l'IN sont représentés en bleu. **B.** Mécanismes impliquant la triade catalytique du CCD de l'IN du VIH-1 en présence d'ions  $Mn^{2+}$  lors des réactions de transfert de brin. La structure de la réaction de clivage 3' est superposée à celle du transfert de brin (Hare *et al.*, 2012 ; codes PDB : 4E7I et 4E7K). Les résidus d'acide aminé de l'IN et l'ADN hôte sont représentés en turquoise (clivage 3') et en rose (transfert de brin). Les ions métalliques sont représentés en violet, avec une inscription en jaune pour le clivage 3' et en blanc pour le transfert de brin. Les liaisons de coordination sont en pointillé. Les attaques nucléophiles sont représentées par des flèches noires. Source : Passos *et al.*, 2021.

La structure du CCD de l'IN du VIH-1 a été obtenue par cristallographie aux rayons X en mutant le résidu hydrophobe Phe185 en lysine afin d'augmenter la solubilité (Jenkins *et al.*, 1995 ; Dyda *et al.*, 1994 ; Maignan *et al.*, 1998). Elle présente une topologie similaire à celle des polynucléotidyl transférases comme la RNase H (Yang et Steitz, 1995 ; Dyda *et al.*, 1994) avec cinq brins  $\beta$  entourés par six hélices  $\alpha$  (**figure 36**).



**Figure 36.** Structure de l'interface dimérique du CCD de l'IN du VIH-1 obtenue par cristallographie aux rayons X (Maignan *et al.*, 1998 ; code PDB : 1BL3). La structure de chaque monomère est représentée en rubans (bleu et violet). Les chaînes latérales des résidus de la triade catalytique sont représentées en bâtonnets. Les ions  $Mg^{2+}$  sont représentés par des sphères rouges. Figure réalisée avec PyMOL.

Le CCD dimérise en solution avec une interface conservée dans toutes les espèces rétrovirales. La seule exception a été observée dans le laboratoire pour un fragment de CCD d'IN aviaire stabilisé par un ion  $Zn^{2+}$  (Ballandras *et al.*, 2011). Les auteurs ont proposé que cette nouvelle interface pouvait favoriser la liaison de l'ARN, en dehors de l'étape catalytique d'intégration avec l'ADN.

#### IV.1.4. Le domaine carboxy-terminal

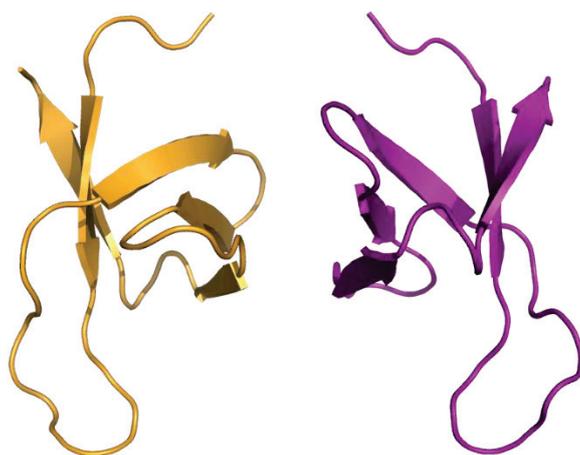
Le CTD contient environ 50 résidus d'acide aminé. Ce domaine est le moins conservé au niveau de la séquence en acides aminés entre les différents genres de rétrovirus. Cependant les structures tertiaires

des CTD sont similaires et forment un tonneau  $\beta$  replié en domaine SH3 (« *Src* homology 3 domain »), qui permet la liaison à l'ADN et la multimérisation (Lodi *et al.*, 1995 ; Eijkelenboom *et al.*, 1995). Le CTD est multifonctionnel, il permet de lier l'ADN viral comme l'ADN hôte cible, il participe également dans les interactions entre les molécules d'IN, avec l'ARN et d'autres partenaires cellulaires (Rocchi *et al.*, 2022).

Dans la littérature, le domaine SH3 est retrouvé chez les kinases ; il joue également un rôle dans la transcription et les nucléosomes ; il est impliqué dans des interactions médiées par des phosphorylations et des acétylations, dans le motif RRM (« *RNA recognition motif* ») (Kurochkina et Guha, 2013 ; Rocchi *et al.*, 2022 ; Kaneko *et al.*, 2008). Le domaine SH3 est souvent impliqué dans les interactions protéines:protéines en reconnaissant, par exemple, un motif riche en proline via une poche hydrophobe (Teyra *et al.*, 2017).

L'extrémité CT du CTD n'est pas critique pour la réplication virale, mais sa présence augmente l'efficacité de réplication et donc l'infectivité du virus. Une délétion de plus de neuf résidus d'acide aminé du CT impacte l'efficacité de réplication (Dar *et al.*, 2009 ; Mohammed *et al.*, 2011).

La structure du CTD de l'IN du VIH-1 a été obtenue en premier par RMN sous forme dimérique. Elle ne comprend pas l'extrémité CT qui affecte la solubilité de la protéine (Lodi *et al.*, 1995 ; Eijkelenboom *et al.*, 1995) (**figure 37**).

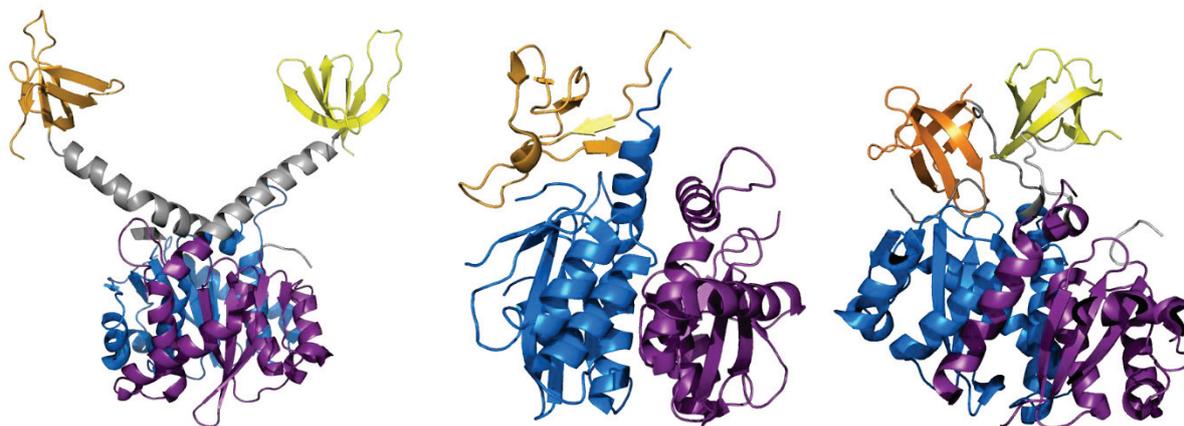


**Figure 37.** Structure de l'interface dimérique du CTD de l'IN du VIH-1 obtenue par RMN (Lodi *et al.*, 1995 ; code PDB : 1IHV). La structure de chaque monomère est représentée en rubans (orange et violet). Figure réalisée avec PyMOL.

#### IV.1.5. Structures « deux domaines » de l'IN

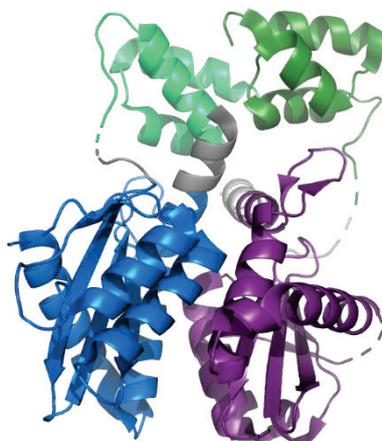
La structure du fragment CCD-CTD (aa 52-288) du VIH-1 a été obtenue en 2000 par cristallographie aux rayons X en réalisant cinq mutations (C56S, W131D, F139D, F185K et C280S) pour solubiliser la construction protéique (Chen *et al.*, 2000). Les structures CCD-CTD du VIS et du RSV ont été obtenues la même année (Chen *et al.*, 2000 ; Yang *et al.*, 2000).

La protéine dimérise dans les trois structures via le CCD (**figure 38**) mais la boucle inter-domaine CCD-CTD du VIH-1 forme une hélice rigide qui n'est pas observée dans le VIS et RSV. Ainsi, les CTD sont éloignés dans le VIH-1 alors qu'ils interagissent dans RSV. Un seul CTD est observé dans le VIS pour cause de flexibilité.



**Figure 38.** Structures CCD-CTD du VIH-1 (Chen *et al.*, 2000 ; code PDB : 1EX4), du VIS (Chen *et al.*, 2000 ; code PDB : 1C6V) et de RSV (Yang *et al.*, 2000 ; code PDB : 1C0M). Le CCD d'un monomère est représenté en bleu, celui de l'autre monomère en violet. Le CTD d'un monomère est représenté en orange, et celui de l'autre en jaune. Les boucles inter-domaines sont en nuance de gris (gris foncé pour un monomère et gris clair pour l'autre monomère). La structure de chaque monomère est représentée en rubans. Figure réalisée avec PyMOL.

Par la suite, la structure du fragment NTD-CCD (aa 1-212) du VIH-1 a été obtenue sous forme dimérique par cristallographie aux rayons X en réalisant trois mutations (W131D, F139D et F185K) (Wang *et al.*, 2001) (**figure 39**). L'interface CCD:CCD est de nouveau conservée alors que l'interface NTD:NTD est différente de celle précédemment observée pour le NTD seul.



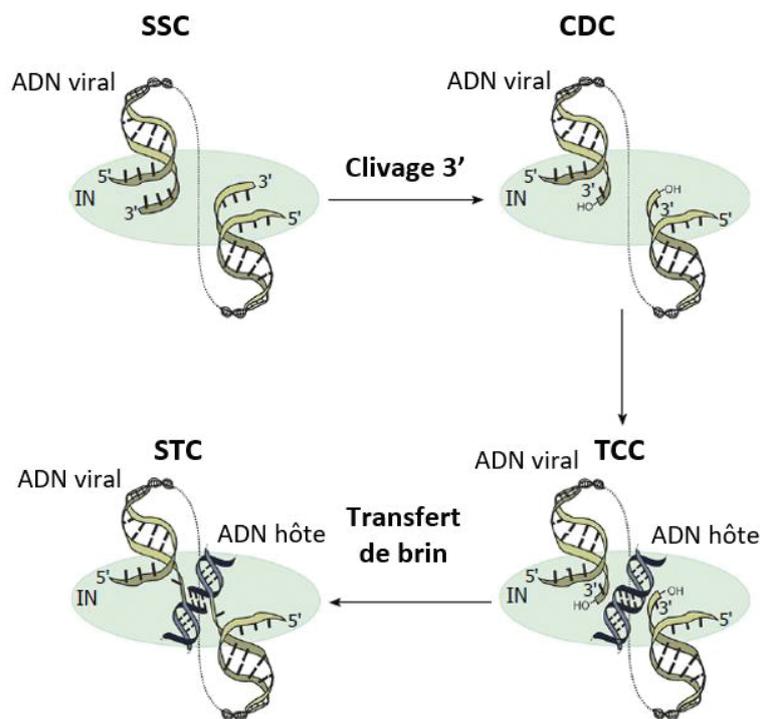
**Figure 39.** Structure NTD-CCD du VIH-1 (Wang *et al.*, 2001 ; code PDB : 1K6Y). Le NTD d'un monomère est représenté en vert clair, et celui de l'autre en vert foncé. Le CCD d'un monomère est représenté en bleu, celui de l'autre monomère en violet. Les boucles inter-domaines sont en nuance de gris (gris foncé pour un monomère et gris clair pour l'autre monomère). La structure de chaque monomère est représentée en rubans. Figure réalisée avec PyMOL.

Ces différentes études par fragments montrent donc une variabilité des interfaces dimériques CTD:CTD et NTD:NTD alors que l'interface CCD:CCD semble bien conservée. L'étude structurale de l'IN entière a ensuite été réalisée en complexe avec de l'ADN viral, pour stabiliser l'ensemble sous la forme d'un complexe nommé intasome. A ce jour, il n'existe pas de structure de l'IN entière sans acide nucléique.

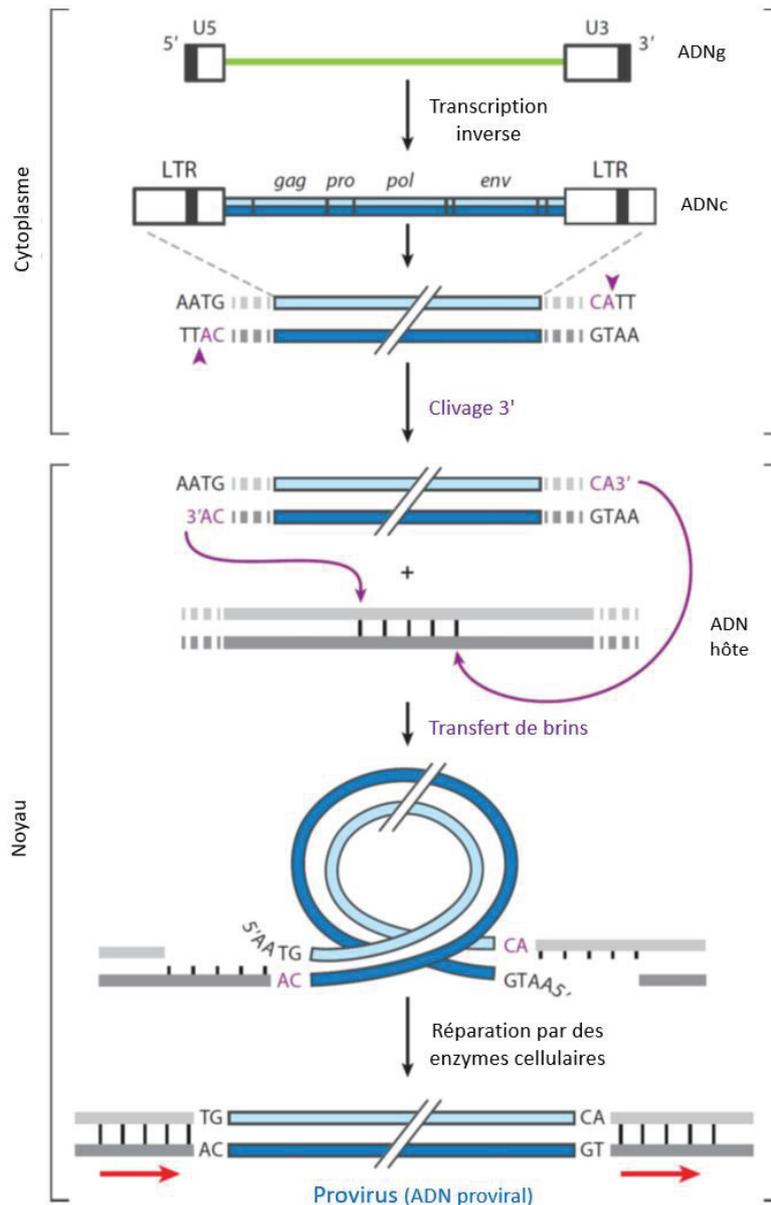
## IV.1.6.1. La réaction d'intégration

Comme expliqué précédemment, la réaction catalytique d'intégration se déroule en deux étapes et débute dans le cytoplasme après la transcription inverse : l'IN qui lie les deux extrémités de l'ADN viral sous forme d'un intasome SSC, « *stable synaptic complex* ») clive en 3' des nucléotides conservées CA pour former un complexe CDC (« *cleaved donor complex* ») (Engelman et Cherepanov, 2012 ; Andrade et Skalka, 2015 ; Grawenhoff et Engelman, 2017) (**figures 40 et 41**).

Puis vient la seconde étape de transfert de brin, qui se déroule dans le noyau et permet l'intégration de l'ADN viral activé dans l'ADN cellulaire sous forme de provirus. L'intasome CDC se lie à l'ADN cellulaire cible pour former le complexe de capture TCC (« *target capture complex* ») (Hare *et al.*, 2012 ; Li et Craigie, 2009 ; Li *et al.*, 2006 ; Grawenhoff et Engelman, 2017). Puis le groupement 3'OH des A aux extrémités 3' des brins d'ADN viral permet une attaque nucléophile sur un phosphate d'une liaison phosphodiester au sein de l'ADN cellulaire (**figure 41**) formant le complexe de transfert de brin (STC, « *strand transfer complex* ») (**figure 40**). L'espace entre les deux sites d'insertion varie selon les virus de quatre à six nucléotides, il est de cinq nucléotides pour le VH-1 (Strauss et Strauss, 2008 ; Andrade et Skalka, 2015 ; Van Cor-Hosmer *et al.*, 2013). Après ce processus de transfert de brin, les enzymes cellulaires permettent de compléter les espaces et d'éliminer les nucléotides non-appariés, notamment les deux nucléotides terminaux de l'ADN viral.



**Figure 40.** Les différents états de l'intasome au cours de la réaction d'intégration catalysée par l'IN multimérique. Source : Grawenhoff et Engelman, 2017.



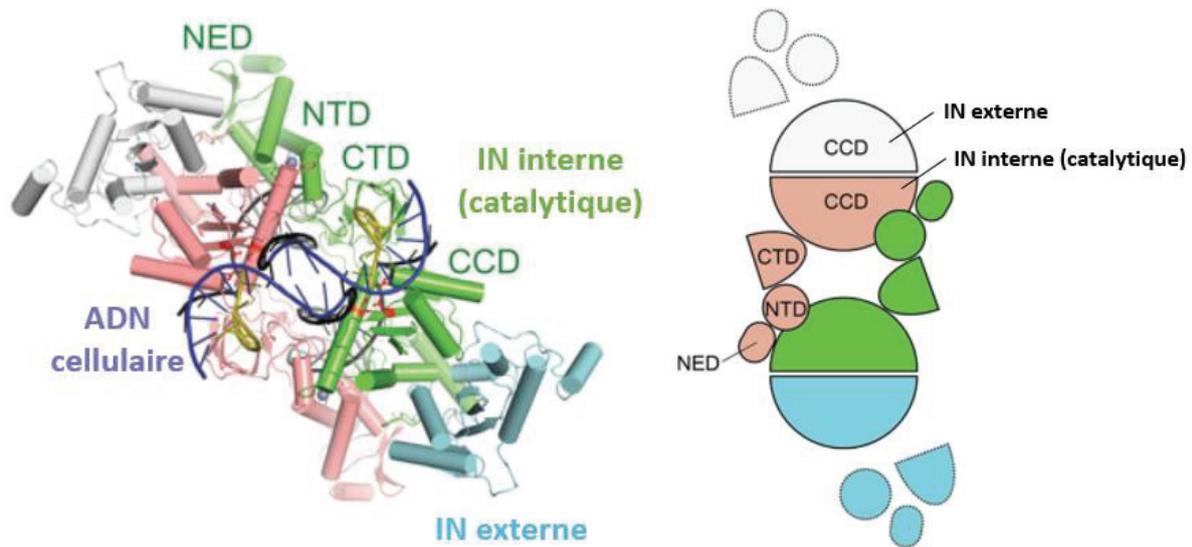
**Figure 41.** Schéma du mécanisme d'intégration. Source : Andrade et Skalka, 2015.

#### IV.1.6.2. Les structures des intasomes

A partir des années 2010, différentes structures d'intasomes rétroviraux ont été résolues par cristallographie aux rayons X et microscopie électronique.

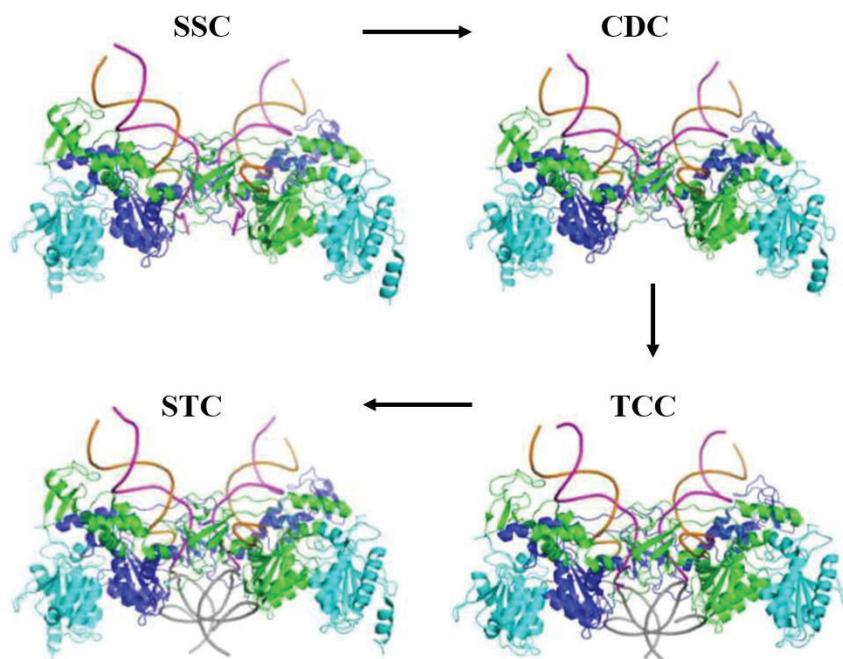
La première structure obtenue est celle de PFV (Hare *et al.*, 2010) qui est une IN soluble caractérisée par une intégration concertée efficace (Valkov *et al.*, 2008 ; Grawenhoff et Engelman, 2017). Cette structure s'assemble en dimère de dimère et correspond au complexe CDC (Hare *et al.*, 2010) (**figure 42**). Elle comporte un dimère interne qui interagit avec l'ADN viral et réalise la catalyse enzymatique. Le CTD de ce monomère se positionne entre les domaines CCD et NTD internes et participe à la liaison de l'ADN viral. Cet agencement n'était pas prévisible d'après les structures de fragments d'IN seule publiées précédemment.

Seuls les domaines CCD des deux monomères externes sont visibles dans les cartes de densité électronique. Ils forment chacun une interface dimérique canonique avec les CCD des monomères internes (**figure 42**).



**Figure 42.** *A gauche.* Structure de l'intasome tétramérique de PFV (Hare *et al.*, 2010 ; code PDB : 3OYM). *A droite.* Représentation schématique de l'organisation des domaines de l'IN au sein de l'intasome de PFV. Source : Grandgenett et Aihara, 2018.

Cet agencement de l'intasome de PFV en dimère de dimère est rigide et est conservé dans toutes les structures d'intermédiaires de PFV (SSC, TTC, STC) publiées par la suite (Hare *et al.*, 2012 ; Maertens *et al.*, 2010). La structure de l'intasome STC de PFV montre que l'ADN cellulaire est « simplement » lié le long d'une crevasse parcourant l'interface synaptique CCD-CTD des dimères internes, sans changement significatif de l'assemblage de l'intasome (**figure 43**) (Maertens *et al.*, 2010).

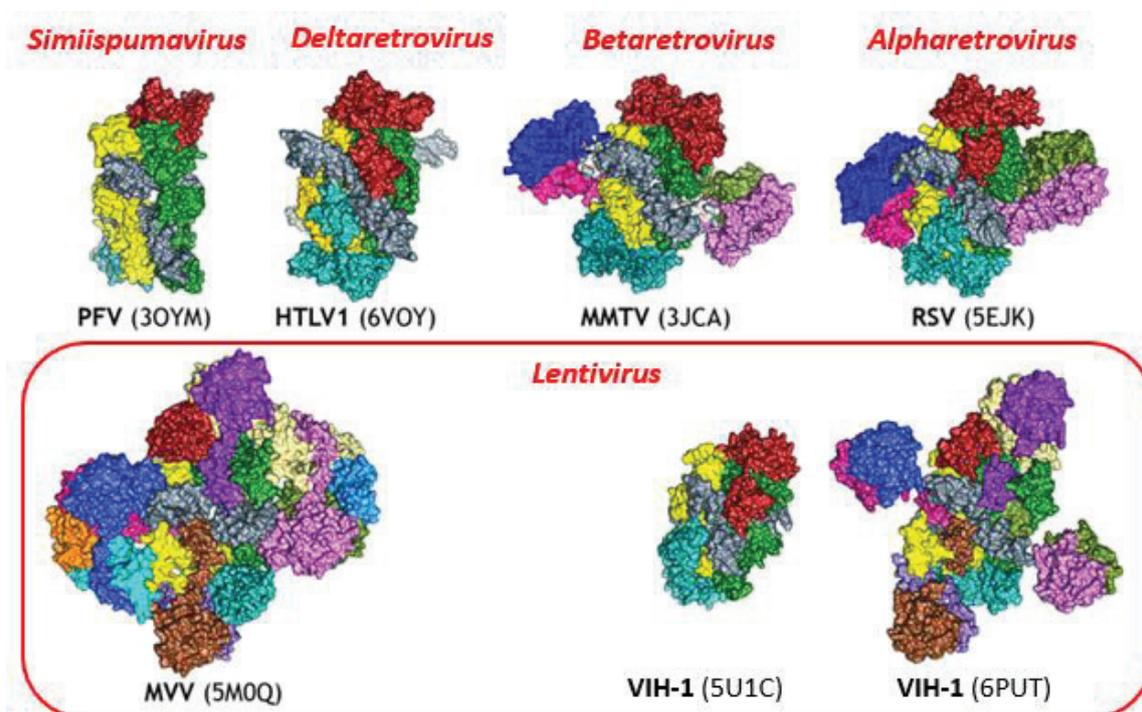


**Figure 43.** Structures des différents intasomes de PFV (SSC, Hare *et al.*, 2012 – code PDB : 4E7H ; CDC, Hare *et al.*, 2010 – code PDB : 3OY9 ; TCC, Maertens *et al.*, 2010 – code PDB : 3OS2 ; STC, Maertens *et al.*, 2010 – code PDB : 3OS0). Source : Engelman et Cherepanov, 2017.

L'arrangement synaptique de l'intasome de PFV est observé dans toutes les structures d'intasomes rétroviraux résolues à ce jour et s'appelle CIC pour « *conserved intasome core* ».

Toutefois, différents degrés d'oligomérisation ont été observés qui semblent reliés à la longueur des boucles inter-domaines. Si la longueur de la boucle CCD-CTD n'est pas suffisante pour permettre au CTD du monomère interne de se positionner entre le CCD et le NTD, le CTD devra être fourni par un monomère extérieur ce qui conduit à un plus haut degré d'oligomérisation (Engelman et Cherepanov, 2017 ; Grawenhoff et Engelman, 2017).

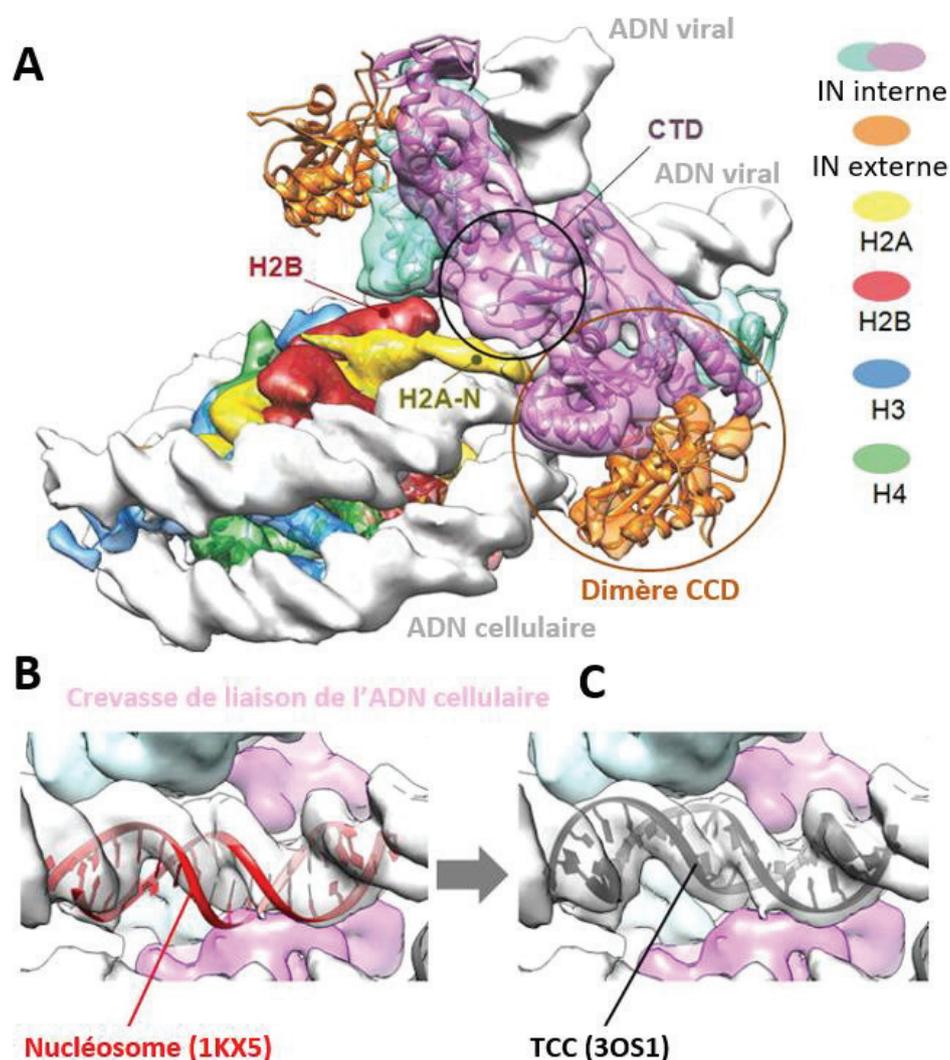
Ainsi l'intasome est sous forme tétramérique pour les simiispumavirus (PFV) et les deltarétrovirus (HTLV) (Bhatt *et al.*, 2020) ; octamérique pour les alpharétrovirus (RSV) (Yin *et al.*, 2016) et les betarétrovirus (MMTV) (Ballandras-Colas *et al.*, 2016) ; et d'un degré d'oligomérisation supérieur pour les lentivirus avec un assemblage hexadécamérique pour le MVV (Ballandras-Colas *et al.*, 2017) et dodécamérique pour le VIH-1 (Passos *et al.*, 2020) (**figure 44**). L'oligomérisation de l'IN *in vivo* permet la protection des extrémités de l'ADN viral (200 – 250 pb) contre les attaques des nucléases cellulaires (Chen *et al.*, 1999).



**Figure 44.** Structures des intasomes de différents genres de rétrovirus. Chaque IN est colorée d'une couleur différente. L'intasome est sous forme tétramérique pour PFV (Hare *et al.*, 2010) et HTLV-1 (Bhatt *et al.*, 2020) ; octamérique pour RSV (Yin *et al.*, 2016) et MMTV (Ballandras-Colas *et al.*, 2016) ; hexadécamérique pour MVV (Ballandras-Colas *et al.*, 2017) ; tétramérique (Passos *et al.*, 2017) et dodécamérique (Passos *et al.*, 2020) pour le VIH-1. Source : Passos *et al.*, 2021.

De plus, la structure de l'intasome de PFV en complexe avec un nucléosome avant l'intégration a été résolue (Maskell *et al.*, 2015) (**figure 45.A**). Le nucléosome contient un octamère de protéines décomposé en un hétérotétramère d'histone H3-H4, deux hétérodimères d'histone H2A-H2B et 147 pb d'ADN cellulaire (Davey *et al.*, 2002). La structure montre que l'intasome interagit avec l'ADN cellulaire et un hétérodimère H2A-H2B. L'ADN cellulaire interagit au niveau d'une surface « *saddle-shaped* » à l'interface dimérique du CCD (Maskell *et al.*, 2015). Le domaine N-terminal de H2A interagit avec un CTD d'une IN interne. Les CCD des IN externes n'ont pas de rôle catalytique et servent de plateforme pour la liaison de l'ADN cellulaire. L'intasome interagit avec le nucléosome et des partenaires cellulaires pour permettre une intégration à des sites préférentiels (Passos *et al.*, 2021),

comme détaillé précédemment (voir **chapitre III.3.3**). Le nucléosome subit un changement conformationnel en se liant à l'intasome (**figure 45.B et C**).



**Figure 45.** *A. Modèle du complexe nucléosome : intasome à partir de la structure du nucléosome (Davey et al., 2002 ; code PDB : 1KX5). Le dimère d'IN interne est représenté en turquoise et violet, et les CCD du dimère d'IN externe sont représentés en orange. L'ADN cellulaire et l'ADN viral sont représentés en gris. Les histones sont représentées en jaune, rouge, bleu et vert. B. Modélisation de la structure du nucléosome (Davey et al., 2002 ; code PDB : 1KX5) au sein de la cavité où est située l'ADN cellulaire dans la structure de l'intasome (Maertens et al., 2010 ; code PDB : 3OS1). C. Structure de l'intasome TCC (Davey et al., 2002 ; code PDB : 1KX5). Source : Maskell et al., 2015.*

#### IV.2. Modifications post-traductionnelles et rôle multifonctionnel de l'IN du VIH-1

L'IN est une protéine pléiotropique qui impacte plusieurs étapes du cycle rétroviral (transcription inverse, transcription virale, morphogénèse) en plus de celle principale d'intégration décrite ci-dessus. Pour réaliser ces différentes fonctions, l'IN du VIH-1 interagit avec différents partenaires grâce à la présence de certaines modifications post-traductionnelles (Zheng et Yao, 2013) (**figure 46**). La majorité de ces modifications post-traductionnelles se situent dans le CTD.

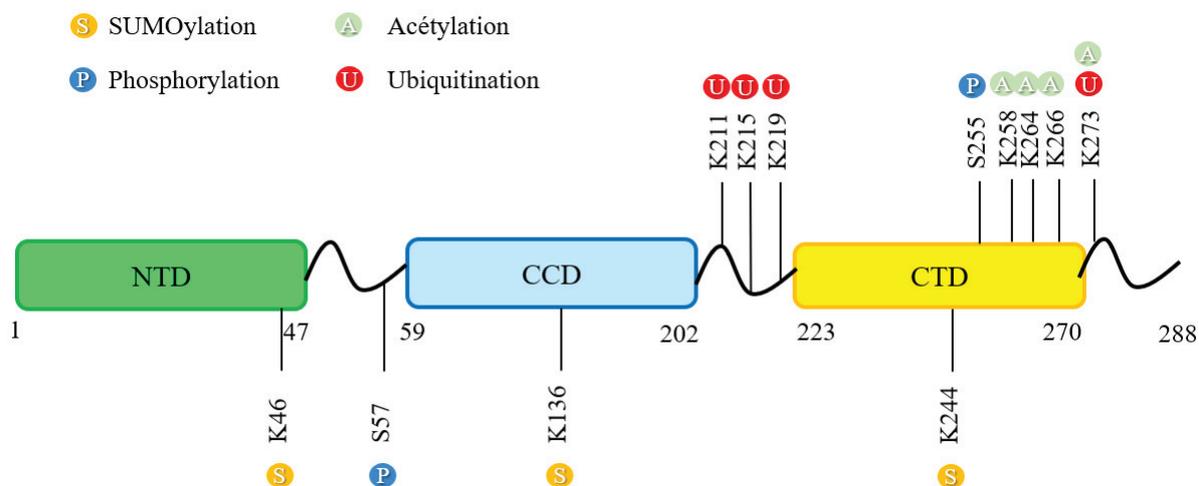
Les acétylations du CTD sont réalisées par des histones acétyltransférases (HAT) cellulaires. Les acétylations induites par la HAT p300 sur les résidus K264, K266 et K273 du CTD augmentent l'affinité

de l'IN pour l'ADN et améliorent l'activité de transfert de brin de l'IN *in vitro* (Cereseto *et al.*, 2005). De même, l'acétylation sur ces résidus peut être réalisées par une autre HAT, la GCN5 (« *general control non-depressible 5* »), qui acétyle aussi le résidu K258 du CTD améliorant l'efficacité de l'intégration provirale (Terreni *et al.*, 2010). La mutation de l'acétylation sur ces résidus abolie la réplication virale en affectant l'intégration.

En plus de ces acétylations, l'IN peut être modifiée par des SUMOylations sur les résidus K46, K136 et K244 (Zamborlini *et al.*, 2011). La fonction de ces modifications n'est pas encore totalement comprise, mais leur absence diminue l'efficacité de l'intégration provirale *in vivo* (Zamborlini *et al.*, 2011).

Les phosphorylations par des kinases cellulaires, JNK (« *c-Jun N-terminal kinase* ») et GCN2 (« *general control non-derepressible 2* ») respectivement, sur les résidus S57 et S255 de l'IN ont des rôles antagonistes (Manganaro *et al.*, 2010 ; Jaspert *et al.*, 2017). En effet, la phosphorylation sur la S57 permet l'interaction de l'IN avec l'enzyme cellulaire Pin1 (« *peptidyl prolyl-isomerase* ») (Manganaro *et al.*, 2010). Cette enzyme catalyse l'isomérisation cis-trans des liaisons peptidyl-prolyl, ce qui modifie la conformation de l'IN et améliore la stabilité de l'IN en la protégeant de la digestion médiée par certaines protéases *in vitro* (Manganaro *et al.*, 2010). Tandis que la phosphorylation sur la S255 impacte négativement l'intégration et diminue la réplication du VIH-1 (Jaspert *et al.*, 2017).

Dernièrement, l'IN présente plusieurs ubiquitinations. L'ubiquitination des résidus K211, K215, K219 et K273 est médiée par l'ubiquitine ligase Cullin2/VHL (« *cullin2-based von Hippel-Lindau* ») et permet la dégradation de l'IN par le protéasome (Mousnier *et al.*, 2007). La mutation de ces lysines en arginines réduit de 60% l'infectivité du virus, en effet la dégradation de l'IN par le protéasome joue un rôle dans la transition entre l'intégration et la transcription du provirus dans le cycle réplcatif du VIH-1.



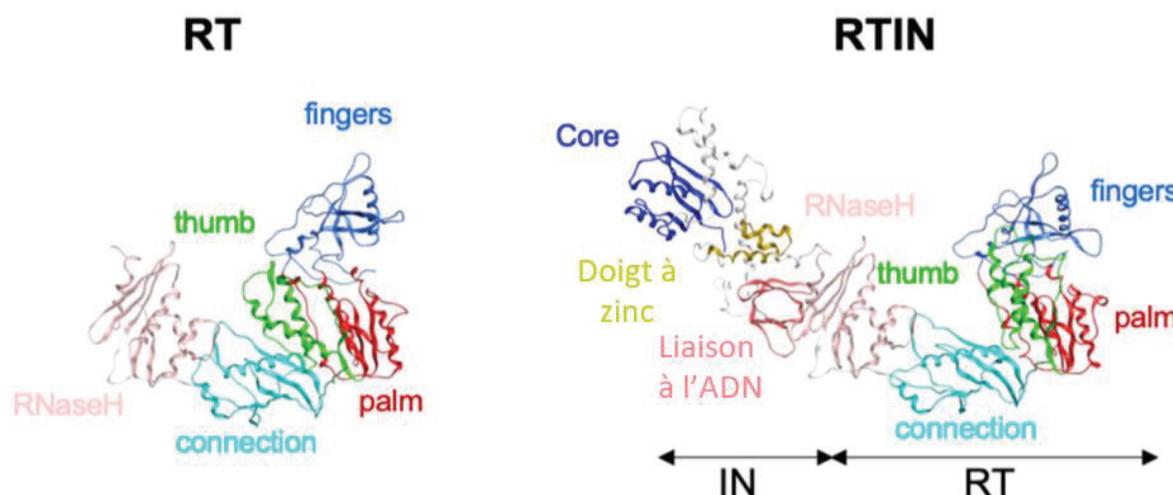
**Figure 46.** Représentations des sites de modifications post-traductionnelles sur l'IN du VIH-1. Source : Zheng *et Yao*, 2013.

#### IV.2.1. Transcription inverse

La RT entière de même que ses sous-unités individuelles, p51 et p66, sont capables d'interagir directement avec l'IN (Wu *et al.*, 1999 ; Hehl *et al.*, 2004). Cette interaction est médiée par le CTD de l'IN et les résidus W243, V250 et K258 sont critiques pour l'interaction avec la RT (Wilkinson *et al.*, 2009). L'interaction entre l'IN et la RT n'affecte pas l'étape de transcription inverse, mais permet d'améliorer la réaction de transfert de brin catalysée par l'IN *in vitro*. Cependant d'autres études montrent que l'IN influence également l'étape de transcription inverse en améliorant l'initiation et l'élongation de la transcription inverse (Dobard *et al.*, 2007). Des expériences *ex vivo* avec des mutants de l'IN pour la liaison à la RT montrent que l'interaction de l'IN avec la RT est critique pour la transcription inverse (Wu *et al.*, 1999 ; Wilkinson *et al.*, 2009 ; Tekeste *et al.*, 2015). De plus, la

résolution récente d'un modèle structural de la polyprotéine RT-IN a permis de mettre en évidence les changements conformationnels de la RT suite à la liaison de l'IN (Masuda *et al.*, 2023) (**figure 47**). Dans ce modèle, la liaison de l'IN à la RT modifie la configuration de la région « thumb » de la RT, permettant l'établissement de contacts entre les régions « thumb » et « fingers ».

Les régions « thumb », « fingers » et « palm » constituent le centre actif de la polymérase et participent à la liaison du substrat ARN et des dNTP. Ainsi, la liaison de l'IN à la RT pourrait induire des changements de conformation de la RT, qui amélioreraient l'activité polymérase de l'enzyme.



**Figure 47.** Structure de la RT seule (à gauche) et modèle structural de la RT liée à l'IN (à droite). Sources : Masuda *et al.*, 2023 ; code PDB pour la RT : 3HVT (Smerdon *et al.*, 1994), et codes PDB pour l'IN : 1K6Y (Wang *et al.*, 2001) et 1EX4 (Chen *et al.*, 2000).

#### IV.2.2. Import nucléaire

Pendant le cycle rétroviral, l'IN qui est localisée au sein de la capsid virale est transportée du cytoplasme au noyau où se déroule la réaction de transfert de brin (Zila *et al.*, 2021). Lors de l'étape d'import nucléaire, l'IN interagit via son CTD avec différents facteurs cellulaires impliqués dans l'import nucléaire. Tout d'abord, il a été mis en évidence que la séquence NLS du CTD de l'IN est nécessaire pour l'import nucléaire (Levin *et al.*, 2009). De plus, la transportine-SR2 (TRN-SR2) interagit directement avec l'IN du VIH-1 au niveau des résidus R262, R263 et K264 du CTD (Tabasi *et al.*, 2021). TRN-SR2 est un membre de la famille des karyophérines, qui sont des protéines permettant l'import ou l'export en liant le signal NLS ou NES (« nuclear export signal ») d'une cargaison. En plus, il a été montré que l'IN interagit également avec les importines  $\alpha$ , notamment les importines  $\alpha 3$  et  $\alpha 7$ , qui sont des membres de la famille des karyophérines (Gallay *et al.*, 1997 ; Ao *et al.*, 2010 ; Zaitseva *et al.*, 2009).

#### IV.2.3. Transcription provirale

Des expériences de séquençage après « crosslinking » et immuno-précipitation (« CLIP-Seq » pour « Cross-Linking ImmunoPrecipitation Sequencing ») et des expériences d'interactions biomoléculaires par « AlphaScreen » (« Amplified Luminescent Proximity Homogeneous Assay ») ont mis en évidence une liaison préférentielle de l'IN du VIH-1 à l'ARN TAR situé dans le 5'UTR de l'ARNg (Kessl *et al.*, 2016). L'ARN TAR joue un rôle majeur dans la transcription provirale avec la protéine Tat, comme décrit précédemment. L'IN lie l'ARN TAR lorsqu'elle est sous forme monomérique, dimérique et tétramérique avec une meilleure affinité (Elliott *et al.*, 2020). L'interaction entre l'ARN TAR et l'IN

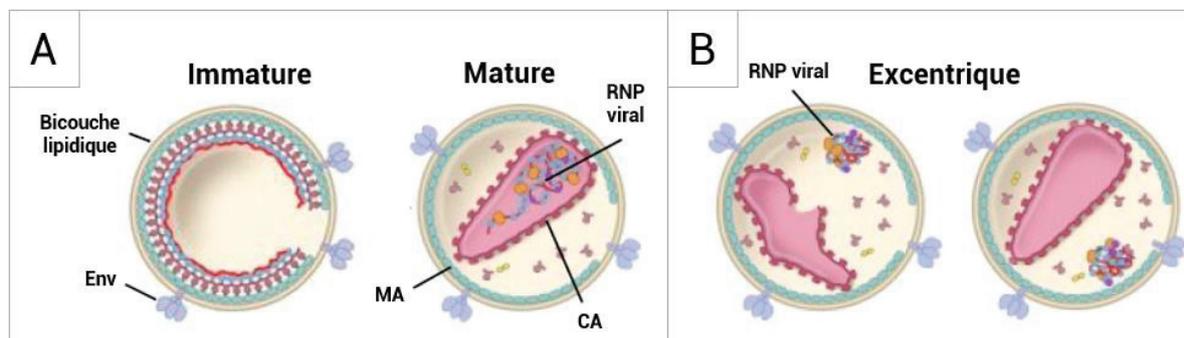
met en jeu des résidus présents dans le CTD de l'IN : K264, K266, R269 et K273. De plus, une étude récente a montré l'importance de l'acétylation de quatre résidus de lysine situés dans le CTD de l'IN du VIH-1 – K258, K264, K266 et K273 – pour la transcription provirale (Winans et Goff, 2020).

Il est donc intéressant de noter que le CTD est critique pour l'interaction avec l'ARN. De plus, dans le cas de l'interaction à l'ARN TAR, les résidus du CTD impliqués présentent pour la plupart des modifications post-traductionnelles qui sont critiques pour la transcription provirale. Cette observation suggère un rôle de l'IN lié à l'ARN lors de la transcription provirale.

#### IV.2.4. Morphogenèse

L'implication de l'IN du VIH-1 dans la morphogenèse de la particule virale a été récemment mise en évidence. L'IN lie l'ARNg sur des régions structurées préférentielles (Kessl *et al.*, 2016 ; Elliott *et al.*, 2020). Plusieurs résidus chargés positivement situés dans le CTD de l'IN participent à cette interaction : R262, R263, K264, K266, R269 et K273 ; ainsi qu'un résidu lysine dans le NTD : K34 (Elliott *et al.*, 2020 ; Mugisha *et al.*, 2022).

Une inhibition de la liaison de l'IN à l'ARNg a également été observée et peut être due à trois mécanismes (Elliott *et al.*, 2020) : (1) une diminution du nombre d'IN dans le virus, conduisant à une limitation de la quantité d'IN disponible pour lier l'ARNg ; (2) une altération de la multimérisation de l'IN cause une perturbation de la liaison de l'IN à l'ARNg ; (3) une inhibition directe de l'interaction entre l'IN et l'ARNg. Cette inhibition entraîne la formation de particules virales excentriques avec une mauvaise localisation de l'ARNg en-dehors de la capsid virale, ce qui conduit à une dégradation précoce de l'ARNg au prochain cycle viral, donc les particules virales sont non-infectieuses (Balakrishnan *et al.*, 2013 ; Fontana *et al.*, 2015 ; Jurado *et al.*, 2013 ; Madison *et al.*, 2017) (**figure 48**). Ce phénotype correspond aux mutants de l'IN de classe II, qui se distinguent des mutants de l'IN de classe I où l'étape d'intégration est inhibée (Sharma *et al.*, 2014).



**Figure 48.** Schéma de particules du VIH-1 immature, mature (A) et excentrique (B). Sources : Fontana *et al.*, 2015 ; Engelman et Kvaratskhelia, 2022.

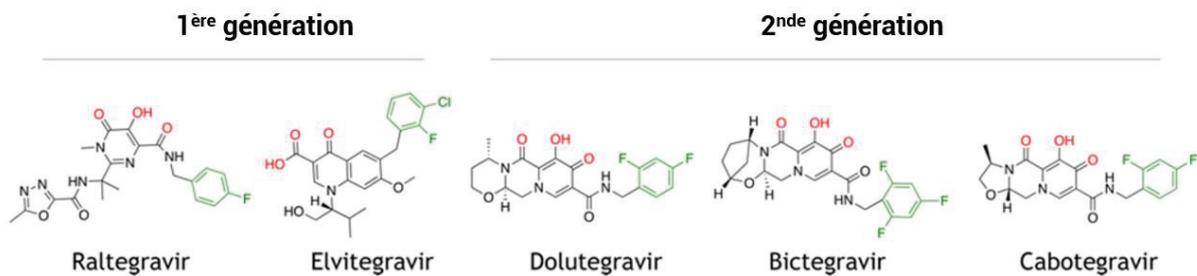
L'IN du VIH-1 contribue donc à la morphogenèse en assurant la bonne localisation des complexes ribonucléoprotéiques ARNg:NC au sein de la capsid virale. Pour mieux comprendre ce rôle, il faudrait déterminer la structure de l'IN en complexe avec l'ARNg et une partie de mes travaux de thèse a porté sur cette étude.

### IV.3. Les inhibiteurs de l'intégrase

L'IN est une enzyme clé du cycle rétroviral et il existe actuellement deux classes d'inhibiteurs de l'IN : les inhibiteurs de classe I qui sont les INSTI et ciblent l'étape d'intégration ; les inhibiteurs de classe II qui sont les ALLINI et ciblent les fonctions non-catalytiques de l'IN comme la liaison à l'ARNg.

#### IV.3.1. INSTI – Inhibiteurs du transfert de brin

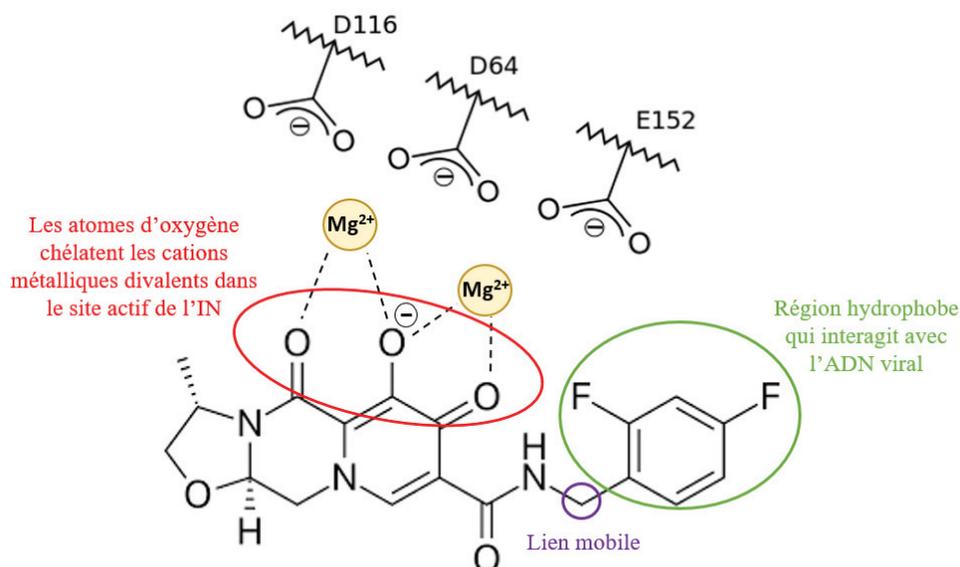
Cinq INSTI sont approuvés par la FDA (« *Food and Drug Administration* ») pour le traitement des patients séropositifs. Il s'agit d'INSTI de première génération (Raltegravir, Elvitegravir) et de seconde génération (Dolutegravir, Bictégravir et Cabotegravir) présentant une meilleure barrière génétique (Delaugerre, 2010 ; Passos *et al.*, 2021 ; Jozwik *et al.*, 2020) (**figure 49**).



**Figure 49.** Formules développées des cinq INSTI cliniquement approuvés pour traiter les patients infectés par le VIH-1. Source : Passos *et al.*, 2021.

Ces inhibiteurs contiennent un pharmacophore avec trois atomes électronégatifs qui chélatent les ions  $Mg^{2+}$  du site actif de l'IN (Jozwik *et al.*, 2020 ; Passos *et al.*, 2021 ; Engelman, 2019) (**figure 50**). De plus, les INSTI présentent à leur extrémité un cycle benzène halogéné qui interagit avec l'ADN viral au sein de l'intasome. Ce cycle déplace l'extrémité 3' de l'ADN pour empêcher la réaction de transfert de brin ; la présence d'un lien mobile entre le cycle benzène et le cœur de la molécule permet d'augmenter la surface de contact avec le site actif de l'IN (Passos *et al.*, 2021).

L'interaction des INSTI avec l'ADN permet ainsi un ciblage spécifique de l'IN au sein des intasomes (Jozwik *et al.*, 2020).



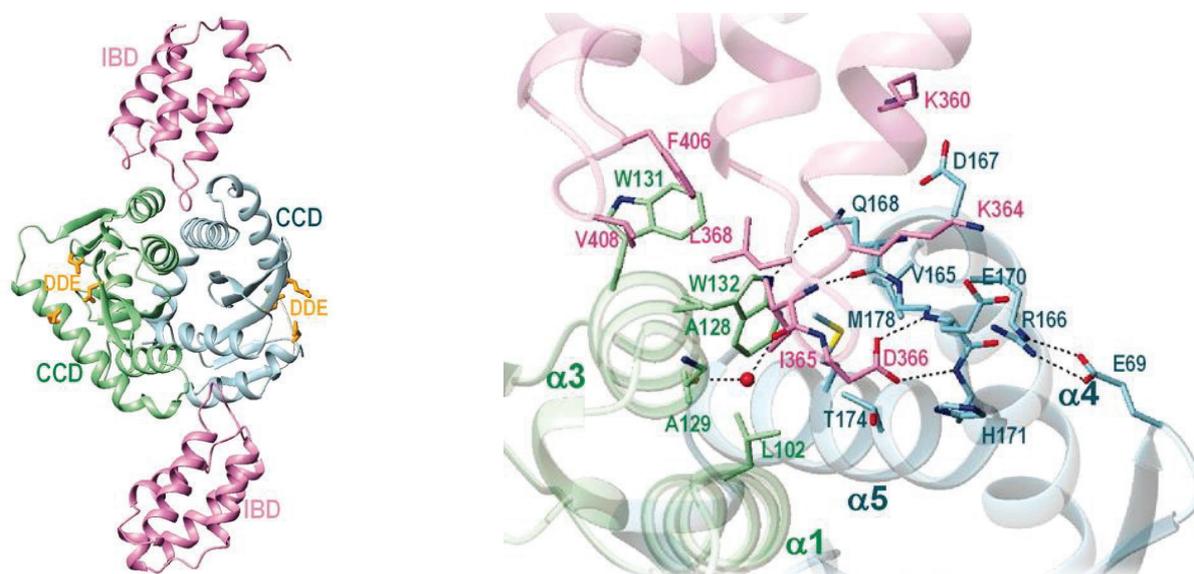
**Figure 50.** Groupements importants et mode d'action des INSTI. La structure développée utilisée est celle du Cabotegravir. Sources : Jozwik *et al.*, 2020 ; Passos *et al.*, 2021.

#### IV.3.2. ALLINI – Inhibiteurs allostériques

La mise sur le marché des ALLINI n'a pas encore été approuvée par la FDA. Ces inhibiteurs avaient été initialement conçus pour empêcher la liaison du partenaire cellulaire LEDGF/p75 à l'IN du VIH-1 (Cherepanov *et al.*, 2005 ; Engelman *et al.*, 2013 ; Grawenhoff et Engelman, 2017) (voir **chapitre III.3.3**).

Cette liaison intermoléculaire se fait entre le domaine IBD de LEDGF/p75 et le CCD de l'IN. La structure de ce complexe a été résolue par cristallographie aux rayons X avec la mutation F185K (Cherepanov *et al.*, 2005) (**figure 51**) et a permis de guider la synthèse des ALLINI.

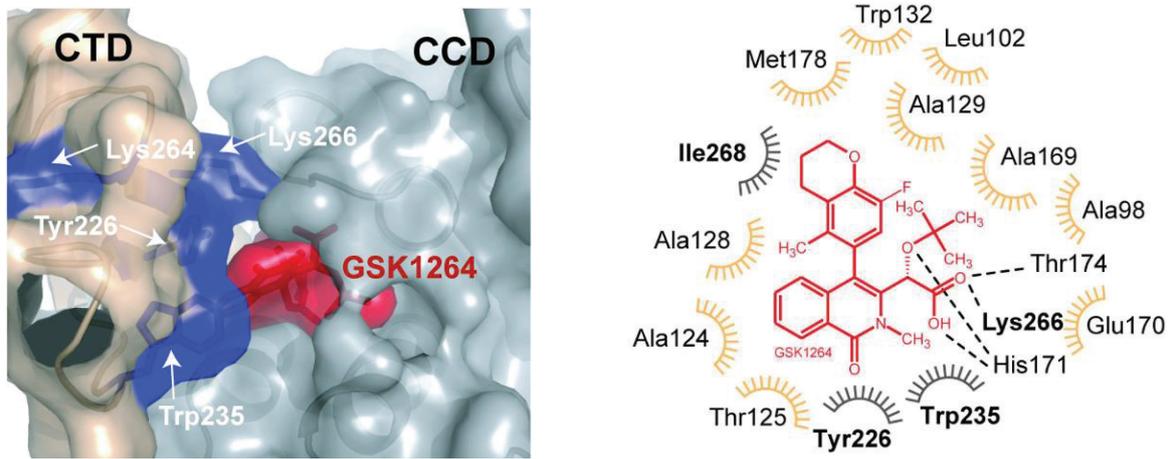
Ces molécules sont en effet des analogues structuraux des résidus Asp366 et Ile365 de l'IBD, qui interagissent respectivement avec les résidus Glu170 et His171, et une poche hydrophobe (Leu102, Ala128 et Trp132) du CCD (Engelman *et al.*, 2013 ; Grawenhoff et Engelman, 2017).



**Figure 51.** Mécanisme moléculaire de l'interaction de l'IN avec LEDGF/p75. **A gauche.** Représentation de la structure IBD : CCD de l'IN du VIH-1 en rubans (Cherepanov *et al.*, 2005 ; code PDB : 2B4J). L'IBD est représenté en rose, les CCD en vert et bleu, et les chaînes latérales de la triade catalytique sont représentées en bâtonnets orange. **A droite.** Représentation en bâtonnets des chaînes latérales des résidus importants pour l'interaction entre l'IBD de LEDGF/p75 et le CCD de l'IN du VIH-1 (Cherepanov *et al.*, 2005 ; code PDB : 2B4J). Source : Cherepanov *et al.*, 2005.

Curieusement, il a ensuite été observé que les ALLINI n'affectaient pas l'intégration sélective médiée par LEDGF/p75 à des concentrations thérapeutiques (Gupta *et al.*, 2014 ; Feng *et al.*, 2015), mais qu'ils inhibaient la morphogénèse du VIH-1. Ils entraînent en effet une hyper-multimérisation de l'IN via des interactions CCD:CTD (**figures 52 et 53**) et une perturbation de la liaison de l'IN à l'ARN, ce qui conduit à des particules virales excentriques non-infectieuses (Jurado *et al.*, 2013 ; Kessler *et al.*, 2012 ; Feng *et al.*, 2015).

Ainsi, les ALLINI agissent comme des stabilisateurs d'interaction IN:IN, notamment via les résidus Lys264 et Lys266 du CTD (**figure 52**). Le composé STP0404 est actuellement en phase d'essai chez l'homme (Maehigashi *et al.*, 2021).



**Figure 52.** *A gauche.* Représentation des surfaces de la structure GSK 1264 : IN du VIH-1 (Gupta et al., 2016 ; code PDB : 5HOT). L'ALLINI GSK 1264 est représenté en rouge. Le CCD est représenté en gris, le CTD en beige et les résidus du CTD à l'interface avec GSK 1264 en bleu foncé. *A droite.* Représentation schématique des contacts entre le CCD (en jaune) et le CTD (en noir et en gras) de l'IN du VIH-1 et l'ALLINI GSK 1264 (en rouge). Source : Gupta et al., 2016.



**Figure 53.** La liaison de l'ALLINI GSK 1264 entraîne une hyper-multimérisation de l'IN du VIH-1. Structure du complexe GSK 1264:IN du VIH-1 (Gupta et al., 2016 ; code PDB : 5HOT). L'ALLINI GSK 1264 est représenté par des sphères rouges. Un dimère CCD-CTD est représenté en gris et l'autre en beige sous forme de rubans. Source : Gupta et al., 2016.

## V. ISG20

Pour ce qui concerne l'état des connaissances sur mon second sujet, le facteur de restriction ISG20, j'ai participé à la rédaction d'une revue en collaboration avec l'équipe du Dr. Andrea Cimarelli (CIRI, Lyon), qui a été acceptée dans le journal FEBS Open Bio en février 2022. Cette revue est insérée ci-dessous et fait un état des lieux des connaissances actuelles sur les mécanismes moléculaires de l'action antivirale d'ISG20, en détaillant les preuves à l'appui et les questions qui restent à résoudre pour chaque mécanisme proposé.

# ISG20: an enigmatic antiviral RNase targeting multiple viruses

S everine Deymier<sup>1</sup> , Camille Louvat<sup>2</sup> , Francesca Fiorini<sup>2</sup>  and Andrea Cimarelli<sup>1</sup> 

<sup>1</sup> Centre International de Recherche en Infectiologie (CIRI), Universit  de Lyon, Inserm, U1111, Universit  Claude Bernard Lyon 1, CNRS, UMR5308,  cole Nationale Sup rieur de Lyon, France

<sup>2</sup> UMR 5086 - CNRS / UCBL, IBCP, Lyon, France

## Keywords

epitranscriptomic modifications; interferon; ISG20; RNase; translational inhibition; virus inhibition

## Correspondence

A. Cimarelli, Centre International de Recherche en Infectiologie, 46 All e d'Italie, Lyon 69007, France  
Tel: +33 472728696  
E-mail: acimarel@ens-lyon.fr

(Received 15 November 2021, revised 31 January 2022, accepted 15 February 2022)

doi:10.1002/2211-5463.13382

Interferon-stimulated gene 20 kDa protein (ISG20) is a relatively understudied antiviral protein capable of inhibiting a broad spectrum of viruses. ISG20 exhibits strong RNase properties, and it belongs to the large family of DEDD exonucleases, present in both prokaryotes and eukaryotes. ISG20 was initially characterized as having strong RNase activity *in vitro*, suggesting that its inhibitory effects are mediated via direct degradation of viral RNAs. This mechanism of action has since been further elucidated and additional antiviral activities of ISG20 highlighted, including direct degradation of deaminated viral DNA and translational inhibition of viral RNA and nonself RNAs. This review focuses on the current understanding of the main molecular mechanisms of viral inhibition by ISG20 and discusses the latest developments on the features that govern specificity or resistance to its action.

## A general canvas for the cellular context of expression of ISG20

ISG20 is an exquisite interferon-regulated protein, and as such, what follows is a very general introduction to interferons (IFNs). The interferon system represents the

first line of response against microbial infections [1–3]. During infection, the cell recognizes pathogen-associated molecular patterns (PAMPs) via pattern recognition receptors (PRRs). Recognition initiates signaling cascades that converge onto several transcription factors (IRF3, IRF7, NF- B...) that migrate in the

## Abbreviations

+ssRNA, positive-strand RNA; AEN, apoptosis-enhancing nuclease; APOBEC3A, apolipoprotein B mRNA-editing enzyme catalytic subunit 3A; CB, cajal bodies; cccDNA, covalently closed circular DNA; CHIKV, chikungunya virus; CrPV, cricket paralysis virus; DENV, dengue virus; dsRNA/DNA, double-stranded RNA or DNA; E box, enhancer box; eif2, eukaryotic initiation factor 2; eif4, eukaryotic initiation factor 4; eIF4F, eukaryotic initiation factor 4F; Exo, exonuclease motif; HBV, hepatitis B virus; HCV, hepatitis C virus; HEM45, human estrogen-regulated transcript 45 protein; HIV-1, human immunodeficiency virus type 1; I, inosine; IAV, influenza A virus; IFIT1, interferon-tetrapeptide repeat protein 1; IFN, interferon; IFNAR1/2, interferon alpha and beta receptor subunit 1 or 2; IFN-I (IFN / ), type I interferon; IFN-II (IFN gamma), type II interferon; IRES, internal ribosome entry site; IRF1/3/7/9, interferon regulatory factor 1, 3, 7, or 9; ISG, interferon-stimulated gene; ISG20, interferon-stimulated gene 20kDa protein; ISG20L1, interferon-stimulated exonuclease gene 20kDa-like 1; ISG20L2, interferon-stimulated exonuclease gene 20kDa-like 2; ISRE, interferon-stimulated response element; JAK-STAT, Janus kinase signal transducer and activator of transcription; m1A, N1-methyladenosine; m5C, 5-methylcytosine; m6A, N6-methyladenosine modification; mISG20, mouse ISG20; mTOR, mammalian target of rapamycin; NP, nucleoprotein; NS, nonstructural; P bodies, processing bodies; PAMP, pathogen-associated molecular pattern; pgRNA, pregenomic RNA; Poly(I:C), polyinosinic:polycytidylic acid; PRR, pattern recognition receptor; rDNA, relaxed circular DNA; RdRp, RNA-dependent RNA polymerase; RIG-I, retinoic acid-inducible gene I; RNaseL, ribonuclease L (for latent); RT, reverse transcriptase; SARS-CoV, severe acute respiratory syndrome coronavirus; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; Sp-1, specificity protein 1; -ssRNA, negative-strand RNA; ssRNA/DNA, single-stranded RNA or DNA; STAT1/2, signal transducer and activator of transcription 1 or 2; TLR3, Toll-like receptor 3; UMP, uridine 5'-monophosphate; USF-1, upstream stimulatory factor 1; VEEV, Venezuelan equine encephalitis virus; VSV, vesicular stomatitis virus; WNV, West Nile virus; WT, wild-type; YTHDF2, YTH-domain family 2;  , epsilon;  , pseudouridine.

nucleus and lead to the synthesis of type I interferons (IFN-I; IFN $\alpha/\beta$ ). Contrary to more specialized interferons (i.e., types II and III, which are specific to immune cells), IFN-Is are instrumental to induce a generalized antiviral state and to start innate immune responses.

In humans, the IFN locus on human chromosome 9p contains 13 functional  $\alpha$  genes (*IFN $\alpha$ 1*, *IFN $\alpha$ 2*, *IFN $\alpha$ 4*, *IFN $\alpha$ 5*, *IFN $\alpha$ 6*, *IFN $\alpha$ 7*, *IFN $\alpha$ 8*, *IFN $\alpha$ 10*, *IFN $\alpha$ 13*, *IFN $\alpha$ 14*, *IFN $\alpha$ 16*, *IFN $\alpha$ 17*, and *IFN $\alpha$ 21*), one  $\beta$  gene situated syntenically (IFN $\beta$ 1) [4,5], and three more distantly related IFN genes (*IFN $\epsilon$* , *IFN $\kappa$* , and *IFN $\omega$* ). Members of the IFN family exhibit from 30 to 80% sequence identity, and all are ligands for the cell surface receptors IFN $\alpha$  receptor subunits 1 and 2 (IFNAR1/2). The ligand–receptor interaction leads to the activation of the Janus kinase signal transducer and activator of transcription (JAK–STAT) signaling pathway and to the upregulation of several hundreds of genes, referred to as interferon-stimulated genes (ISGs). These genes exert numerous effects aimed at blocking the microbe at multiple steps (entry into the cell, RNA processing, translation, etc.) [4–6].

In a very simplified manner then, this innate defense mechanism recognizes common features present in a given class of pathogens (reverse transcription as a blueprint for retroviruses, lipopolysaccharide for bacteria, etc.) in cells that are compromised. Through the rapid diffusion of signals (the IFNs), this leads to a generalized antimicrobial cellular state in which cells become more resistant to infection. This intrinsic cellular resistance is ultimately mediated by interferon-sensitive genes (ISGs) that act at multiple layers to inhibit pathogen replication either by directly targeting it or by indirectly affecting cellular processes normally required for its replication.

If the mechanism of action of some ISGs is well described in the literature, the role of many others remains either unknown or unclear.

Among ISGs that act as antiviral effectors and that are thus thought to inhibit virus infection by directly binding or altering one of their components is the interferon-stimulated gene 20 kDa protein (ISG20), also known as human estrogen-regulated transcript 45 protein (HEM45) [7,8]. ISG20 is the second RNase described to be induced by the IFN system after RNase L [9]. Despite the fact that initial studies reported a simple and direct explanation for viral inhibition by ISG20 (i.e., ISG20 directly degrades viral RNA thanks to its RNase enzymatic properties) [10], the following studies have somewhat complexified our view of how ISG20 may act. The following paragraphs will specifically focus on the discovery of ISG20, on the mechanisms

that over the years have been ascribed to ISG20 and to open questions that may help decorticate the antiviral functions of this broad-spectrum antiviral factor.

## A historical overview of the discovery of ISG20

ISG20 was discovered in 1997 in Daudi cells as a new interferon-regulated protein [7]. Later studies demonstrated that ISG20 could be induced transcriptionally by both type I ( $\alpha/\beta$ ) and type II ( $\gamma$ ) IFN through a unique interferon-stimulated response element (ISRE) present in its promoter region, under the control of the IFN regulatory factor 1 (IRF1) [11–13]. Despite being clearly upregulated by IFNs, ISG20 exhibits also variable expression levels at the basal state in several cell types under the control of different transcription factor, as the specificity protein 1 (SP-1) or the upstream stimulatory factor 1 (USF-1), so that ISG20 could have a physiological impact on cellular functions also in the absence of ongoing interferon responses [11]. Independently from the study of Gongora *et al.*, ISG20 was identified as a human estrogen-regulated transcript (HEM45) in breast cancer cell lines [8], hence its double name ISG20/HEM45. Several reports suggest a link between the tumorigenic process and ISG20 [14–17], although the exact mechanism through which ISG20 could influence this process remains unclear [18–20].

In humans, *isg20* is located on the chromosome 15q26.1 and codes for a 181 amino acid protein [10,21]. Humans code for two additional proteins bearing high homology with ISG20: the interferon-stimulated exonuclease gene 20 kDa-like 1 and 2 (ISG20L1 and ISG20L2, which share 53 and 48% identity with ISG20, respectively) (Fig. 1A). *isg20l1* lays just next to *isg20* and has been involved in apoptosis, hence its name of apoptosis-enhancing nuclease (AEN), while *isg20l2* is located on chromosome 1q23.1 and has been involved in ribosome biogenesis. At this time, neither ISG20L1 nor ISG20L2 has been reported to possess antiviral activities [22–24].

## ISG20, a member of the DnaQ-like (or DEDD) 3′-5′ exonuclease domain superfamily

The DEDD superfamily of 3′-exonucleases includes a large number of prokaryotic and eukaryotic enzymes that cleave one at a time from the 3′-RNA and/or DNA terminus [25]. Exoribonucleases of this family can be autonomous enzymes or integrated into larger replicative complexes, supporting, for example, the proofreading functions of DNA polymerases, of DNA



Asp94 (Fig. 1C, adapted from Horio *et al.* [29], PDB:1WLJ). Only the  $Mn^{2+}$  showed stimulatory effects on ISG20 activity, whereas other bivalent cations such  $Mg^{2+}$ ,  $Ca^{2+}$ , or  $Zn^{2+}$  did not [10]. Importantly, the conserved His149 interacts with the O2P of the UMP phosphate group, suggesting that this residue cannot be responsible for the preference shown by ISG20 for ssRNA to ssDNA [10,29]. This preference might be exerted by Met14 and Arg53 (Fig. 1C), because they directly interact with the ribose 2'-OH group of the UMP [10,29]. However, deeper mutational analyses or the structure of RNA-bound ISG20 are needed to better understand the molecular bases of ISG20 preference for RNA substrates. Furthermore, concerning the enzymatic mechanism, whether ISG20 acts as a processive or a distributive enzyme has also not been properly investigated so far.

While the purified ISG20, free from any cellular interactor, cleaves essentially ssRNA *in vitro*, the 3'-structured substrates are likely resistant [10,30]. This resistance has been well characterized in the case of the hepatitis B virus (HBV) RNA encapsidation signal called epsilon ( $\epsilon$ ) [30,31]. The  $\epsilon$  RNA is located near the 5' end of the HBV pregenomic RNA, and it exhibits a long 13-bp basal stem followed by a 6-nt bulge region also called priming loop, a 11-bp upper stem and a terminal hexaloop [32,33] (Fig. 1D). The first full-length structure of HBV  $\epsilon$  RNA has been recently solved using solution nuclear magnetic resonance (NMR) spectroscopy [34] (PDB:6VAR; Fig. 1D). This structure is of interest in light of the fact that the specific binding of ISG20 to the HBV  $\epsilon$  RNA takes place at the level of the lower stem, in a manner independent of the bulge sequence but dependent of the stem length. More than nine RNA base pairs of  $\epsilon$  lower stem seem to be necessary for stable ISG20: $\epsilon$  interaction. Despite this efficient binding, ISG20 was not able to cleave  $\epsilon$  RNA substrate [30]. Interestingly, a C-terminal truncated version of ISG20, with the ExoIII motif deleted, has been shown to abolish RNA binding [30]. This suggests that ExoIII could harbor the RNA-binding domain or, alternatively, that the coordination of the MnA by Asp154 is critical for RNA interaction (Fig. 1C).

Recently, a strong exonuclease activity of ISG20 has been reported on deoxyuridine-containing ssDNA. This finding is of interest because deoxyuridine-containing DNA is the result of the editing activity of a second important antiviral factor, the human apolipoprotein B mRNA-editing catalytic polypeptide-like 3A protein (APOBEC3A; in this case, the hepatitis B virus, HBV, and more generally of all members of this family) that deaminates

cytosines on viral genomes [35]. This activity is of interest because it connects the concerted action of two antiviral factors in the mutagenesis and degradation of viral genomes, as we will discuss more extensively below.

## Virus inhibition by ISG20

In light of its exquisite activity against RNA, it is perhaps not surprising, although simplistic, that ISG20 has been described to interfere with the replication of a large spectrum of RNA viruses. This list ranges from negative- or positive-strand RNA viruses such as *Rhabdoviridae* or *Flaviviridae* (for instance, the vesicular stomatitis virus, VSV, or the hepatitis C virus, HCV) to *Retroviridae* such as the human immunodeficiency type 1 virus (HIV-1), the genome of which is reverse-transcribed from a single-stranded RNA to double-stranded DNA [36–39] (for an exhaustive list, Table 1).

From a very schematic manner, negative- and positive-strand RNA viruses undergo an obligatory step of antigenomic RNA synthesis that leads to viral RNA amplification, translation of viral proteins, and ultimately the formation of novel virion particles (as schematically presented in Fig. 2A). The first round of viral translation, which is key to ensure the RNA amplification cycle, occurs either directly from the viral genome in the case of positive-strand RNA viruses or after pioneer transcription in the case of the negative-strand ones. The replication cycle of the viral genome is instead quite distinct in the case of *Retroviridae* and *Hepadnaviridae* that undergo a reverse transcription phase that leads to the formation of linear or circular DNA (Fig. 2B). These nucleic acids are first transported into the nucleus of host cells and then transcribed with the concurrence of viral components, similar to cellular genes. In this case therefore, genome amplification is promoted by viral-coded proteins, but is essentially assured by the cellular transcription machinery, and only RNAs of positive polarity are produced during infection [40,41]. Of note, antigenome RNA transcription has been reported for some retroviruses, but the role of these transcripts is either unclear or transcriptional regulation [42,43].

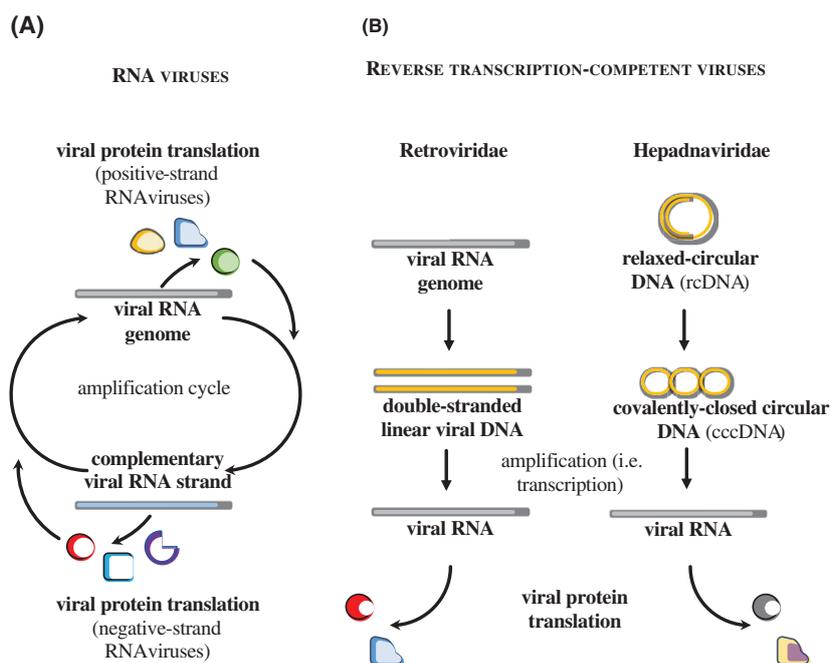
Even through the lenses of this simplified view of their mechanisms of replication, it is difficult to reconcile a single mechanism of inhibition based on the direct degradation of viral RNA against such diverse viruses. Indeed, while most RNA viruses targeted by ISG20 carry out their replication cycle entirely in the cell cytoplasm, some, as *Orthomyxoviridae*, replicate in the nucleus [44–47]. Also replicating away from the

**Table 1.** Virus susceptibility to ISG20 inhibition and known modifications of viral RNAs.

	ISG20 inhibition (Refs.)	Known viral RNA features			Refs.
		7mG cap (cap 1)	Epitranscr. Modif.	Poly(A) tail	
Negative-polarity RNA viruses					
<b>Orthomyxoviridae</b>	Yes; [37,96]	Yes <i>Cap-snatching</i>	m6A	Yes	[97,98]
Influenza A virus (IAV)					
<b>Peribunyaviridae</b>	Yes; [80]	Yes <i>Cap-snatching</i>	ND	No	[99,100]
Bunyamwera (BUNV)	No; [80]				
Cache Valley virus (CVV)					
Kairi virus (KRIV)					
Oropouche virus (OROV)					
Schmallenberg virus (SBV)					
Batama virus (BMAV)					
Boraceia virus (BORV)					
Tacaiuma virus (TCMV)					
Anopheles A virus (ANAV)					
Capim virus (CAPV)					
<b>Hantaviridae</b>	Yes; [80]				
Puumala virus (PUUV)					
<b>Nairoviridae</b>					
Dugbe virus (DUGV)					
<b>Phenuiviridae</b>	No; [80]				
Rift Valley fever virus (RVFV)					
Severe fever with thrombocytopenia syndrome virus (SFTSV)					
Heartland virus (HRTV)					
<b>Rhabdoviridae</b>	Yes; [37,39]	Yes	m6A	Yes	[101–103]
Vesicular stomatitis virus (VSV)					
Positive-polarity RNA viruses					
<b>Flaviviridae</b>	Yes; [24,104,105]	No	m6A, m5C, $\psi$ , m1A, I	No	[89,106–109]
Hepatitis C virus (HCV)					
Bovine viral diarrhea virus (BVDV)	Yes; [24]		ND	No	[107,110]
West Nile virus (WNV)	Yes; [38]	Yes	m6A (WNV)	No	[89,108,111–115]
Dengue virus (DENV)			m6A, m5C, $\psi$ , m1A, I (DENV)		
Yellow fever virus (YFV)					
Zika virus (ZIKV)	Yes; [24] Yes; [93]		m6A, I m6A, m5C, $\psi$ , m1A		
<b>Picornaviridae</b>	Yes; [24,37]	No	ND	Yes	[116–118]
Hepatitis A virus HAV					
Encephalomyocarditis virus (ECMV)					
<b>Togaviridae</b>	Yes; [12,119]	No <i>Cap 0*</i>	ND	Yes	[120,121]
Sindbis virus (SINV)					
Chikungunya virus (CHIKV)					
Venezuelan equine encephalitis virus (VEEV)					
<b>Coronaviridae</b>	No (1); [24] Yes (2); [122]	Yes	ND m6A (SARS-CoV-2)	Yes	[123–125]
Severe acute respiratory syndrome coronavirus (SARS-CoV-1) and SARS-CoV-2					
Viruses containing retroelements					
<b>Retroviridae</b>	Yes; [36,126,127]	Yes	m6A, m5C, $\psi$ , m1A, I	Yes	[108,128–131]
Human immunodeficiency virus type 1 (HIV-1)					
<b>Hepadnaviridae</b>	Yes; [21,132,133]	Yes	m6A	Yes	[50,133–135]
Hepatitis B virus (HBV)					

Virus families are indicated in bold.

\*Cap 0: m7GpppNp. Di- and trimethylated caps have also been described (m2,7G- and m2,2,7G-caps in the case of SINV). ND, not determined.



**Fig. 2.** Strategies for viral genome amplification used by viruses reported to be inhibited by ISG20. (A) Simplified overview of the RNA amplification cycle used by positive- and negative-strand RNA viruses. Upon entry, the RNA genome of positive-strand RNA viruses is translated into viral proteins that promote synthesis of the complementary RNA strand and ensure the amplification cycle. In the case of the latter instead, a first round of pioneer RNA synthesis is needed to ensure the first round of viral protein translation. (B) Schematic overview of the strategy used by *Retroviridae* and *Hepadnaviridae* to replicate through an intermediate reverse transcription step. In this case, viral RNA is amplified through transcription of RNAs of positive polarity in the cell nucleus, a process that co-opts the cellular RNA polymerase II.

cell cytoplasm are *Retroviridae* and *Hepadnaviridae*, the viral RNAs of which are transcribed in the cell nucleus as *bona fide* cellular mRNAs, endowed with a 5' cap (7-methylguanylate, m7G) and a 3' poly A tail.

How can this enzyme distinguish between cellular and viral RNAs? One possible hypothesis would be that ISG20 is capable of discriminating RNAs based on specific features present, absent, or differentially present in one versus the other RNA species, as, for example, epitranscriptomic modifications, 5' cap, or poly A tails. However, while some of the viruses targeted by ISG20 lack some of these canonical features present on cellular RNAs, other viruses do possess them (Table 1). As such, the exact mechanism through which ISG20 identifies its RNA targets remains unclear.

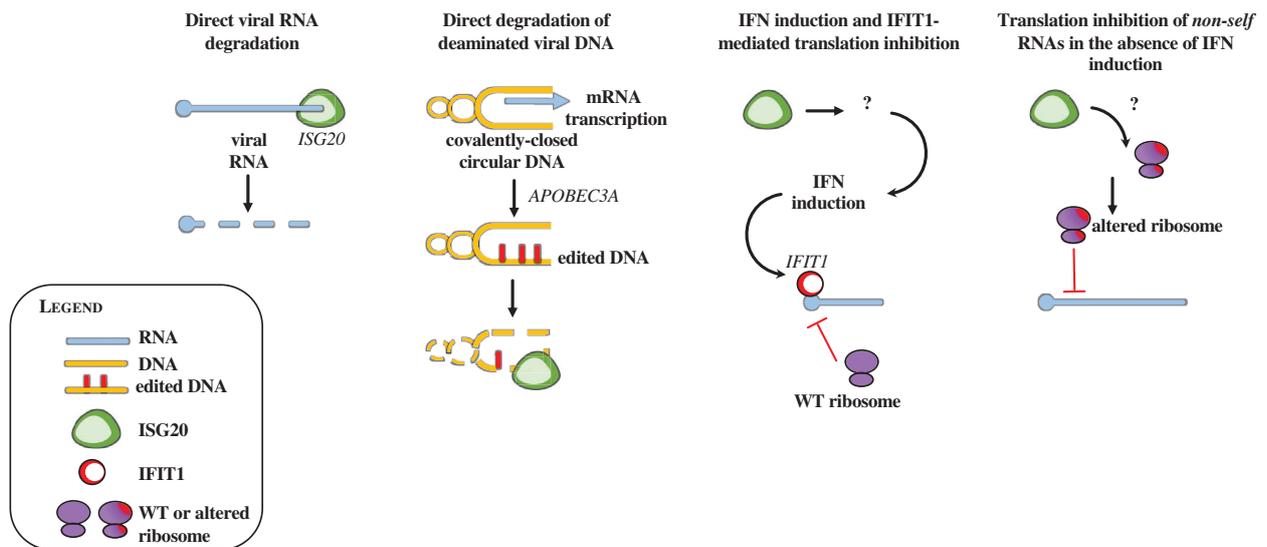
At present, the results obtained and confirmed by numerous laboratories concur in three experimental evidences: ISG20 behaves as a potent RNase *in vitro*; this enzyme inhibits the replication of numerous viruses and its enzymatic activity is key for this property, given that a single point mutation that disrupts the catalytic site of ISG20 both abolishes its ability to degrade RNA *in vitro* and impairs its antiviral behavior. The explanation of how this behavior underlines the antiviral mechanism of action of ISG20 is still a matter of debate, and currently, four models have been proposed: two in which ISG20 directly degrades its viral target and two in which ISG20 blocks the translation from viral RNA without degrading them. Given

that the RNase activity of ISG20 remains nonetheless important for this translational block, it follows that ISG20 must target a cellular RNA, whose identity remains, however, unknown.

### Mechanism 1: direct viral RNA degradation

The first model proposed on the basis of the experimental evidence described above was that ISG20 directly degraded viral RNA [30,36,37,48,49] (Fig. 3). Indeed, the expression of ISG20 led to lower accumulation of viral RNA and synthetic viral RNA mimics were also efficiently degraded *in vitro*.

In addition to the fact that the direct binding of WT ISG20 to viral RNA has not been formally demonstrated in cells, the major caveat of this model is that it fails to provide the basis through which ISG20 can distinguish viral from cellular RNA substrates. Indeed, ISG20 is capable of degrading most ssRNA *in vitro* [10,29], not just viral substrates so that the question of how ISG20 spares cellular RNAs remains unaddressed. At present, a single study reported the specific binding of ISG20 to the adenosine-methylated (m6A, adenosine in position 1907) form of the epsilon structure of the HBV RNA ( $\epsilon$ ), a particular stem-loop structure and unique site for m6A modification on the HBV transcripts (or two sites within two identical  $\epsilon$  in 5' or 3' ends for the pregenomic RNA) [50]. However, the association between ISG20 and this methylated



**Fig. 3.** Proposed mechanisms of viral inhibition by ISG20. The figure presents the four models of action that have been ascribed to ISG20. From left to right, in the first two models ISG20 has been proposed to directly degrade viral nucleic acid in its RNA or deaminated DNA forms (models 1 and 2, respectively). In models 3 and 4, ISG20 has been proposed to inhibit viral translation but not to promote direct viral RNA degradation. In model 3, translation inhibition is promoted through an unknown mechanism by ISG20 by the induction of IFN and more specifically of IFIT1, a well-known translation inhibitor that targets viral RNAs devoid of a proper 5' cap. In model 4, translation inhibition occurs independently from IFN induction and results in a defect in translation efficiency, likely due to the direct modification of the translation machinery. Also, in this case, the identity of the targeted RNAs remains unknown.

structure was mediated through a direct protein–protein interaction with the m6A reader YTH-domain family 2 (YTHDF2). Given the extent of the m6A modifications on cellular mRNAs [51], it is difficult to envision how YTHDF2 binding, alone, could help ISG20 discriminate between cellular and viral mRNAs. However, this finding is of interest because it raises the possibility that ISG20 connects with the complex epitranscriptomic machinery, perhaps in combination with other RNA features as, for instance, the polyA tail or the 5' cap. The heterogeneous modifications and variable presence of these modifications among the large panel of viruses targeted by ISG20 do not allow for the moment the identification of a unique feature driving susceptibility or resistance to ISG20.

### Mechanism number 2. Direct degradation of deaminated viral DNA

While the activity of ISG20 on ssDNA is generally lower when compared to ssRNA *in vitro*, a recent study indicated that it could be increased in the presence of deoxyuridines (dU) [35]. In so doing, the authors of the study proposed an interesting model of cooperativity between two cellular antiviral factors against HBV, connecting the cytosine deamination activity of the apolipoprotein B mRNA-editing enzyme

catalytic subunit 3A (APOBEC3A) to the ensuing degradative activity of ISG20.

Members of the APOBEC3 family (A through H in humans) are prototype antiviral factors well studied in the context of retroviral infection [52,53]. In this case, APOBEC3s are incorporated into virion particles and act during the process of reverse transcription in the cytoplasm of target cells by deaminating cytosines in the first strand of reverse-transcribed viral DNA (the negative-strand). As a result, deaminated C > U sites are misread when this strand is used as a template for the synthesis of the second strand (the positive-strand) resulting in G-to-A mutations along the viral genome. For this family of viruses, the extent of accumulation of deleterious mutations along the viral genome (stop codons, mutations with negative effect on the protein functionality, etc.) can be so considerable that viral replication is severely impaired in a mechanism that has been referred to as death by mutagenesis [53,54].

The mechanism of cytosine deamination described here in the case of HBV is distinct from the one of *Retroviridae*: first because it occurs in the nucleus and is carried out by APOBEC3A, which is together with APOBEC3B, the sole nuclear APOBEC3 member; and second, because it is carried out on double-stranded covalently closed circular DNA (cccDNA) [55]. Its outcome is therefore not the introduction of mutations

in the viral genome, but rather the stamping of viral DNA for ISG20-mediated degradation [35]. Given that APOBEC3 proteins require a single-strand DNA intermediate to act [56,57], it is likely that APOBEC3A acts on transcriptionally active cccDNA in which the transcription machinery opens up the double-stranded DNA (Fig. 3).

As said before, this mechanism of action is of interest because it models the concerted action of two antiviral factors. Several questions remain, however, unaddressed around the specificity of action of ISG20. First, the major enzymatic functions of ISG20 are directed against RNA, and second, most DNA viruses do replicate in the nucleus, but for the most part seem resistant to ISG20 [37,39]. It will be therefore important to understand how these viruses avoid the combined action of APOBEC3A-ISG20. The restricted pattern of expression of APOBEC3A to cells of the myeloid lineage could provide a first line of specificity to the action of ISG20 [58], but certainly cannot explain it alone. Third and more importantly, the cellular genome is constantly exposed to spontaneous deamination [59,60]. It will be thus imperative to determine how the cellular genome protects itself from ISG20.

### **Mechanism 3: IFN induction and IFIT1-mediated translation inhibition of viral RNA**

For most highly replicating RNA viruses, the first round of translation of viral proteins is capital to ensure viral RNA amplification. As such, an alternative explanation for the lower accumulation of viral RNA observed in numerous studies in the presence of ISG20 could be an inhibitory effect of the first round of viral translation, which would severely undermine the extent of viral RNA amplification.

Translation inhibition has been reported in mouse embryonic fibroblasts (MEF) overexpressing the murine ISG20 (mISG20) that shares 82% identity with human ISG20 (Fig. 3). In this study and in agreement with other studies, the authors found that ISG20-mediated inhibition of the chikungunya virus (CHIKV) or Venezuelan equine encephalitis virus (VEEV) did not involve the degradation of viral RNAs, but rather the inhibition of translation of viral proteins [12]. Interestingly, translation inhibition was indirectly due to the stimulation of IFN-I responses and more specifically to the expression of the interferon-induced protein with tetratricopeptide 1 (IFIT1), a well-known translation suppressor protein that targets uncapped viral RNAs [12,61,62]. In this

model therefore, ISG20 inhibits viral translation indirectly through the induction of an IFN response and IFIT1. Despite the fact that a later study failed to confirm IFN and IFIT1 induction by ISG20 [39], this study opens up the possibility that ISG20 can drive larger responses besides direct viral inhibition.

This model leaves for the moment unanswered two major questions. First, ISG20 is itself an IFN-regulated protein so that its reported induction of IFN would necessarily occur in an environment already primed by IFN. As such, either ISG20 acts as a positive regulator to potentiate or sustain IFN responses over time (perhaps by processing viral RNAs as RNaseL), or this mechanism is restricted to cell types in which ISG20 expression is induced by other signals, as, for example, estrogens. Second, this model does not explain the broad mechanism of viral inhibition of ISG20, as the spectrum of viruses inhibited by ISG20 is larger than those targeted by IFIT1 and it extends to viruses with a cellular-like m7G-5' cap. Importantly, given that the exonuclease activity of ISG20 is important for this behavior, the identity of the RNAs (cellular or viral) targeted by ISG20 to induce IFN responses remains to be identified.

Mechanism 4: Translation inhibition of nonself RNAs in the absence of IFN induction. This mechanism was proposed recently by our laboratory following experimental observations that ISG20 exerted a strong impairment in the translation of viral proteins following VSV infection in spite of a very mild decrease in viral RNAs (Fig. 3) and in the absence of IFN-I or IFIT1 induction [39].

Interestingly, a similar behavior could be reproduced upon ectopic transfection of plasmid DNAs independently from the nature of the reporter gene examined (i.e., cellular or viral, spliced or unspliced). On the contrary, ISG20 had no effect when this exact reporter cassette was stably integrated into the host genome.

Translation inhibition was independent from initiation factors as it affected equally reporter genes under the control of the cricket paralysis virus (*CrPV*) 5'-untranslated region that drives initiation of factorless translation [63]. Interestingly, the suppressive effects of ISG20 were accompanied by its partial recruitment to cytoplasmic processing bodies (P bodies), important and dynamic sites of RNA metabolism [64], although it was not possible to determine whether this association was causal or secondary to the translational block itself [39].

VSV infection and transient DNA transfection do not share many features in common. However, they both represent the massive flooding of the cell with *nonself* genetic material, and in both cases, the process

of cellular translation is largely redirected toward a single (in the case of ectopic DNA transfection) or a narrow set (in the case of viral infection) of RNA transcripts, a situation that is unusual for the cell and often associated with viral infection.

Given the absence of target RNA degradation, but in light of the importance of an intact RNase activity, we proposed that ISG20 could affect cellular translation by degrading a cellular RNA substrate, which affected proportionally the most actively translated RNAs, working as a threshold translational controller.

This model is of interest because it is relatively independent from a single RNA feature. However, it can be influenced by all of them and in particular by the extent of epitranscriptomic modifications that are added cotranscriptionally and thus depend on the chromatin status of the target gene.

In this respect, rRNAs and small nucleolar RNAs (rRNAs and snoRNAs, respectively) represent interesting potential targets with which ISG20 could alter the overall translational abilities of the cell. snoRNAs are small nonprotein coding RNAs (60–300 nt) involved in a number of physiological processes among the post-translational modification of rRNAs, process in which they guide the methylase fibrillarin on specific sites of rRNAs promoting 2'-O-methylation modifications [65,66]. The weight of these modifications in the activity of ribosomes remains to be fully apprehended. However, an increasing number of reports support the notion that specialized subsets of ribosomes are formed in response to certain stimuli [67–69], and in this context, ISG20 could promote the formation of specialized pools of ribosomes, thus influencing the translational capacity of the cell.

This model would accommodate inhibition of a broad spectrum of viruses targeted by ISG20 according to a single mechanism of action. However, also in this case the identity of the RNAs targeted by ISG20 remains unidentified.

### Open question 1: the intracellular whereabouts of ISG20

The intracellular distribution of ISG20 has not been the subject of intense studies so far. Its small size (around 20 kDa) allows it to be fairly well distributed in both cell cytoplasm and nucleus. A subtler localization of ISG20 has been described upon strong detergent permeabilization in promyelocytic leukemia (PML) bodies, nucleolus, and Cajal bodies [7,70]. PML bodies are involved in several physiological processes including antiviral responses, and they are often disrupted by viral infection [71–73]. Instead, nucleolus

and CB are sites of formation and processing of rRNA and small nuclear ribonucleoproteins (snRNPs). An extensive literature exists on these discrete nuclear bodies, and the reader is invited to read some of the following reviews [74–78]. In light of the association of a fraction of ISG20 with these structures, it is thus possible that ISG20 is involved in small nuclear RNA (snRNA) and rRNA biogenesis and/or maturation [70], although this would support an indirect mechanism of viral inhibition by ISG20, rather than a direct one.

In the case of translation inhibition, our laboratory also reported that a fraction of cytoplasmic ISG20 could migrate (or induce) processing bodies (P bodies), which are cytoplasmic membraneless compartments involved in RNA decay and/or control of translation by RNA storage [39,79], in line with their effect in translation inhibition.

However, the dynamics of ISG20 and an extensive characterization of its intracellular distribution during homeostasis or during the context of infection are cruelly lacking and this may yield important clues as to the mechanisms of inhibition of ISG20. For instance, with few exceptions [39], a single mutant is often used in several studies (D94, or similar mutations in the DEDD residues), but the use of additional mutants could extend our comprehension of the links between cellular localization of ISG20 and its antiviral activities. Similarly, although ISG20 is distributed in both cell cytoplasm and nucleus it is unclear whether trafficking between one and the other location is important for the antiviral activity. Certainly, this will be an interesting area of investigation for future studies.

### Open question 2: the specificity determinants of ISG20

While a large number of viruses are susceptible to the action of ISG20, several remain resistant to it. This includes not only DNA viruses such as adenoviruses or the Epstein–Barr virus but also RNA viruses such as several, although not all, members of *Phenuiviridae* [80], as well as SARS-CoV [24]. *in vitro*-driven evolution toward resistance to ISG20 in Bunyamwera orthobunyavirus (BUNV), prototype virus of *Peribunyaviridae* family, has led to interesting hints about the determinants that may modulate susceptibility and resistance to this restriction factors [80]. Several mutations were indeed identified in BUNVs that adapted to replicate in the presence of ISG20 expression. These mutations occurred in the three segments of the viral genome (short (S), large (L), and medium (M)). Despite the fact that mutations in the different genome

segments exhibited unequal contributions to the extent of viral replication likely due to their respective roles in this process, adaptive mutations were observed in potential structural RNA elements (mutations outside open reading frames, ORFs), as well as in the N and NSs proteins. Although the relative contribution of these mutations to ISG20 resistance has not been further dissected, these results indicate that both RNA-binding proteins as the viral N protein and *cis*-acting elements on the viral RNA may influence susceptibility to ISG20.

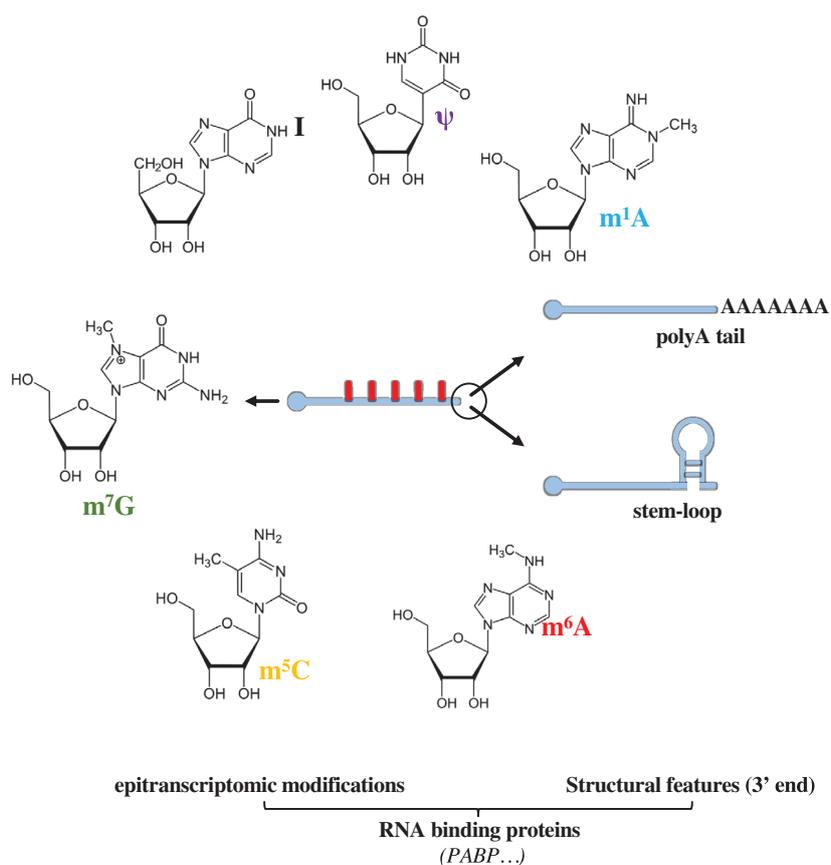
Viral proteins may indeed, as cellular proteins, shield viral RNAs from the degradative action of ISG20, and similarly, particular stem-loop structures may also protect RNA, as ISG20 displays no reported activity against double-stranded RNA. In this respect, the RNA of bunyaviruses is not polyadenylated, but contains a conserved 3' stem-loop structure that may be important to mediate the resistance of some, although not all, members of this viral order.

In addition to the already-mentioned 5'-cap and poly(A) tail in the 3' end, cellular mRNAs are also heavily modified either cotranscriptionally or shortly after transcription by the addition of the following

modifications: ribose 2'-O-methylation, pseudouridine, inosine, m<sup>6</sup>A, and 5-methylcytosine (Fig. 4). Collectively, these dynamic modifications are referred to as epitranscriptomic modifications and are involved in all aspects of the RNA biology, such as stability or efficiency of translation [81,82], aspects that they influence either directly or indirectly through the recruitment of a large number of cellular proteins [83,84]. Readers are invited to read the following reviews for a more complete overview of epitranscriptomic modifications [84–88].

These modifications are mentioned here because an increasing number of reports indicate that viruses themselves can hijack these processes to their advantage to modulate the metabolism of their own RNAs, in addition to negative impact cellular defenses [89–91]. It is for the moment unclear whether the activity of ISG20 can be altered by the variety or the abundance of specific epitranscriptomic marks, because exhaustive studies of all these modifications in the context of specific viruses are lacking (Table 1).

However, we can hypothesize that the extent and or abundance of epitranscriptomic marks in a given RNA could modulate either directly or indirectly through



**Fig. 4.** Features expected to modulate RNA susceptibility to ISG20. Intrinsic features of the RNA as poly A tail or secondary structures, as well as epitranscriptomic modifications that may in principle affect susceptibility to ISG20. The RNA-binding proteins that can dock on the RNA as a consequence of these modifications are not shown for simplicity. m6A: N6-methyladenosine, m5C: 5-methylcytosine, m7G: N7-methylguanosine, I: inosine,  $\psi$ : pseudouridine, m1A: N1-methyladenosine.

RNA-binding proteins the action of ISG20. The weight of these modifications in the antiviral functions of ISG20 can be direct, for instance, by providing protection from direct viral RNA degradation, or indirect, for example, by altering the efficiency of translation by specialized subsets of ribosomes, as proposed in the translation inhibition models.

### Concluding remarks

Overall, several studies indicate that ISG20 is an important and broad antiviral effector during innate immune responses. While the exonuclease activity of ISG20 is deemed necessary for its antiviral activity toward multiple viruses, the identity of the nucleic acids targeted by ISG20 remains unclear, ranging from viral RNA to deaminated single-stranded DNA or to cellular RNA targets that may influence translation. Rarely mentioned, high levels of ISG20 expression lead to cell death [39,92] and it is at present unknown whether regulatory mechanisms exist that keep in check this deleterious effect. In addition to highlighting the importance of carefully controlling ISG20 expression levels during viral replication studies, these results indicate that the specificity of ISG20 for its viral targets is not absolute. Indeed, under certain circumstances the cell can also be susceptible to ISG20, stressing all the more the need to identify the RNA features that govern susceptibility to this factor.

Despite the fact that the underlying molecular mechanism leading to viral inhibition remains to be defined precisely, ISG20 has shown potential for use as broad antiviral inhibitor after vehiculation inside the cell as fusion protein with antibodies (Fc fragment) or with Tat cell-penetrating peptides [93,94]. In addition, ISG20 expression has been also associated with glioma and radioresistance in oral cancer cells and other types of cancer [18,48,95]. As such, the molecular definition of its mechanism of action in the cell could spur novel developments in antiviral treatment and yield clues to other cellular processes of broader interest. Future studies will certainly shed new light on several key aspects of the biology of ISG20.

### Acknowledgments

Work on ISG20, as well as the PhD fellowship of SD, is supported by a grant from the French Agence Nationale de la Recherche (project ANR-20-CE15-0025-01). AC is a permanent researcher of the Centre Nationale de la Recherche Scientifique (CNRS).

### Conflict of interest

The authors have declared that there is no conflict of interest.

### Author contribution

SD, CL, FF, and AC conceived and wrote the manuscript.

### References

- 1 Isaacs A, Lindenmann J. Virus interference: I. The interferon. By Alick Isaacs and Jean Lindenmann, 1957. *CA Cancer J Clin.* 1957;**38**:280–90.
- 2 Schoggins JW. Interferon-stimulated genes: what do they all do? *Annu Rev Virol.* 2019;**6**:567–84.
- 3 McNab F, Mayer-Barber K, Sher A, Wack A, O'Garra A. Type I interferons in infectious disease. *Nat Rev Immunol.* 2015;**15**:87–103.
- 4 Manry J, Laval G, Patin E, Fornarino S, Itan Y, Fumagalli M, et al. Evolutionary genetic dissection of human interferons. *J Exp Med.* 2011;**208**:2747–59.
- 5 Capobianchi MR, Uleri E, Caglioti C, Dolei A. Type I IFN family members: Similarity, differences and interaction. *Cytokine Growth Factor Rev.* 2015;**26**:103–11.
- 6 Sadler AJ, Williams BRG. Interferon-inducible antiviral effectors. *Nat Rev Immunol.* 2008;**8**:559–68.
- 7 Gongora C, David G, Pintard L, Tissot C, Hua TD, Dejean A, et al. Molecular cloning of a new interferon-induced PML nuclear body-associated protein. *J Biol Chem.* 1997;**272**:19457–63.
- 8 Pentecost BT. Expression and estrogen regulation of the HEM45 mRNA in human tumor lines and in the rat uterus. *J Steroid Biochem Mol Biol.* 1998;**64**:25–33.
- 9 Li X-L, Blackford JA, Hassel BA. RNase L Mediates the Antiviral Effect of Interferon through a Selective Reduction in Viral RNA during Encephalomyocarditis Virus Infection. *J Virol.* 1998;**72**:2752–9.
- 10 Nguyen LH, Espert L, Mechti N, Wilson DM. The human interferon- and estrogen-regulated *ISG20/HEM45* Gene product degrades single-stranded RNA and DNA in Vitro †. *Biochemistry.* 2001;**40**:7174–9.
- 11 Gongora C, Degols G, Espert L, Hua TD, Mechti N. A unique ISRE, in the TATA-less human *Isg20* promoter, confers IRF-1-mediated responsiveness to both interferon type I and type II. *Nucleic Acids Res.* 2000;**28**:2333–41.
- 12 Weiss CM, Trobaugh DW, Sun C, Lucas TM, Diamond MS, Ryman KD, et al. The interferon-induced exonuclease ISG20 exerts antiviral activity through upregulation of type I interferon response proteins. *mSphere.* 2018;**3**:18.

- 13 Imaizumi T, Mechti N, Matsumiya T, Sakaki H, Kubota K, Yoshida H, et al. Expression of interferon-stimulated gene 20 in vascular endothelial cells. *Microbiol Immunol.* 2008;**52**:30–5.
- 14 Wang J, Wang Y, Xing P, Liu Q, Zhang C, Sui Y, et al. Development and validation of a hypoxia-related prognostic signature for breast cancer. *Oncol Lett.* 2020;**20**:1906–14.
- 15 Alsheikh HAM, Metge BJ, Pruitt HC, Kammerud SC, Chen D, Wei S, et al. Disruption of STAT5A and NMI signaling axis leads to ISG20-driven metastatic mammary tumors. *Oncogenesis.* 2021;**10**:45.
- 16 Qu X, Shi Z, Guo J, Guo C, Qiu J, Hua K. Identification of a novel six-gene signature with potential prognostic and therapeutic value in cervical cancer. *Cancer Med.* 2021;**10**:6881–96.
- 17 Yu J, Liu T-T, Liang L-L, Liu J, Cai H-Q, Zeng J, et al. Identification and validation of a novel glycolysis-related gene signature for predicting the prognosis in ovarian cancer. *Cancer Cell Int.* 2021;**21**:353.
- 18 Miyashita H, Fukumoto M, Kuwahara Y, Takahashi T, Fukumoto M. ISG20 is overexpressed in clinically relevant radioresistant oral cancer cells. *Int J Clin Exp Pathol.* 2020;**13**:1633–9.
- 19 Xu T, Ruan H, Gao S, Liu J, Liu Y, Song Z, et al. ISG20 serves as a potential biomarker and drives tumor progression in clear cell renal cell carcinoma. *AGING.* 2020;**12**:1808–27.
- 20 Lin S-L, Wu S-M, Chung I-H, Lin Y-H, Chen C-Y, Chi H-C, et al. Stimulation of interferon-stimulated gene 20 by thyroid hormone enhances angiogenesis in liver cancer. *Neoplasia.* 2018;**20**:57–68.
- 21 Van Tong H, Hoan NX, Binh MT, Quyen DT, Meyer CG, Song LH, et al. Interferon-stimulated gene 20 kDa protein serum levels and clinical outcome of hepatitis B virus-related liver diseases. *Oncotarget.* 2018;**9**:27858–71.
- 22 Lee J-H, Koh YA, Cho C-K, Lee S-J, Lee Y-S, Bae S. Identification of a novel ionizing radiation-induced nuclease, AEN, and its functional characterization in apoptosis. *Biochem Biophys Res Comm.* 2005;**337**:39–47.
- 23 Couté Y, Kindbeiter K, Belin S, Dieckmann R, Duret L, Bezin L, et al. ISG20L2, a novel vertebrate nucleolar exoribonuclease involved in ribosome biogenesis. *Mol Cell Proteomics.* 2008;**7**:546–59.
- 24 Zhou Z, Wang N, Woodson SE, Dong Q, Wang J, Liang Y, et al. Antiviral activities of ISG20 in positive-strand RNA virus infections. *Virology.* 2011;**409**:175–88.
- 25 Shevelev IV, Hübscher U. The 3'–5' exonucleases. *Nat Rev Mol Cell Biol.* 2002;**3**:364–76.
- 26 Mason PA, Cox LS. The role of DNA exonucleases in protecting genome stability and their impact on ageing. *AGE.* 2012;**34**:1317–40.
- 27 Zuo Y. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res.* 2001;**29**:1017–26.
- 28 Viswanathan M, Lovett ST. Exonuclease X of *Escherichia coli*. *J Biol Chem.* 1999;**274**:30094–100.
- 29 Horio T, Murai M, Inoue T, Hamasaki T, Tanaka T, Ohgi T. Crystal structure of human ISG20, an interferon-induced antiviral ribonuclease. *FEBS Lett.* 2004;**577**:111–6.
- 30 Liu Y, Nie H, Mao R, Mitra B, Cai D, Yan R, et al. Interferon-inducible ribonuclease ISG20 inhibits hepatitis B virus replication through directly binding to the epsilon stem-loop structure of viral RNA. *PLoS Pathog.* 2017;**13**:e1006296.
- 31 Beck J. Hepatitis B virus replication. *WJG.* 2007;**13**:48.
- 32 Kramvis A, Kew MC. Structure and function of the encapsidation signal of hepadnaviridae. *J Viral Hepat.* 1998;**5**:357–67.
- 33 Flodell S. The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. *Nucleic Acids Res.* 2002;**30**:4803–11.
- 34 LeBlanc RM, Kasprzak WK, Longhini AP, Oleginski LT, Abulwerdi F, Ginocchio S, et al. Structural insights of the conserved “priming loop” of hepatitis B virus pre-genomic RNA. *J Biomol Struct Dyn.* 2021;**44**:1–13.
- 35 Stadler D, Kächele M, Jones AN, Hess J, Urban C, Schneider J, et al. (Interferon-induced degradation of the persistent hepatitis B virus cccDNA form depends on ISG20. *EMBO Rep.* 2021;**22**:68.
- 36 Espert L, Degols G, Lin Y-L, Vincent T, Benkirane M, Mechti N. Interferon-induced exonuclease ISG20 exhibits an antiviral activity against human immunodeficiency virus type 1. *J Gen Virol.* 2005;**86**:2221–9.
- 37 Espert L, Degols G, Gongora C, Blondel D, Williams BR, Silverman RH, et al. ISG20, a New interferon-induced rnaase specific for single-stranded RNA, defines an alternative antiviral pathway against RNA genomic viruses. *J Biol Chem.* 2003;**278**:16151–8.
- 38 Jiang D, Weidner JM, Qing M, Pan X-B, Guo H, Xu C, et al. Identification of five interferon-induced cellular proteins that inhibit west nile virus and dengue virus infections. *J Virol.* 2010;**84**:8332–41.
- 39 Wu N, Nguyen X-N, Wang L, Appourchaux R, Zhang C, Panthou B, et al. The interferon stimulated gene 20 protein (ISG20) is an innate defense antiviral factor that discriminates self versus non-self translation. *PLoS Pathog.* 2019;**15**:e1008093.
- 40 Hu J, Seeger C. Hepadnavirus genome replication and persistence. *Cold Spring Harb Perspect Med.* 2015;**5**:a021386.
- 41 Tsukuda S, Watashi K. Hepatitis B virus biology and life cycle. *Antiviral Res.* 2020;**182**:104925.

- 42 Gaudray G, Gachon F, Basbous J, Biard-Piechaczyk M, Devaux C, Mesnard J-M. The Complementary Strand of the Human T-Cell Leukemia Virus Type 1 RNA Genome Encodes a bZIP Transcription Factor That Down-Regulates Viral Transcription. *J Virol.* 2002;**76**:12813–22.
- 43 Savoret J, Mesnard J-M, Gross A, Chazal N. Antisense transcripts and antisense protein: a new perspective on human immunodeficiency virus type 1. *Front Microbiol.* 2021;**11**:625941.
- 44 Blondel D, Maarifi G, Nisole S, Chelbi-Alix M. Resistance to rhabdoviridae infection and subversion of antiviral responses. *Viruses.* 2015;**7**:3675–702.
- 45 Guu TSY, Zheng W, Tao YJ. Bunyavirus: Structure and Replication. In: Rossmann MG, Rao VB, editors. *Viral Molecular Machines.* US, Boston, MA: Springer; 2012. p. 245–66.
- 46 Sun X, Li D, Wang Z, Liu Q, Wei Y, Liu T. A dimorphism shift of hepatitis B virus capsids in response to ionic conditions. *Nanoscale.* 2018;**10**:16984–9.
- 47 Couch RB. Orthomyxoviruses. In: Baron S, editors. *Medical Microbiology.* 4th ed. Galveston, TX: University of Texas Medical Branch at Galveston; 1996.
- 48 Zheng Z, Wang L, Pan J. Interferon-stimulated gene 20-kDa protein (ISG20) in infection and disease: Review and outlook. *Intractable Rare Dis Res.* 2017;**6**:35–40.
- 49 Leong CR, Funami K, Oshiumi H, Mengao D, Takaki H, Matsumoto M, et al. Interferon-stimulated gene of 20 kDa protein (ISG20) degrades RNA of hepatitis B virus to impede the replication of HBV *in vitro* and *in vivo*. *Oncotarget.* 2016;**7**:68179–93.
- 50 Imam H, Khan M, Gokhale NS, McIntyre ABR, Kim G-W, Jang JY, et al. N<sup>6</sup>-methyladenosine modification of hepatitis B virus RNA differentially regulates the viral life cycle. *Proc Natl Acad Sci USA.* 2018;**115**:8829–34.
- 51 Shi H, Wei J, He C. Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Mol Cell.* 2019;**74**:640–50.
- 52 Malim MH. APOBEC proteins and intrinsic resistance to HIV-1 infection. *Phil Trans R Soc B.* 2009;**364**:675–87.
- 53 Harris RS, Dudley JP. APOBECs and virus restriction. *Virology.* 2015;**479–480**:131–45.
- 54 Vartanian J-P, Sommer P, Wain-Hobson S. Death and the retrovirus. *Trends Mol Med.* 2003;**9**:409–13.
- 55 Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat Struct Mol Biol.* 2010;**17**:222–9.
- 56 Bohn M-F, Shandilya SMD, Silvas TV, Nalivaika EA, Kouno T, Kelch BA, et al. The ssDNA Mutator APOBEC3A is regulated by cooperative dimerization. *Structure.* 2015;**23**:903–11.
- 57 Hoopes JI, Cortez LM, Mertz TM, Malc EP, Mieczkowski PA, Roberts SA. APOBEC3A and APOBEC3B preferentially deaminate the lagging strand template during DNA replication. *Cell Rep.* 2016;**14**:1273–82.
- 58 Berger G, Durand S, Fargier G, Nguyen X-N, Cordeil S, Bouaziz S, et al. APOBEC3A Is a Specific Inhibitor of the Early Phases of HIV-1 Infection in Myeloid Cells. *PLoS Pathog.* 2011;**7**:e1002221.
- 59 Duncan BK, Miller JH. Mutagenic deamination of cytosine residues in DNA. *Nature.* 1980;**287**:560–1.
- 60 Ehrlich M, Zhang X-Y, Inamdar NM. Spontaneous deamination of cytosine and 5-methylcytosine residues in DNA and replacement of 5-methylcytosine residues with cytosine residues. *Mutation Res/Rev Genetic Toxicol.* 1990;**238**:277–86.
- 61 Diamond MS, Farzan M. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat Rev Immunol.* 2013;**13**:46–57.
- 62 Diamond MS. IFIT1: A dual sensor and effector molecule that detects non-2'-O methylated viral RNA and inhibits its translation. *Cytokine Growth Factor Rev.* 2014;**25**:543–50.
- 63 Thompson SR, Gulyas KD, Sarnow P. Internal initiation in *Saccharomyces cerevisiae* mediated by an initiator tRNA/eIF2-independent internal ribosome entry site element. *Proc Natl Acad Sci.* 2001;**98**:12972–7.
- 64 Eulalio A, Behm-Ansmant I, Izaurralde E. P bodies: at the crossroads of post-transcriptional pathways. *Nat Rev Mol Cell Biol.* 2007;**8**:9–22.
- 65 Rodriguez-Corona U, Sobol M, Rodriguez-Zapata LC, Hozak P, Castano E. Fibrillarin from Archaea to human: Review on fibrillarin. *Biol Cell.* 2015;**107**:159–74.
- 66 Dimitrova DG, Teyssset L, Carré C. RNA 2'-O-Methylation (Nm) Modification in Human Diseases. *Genes.* 2019;**10**:117.
- 67 Genuth NR, Barna M. The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Mol Cell.* 2018;**71**:364–74.
- 68 Monaco P, Marcel V, Diaz J-J, Catez F. 2'-O-Methylation of Ribosomal RNA: Towards an Epitranscriptomic Control of Translation? *Biomolecules.* 2018;**8**:106.
- 69 Ferretti MB, Karbstein K. Does functional specialization of ribosomes really exist? *RNA.* 2019;**25**:521–38.
- 70 Espert L, Eldin P, Gongora C, Bayard B, Harper F, Chelbi-Alix MK, et al. The exonuclease ISG20 mainly localizes in the nucleolus and the Cajal (Coiled) bodies

- and is associated with nuclear SMN protein-containing complexes. *J Cell Biochem.* 2006;**98**:1320–33.
- 71 Kelly C, Van Driel R, Wilkinson GWG. Disruption of PML-associated nuclear bodies during human cytomegalovirus infection. *J Gen Virol.* 1995;**76**:2887–93.
- 72 Regad T, Chelbi-Alix MK. Role and fate of PML nuclear bodies in response to interferon and viral infections. *Oncogene.* 2001;**20**:7274–86.
- 73 Everett RD, Chelbi-Alix MK. PML and PML nuclear bodies: Implications in antiviral defence. *Biochimie.* 2007;**89**:819–30.
- 74 Bernardi R, Pandolfi PP. Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat Rev Mol Cell Biol.* 2007;**8**:1006–16.
- 75 Dunder M, Misteli T. Biogenesis of Nuclear Bodies. *Cold Spring Harbor Perspect Biol.* 2010;**2**:a000711.
- 76 Mao YS, Zhang B, Spector DL. Biogenesis and function of nuclear bodies. *Trends Genet.* 2011;**27**:295–306.
- 77 Courchaine EM, Lu A, Neugebauer KM. Droplet organelles? *EMBO J.* 2016;**35**:1603–12.
- 78 Trinkle-Mulcahy L, Sleeman JE. The Cajal body and the nucleolus: “In a relationship” or “It’s complicated”? *RNA Biol.* 2017;**14**:739–51. <https://doi.org/10.1080/15476286.2016.1236169>
- 79 Decker CJ, Parker R. P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harbor Perspect Biol.* 2012;**4**:a012286.
- 80 Feng J, Wickenhagen A, Turnbull ML, Rezelj VV, Kreher F, Tilston-Lunel NL, et al. Interferon-Stimulated Gene (ISG)-Expression Screening Reveals the Specific Antibunyaviral Activity of ISG20. *J Virol.* 2018;**92**:e02140–17.
- 81 Fu Y, Dominissini D, Rechavi G, He C. Gene expression regulation mediated through reversible m<sup>6</sup>A RNA methylation. *Nat Rev Genet.* 2014;**15**:293–306.
- 82 Frye M, Harada BT, Behm M, He C. RNA modifications modulate gene expression during development. *Science.* 2018;**361**:1346–9.
- 83 Yang Y, Hsu PJ, Chen Y-S, Yang Y-G. Dynamic transcriptomic m<sup>6</sup>A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Res.* 2018;**28**:616–24. <https://doi.org/10.1038/s41422-018-0040-8>
- 84 Wiener D, Schwartz S. The epitranscriptome beyond m<sup>6</sup>A. *Nat Rev Genet.* 2021;**22**:119–31.
- 85 Wang X, He C. Dynamic RNA Modifications in Posttranscriptional Regulation. *Mol Cell.* 2014;**56**:5–12.
- 86 Kan RL, Chen J, Sallam T. Crosstalk between epitranscriptomic and epigenetic mechanisms in gene regulation. *Trends Genet.* 2022;**38**:182–93. <https://doi.org/10.1016/j.tig.2021.06.014>
- 87 Harcourt EM, Kietrys AM, Kool ET. Chemical and structural effects of base modifications in messenger RNA. *Nature.* 2017;**541**:339–46.
- 88 Boo SH, Kim YK. The emerging role of RNA modifications in the regulation of mRNA stability. *Exp Mol Med.* 2020;**52**:400–8.
- 89 Netzband R, Pager CT. Epitranscriptomic marks: Emerging modulators of RNA virus gene expression. *WIREs RNA.* 2020;**11**. <https://doi.org/10.1002/wrna.1576>
- 90 Daffis S, Szretter KJ, Schriewer J, Li J, Youn S, Errett J, et al. 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature.* 2010;**468**:452–6.
- 91 Tsai K, Cullen BR. Epigenetic and epitranscriptomic regulation of viral replication. *Nat Rev Microbiol.* 2020;**18**:559–70.
- 92 Degols G, Eldin P, Mechti N. ISG20, an actor of the innate immune response. *Biochimie.* 2007;**89**:831–5.
- 93 Ding J, Aldo P, Roberts CM, Stabach P, Liu H, You Y, et al. Placenta-derived interferon-stimulated gene 20 controls ZIKA virus infection. *EMBO Rep.* 2021;**22**:2450.
- 94 Liu K, Li Y, Zhou B, Wang F, Huan B, Shao D, et al. A conjugate protein containing HIV TAT, ISG20, and a PRRSV polymerase binding inhibits PRRSV replication and may be a novel therapeutic platform. *Res Vet Sci.* 2017;**113**:13–20.
- 95 Gao M, Lin Y, Liu X, Li Y, Zhang C, Wang Z, et al. ISG20 promotes local tumor immunity and contributes to poor survival in human glioma. *OncoImmunology.* 2019;**8**:e1534038.
- 96 Qu H, Li J, Yang L, Sun L, Liu W, He H. Influenza A Virus-induced expression of ISG20 inhibits viral replication by interacting with nucleoprotein. *Virus Genes.* 2016;**52**:759–67.
- 97 Krug RM, Morgan MA, Shatkin AJ. Influenza viral mRNA contains internal N<sup>6</sup>-methyladenosine and 5'-terminal 7-methylguanosine in cap structures. *J Virol.* 1976;**20**:45–53.
- 98 Courtney DG, Kennedy EM, Dumm RE, Bogerd HP, Tsai K, Heaton NS, et al. Epitranscriptomic Enhancement of Influenza A Virus Gene Expression and Replication. *Cell Host Microbe.* 2017;**22**:377–386.e5.
- 99 Olschewski S, Cusack S, Rosenthal M. The Cap-Snatching Mechanism of Bunyaviruses. *Trends Microbiol.* 2020;**28**:293–303.
- 100 Dunn EF, Pritlove DC, Elliott RM. The S RNA genome segments of Batai, Cache Valley, Guaroa, Kairi, Lumbo, Main Drain and Northway bunyaviruses: sequence determination and analysis. *J Gen Virol.* 1994;**75**:597–608.
- 101 Wang JT, McElvain LE, Whelan SPJ. Vesicular Stomatitis Virus mRNA Capping Machinery Requires

- Specific *cis*-Acting Signals in the RNA. *J Virol.* 2007;**81**:11499–506.
- 102 Qiu W, Zhang Q, Zhang R, Lu Y, Wang X, Tian H, et al. N6-methyladenosine RNA modification suppresses antiviral innate sensing pathways via reshaping double-stranded RNA. *Nat Commun.* 2021;**12**:1582.
- 103 Hwang LN, Englund N, Pattnaik AK. Polyadenylation of Vesicular Stomatitis Virus mRNA Dictates Efficient Transcription Termination at the Intercistronic Gene Junctions. *J Virol.* 1998;**72**:1805–13.
- 104 Jiang D, Guo H, Xu C, Chang J, Gu B, Wang L, et al. Identification of three interferon-inducible cellular enzymes that inhibit the replication of hepatitis C virus. *J Virol.* 2008;**82**:1665–78.
- 105 Jia Y, Wei L, Jiang D, Cong X, Fei R. Investigating the inhibitory effects of interferon-alpha on the replication of hepatitis C virus replicon. *Zhonghua Yi Xue Za Zhi.* 2005;**85**:2065–9.
- 106 Bradrick SS. The hepatitis C virus 3'-untranslated region or a poly(A) tract promote efficient translation subsequent to the initiation phase. *Nucleic Acids Res.* 2006;**34**:1293–303.
- 107 Moes L, Wirth M. The internal initiation of translation in bovine viral diarrhoea virus RNA depends on the presence of an RNA pseudoknot upstream of the initiation codon. *Virol J.* 2007;**4**:124.
- 108 McIntyre W, Netzband R, Bonenfant G, Biegel JM, Miller C, Fuchs G, et al. Positive-sense RNA viruses reveal the complexity and dynamics of the cellular and viral epitranscriptomes during infection. *Nucleic Acids Res.* 2018;**46**:5776–91.
- 109 Kolykhalov AA, Feinstone SM, Rice CM. Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA. *J Virol.* 1996;**70**:3363–71.
- 110 Collett MS, Anderson DK, Retzel E. Comparisons of the pestivirus bovine viral diarrhoea virus with members of the flaviviridae. *J Gen Virol.* 1988;**69**:2637–43.
- 111 Zhou Y, Ray D, Zhao Y, Dong H, Ren S, Li Z, et al. Structure and Function of Flavivirus NS5 Methyltransferase. *J Virol.* 2007;**81**:3891–903.
- 112 Dong H, Ren S, Zhang B, Zhou Y, Puig-Basagoiti F, Li H, et al. West Nile virus methyltransferase catalyzes two methylations of the viral RNA cap through a substrate-repositioning mechanism. *J Virol.* 2008;**82**:4295–307.
- 113 Saeedi BJ, Geiss BJ. Regulation of flavivirus RNA synthesis and capping: Regulation of flavivirus RNA synthesis and capping. *Wires RNA.* 2013;**4**:723–35.
- 114 Song Y, Mugavero J, Stauff CB, Wimmer E. Dengue and Zika Virus 5' Untranslated Regions Harbor Internal Ribosomal Entry Site Functions. *MBio.* 2019;**10**:19.
- 115 Gokhale NS, McIntyre ABR, McFadden MJ, Roder AE, Kennedy EM, Gandara JA, et al. N6-methyladenosine in flaviviridae viral RNA genomes regulates infection. *Cell Host & Microbe.* 2016;**20**:654–65. <https://doi.org/10.1016/j.chom.2016.09.015>
- 116 Bergamini G, Preiss T, Hentze MW. Picornavirus IRESes and the poly(A) tail jointly promote cap-independent translation in a mammalian cell-free system. *RNA.* 2000;**6**:1781–90.
- 117 Jang SK, Pestova TV, Hellen CUT, Witherell GW, Wimmer E. Cap-Independent Translation of Picornavirus RNAs: Structure and Function of the Internal Ribosomal Entry Site. *Enzyme.* 1990;**44**:292–309.
- 118 Decroly E, Ferron F, Lescar J, Canard B. Conventional and unconventional mechanisms for capping viral mRNA. *Nat Rev Microbiol.* 2012;**10**:51–65.
- 119 Zhang Y, Burke CW, Ryman KD, Klimstra WB. Identification and characterization of interferon-induced proteins that inhibit alphavirus replication. *J Virol.* 2007;**81**:11246–55.
- 120 Hyde JL, Diamond MS. Innate immune restriction and antagonism of viral RNA lacking 2'-O methylation. *Virology.* 2015;**479–480**:66–74.
- 121 Hsueh C-C, Dubin DT. Di- and trimethylated congeners of 7-methylguanine in Sindbis virus mRNA. *Nature.* 1976;**264**:190–1.
- 122 Furutani Y, Toguchi M, Higuchi S, Yanaka K, Gailhouse L, Qin X-Y, et al. Establishment of a Rapid Detection System for ISG20-Dependent SARS-CoV-2 Subreplicon RNA Degradation Induced by Interferon- $\alpha$ . *IJMS.* 2021;**22**:11641.
- 123 Nakagawa K, Lokugamage KG, Makino S. *Viral and Cellular mRNA Translation in Coronavirus-Infected Cells.* *Advances in Virus Research.* Amsterdam: Elsevier; 2016. p. 165–92.
- 124 Kim D, Lee J-Y, Yang J-S, Kim JW, Kim VN, Chang H. The Architecture of SARS-CoV-2 Transcriptome. *Cell.* 2020;**181**:914–921.e10.
- 125 Liu J, Xu Y-P, Li K, Ye Q, Zhou H-Y, Sun H, et al. The m6A methylome of SARS-CoV-2 in host cells. *Cell Research.* 2021;**31**:404–14. <https://doi.org/10.1038/s41422-020-00465-7>
- 126 Zahoor MA, Xue G, Sato H, Murakami T, Takeshima S, Aida Y. HIV-1 Vpr induces interferon-stimulated genes in human monocyte-derived macrophages. *PLoS One.* 2014;**9**:e106418.
- 127 Zahoor MA, Xue G, Sato H, Aida Y. Genome-wide transcriptional profiling reveals that HIV-1 Vpr differentially regulates interferon-stimulated genes in human monocyte-derived dendritic cells. *Virus Res.* 2015;**208**:156–63.

- 128 Wilusz J. Putting an 'End' to HIV mRNAs: capping and polyadenylation as potential therapeutic targets. *AIDS Res Ther.* 2013;**10**:31.
- 129 Yedavalli VSRK, Jeang K-T. Trimethylguanosine capping selectively promotes expression of Rev-dependent HIV-1 RNAs. *Proc Natl Acad Sci.* 2010;**107**:14787–92.
- 130 Courtney DG. Post-Transcriptional Regulation of Viral RNA through Epitranscriptional Modification. *Cells.* 2021;**10**:1129.
- 131 Doria-Rose NA, Klein RM, Manion MM, O'Dell S, Phogat A, Chakrabarti B, et al. Frequency and Phenotype of Human Immunodeficiency Virus Envelope-Specific B Cells from Patients with Broadly Cross-Neutralizing Antibodies. *J Virol.* 2009;**83**:188–99.
- 132 Wieland S, Thimme R, Purcell RH, Chisari FV. Genomic analysis of the host response to hepatitis B virus infection. *Proc Natl Acad Sci.* 2004;**101**:6669–74.
- 133 Imam H, Kim G-W, Mir SA, Khan M, Siddiqui A. Interferon-stimulated gene 20 (ISG20) selectively degrades N6-methyladenosine modified Hepatitis B Virus transcripts. *PLoS Pathog.* 2020;**16**:e1008338.
- 134 Seeger C, Mason WS. Molecular biology of hepatitis B virus infection. *Virology.* 2015;**479–480**:672–86.
- 135 Block TM, Young JAT, Javanbakht H, Sofia MJ, Zhou T. Host RNA quality control as a hepatitis B antiviral target. *Antiviral Res.* 2021;**186**:104972.

## Le projet doctoral – les objectifs de l'étude

Au cours de ma thèse j'ai été encadré par le Dr. Francesca Fiorini et le Prof. Patrice Gouet au sein de l'équipe Rétrovirus et Biochimie Structurale dans l'unité MMSB (UMR 5086) à l'Institut de Biologie et Chimie des Protéines. Mon projet doctoral porte sur l'étude biochimique de l'interaction ARN–protéine lors de l'infection virale par le VIH-1 et de la réponse antivirale cellulaire. J'ai travaillé sur l'interaction entre l'ARNg viral et l'IN rétrovirale dans le contexte d'une infection par le VIH-1, et celle entre l'ARN viral et le facteur de restriction ISG20.

Dans les deux cas, il s'agit de complexes ribonucléoprotéiques qui constituent de nouvelles cibles thérapeutiques avec l'idée : 1/ d'inhiber spécifiquement l'interaction de l'IN avec l'ARNg viral ; et 2/ de favoriser l'activité d'ISG20 dans la réponse immunitaire lors de l'infection par le VIH-1.

### **PARTIE I : Etude fonctionnelle et structurale de l'interaction entre l'IN du VIH-1 et l'ARNg viral**

#### Contexte

Le travail de thèse a pour objectif d'étudier l'interaction de l'IN du VIH-1 avec l'ARN viral afin de comprendre son implication dans la morphogenèse rétrovirale (Kessl *et al.*, 2016 ; Elliott *et al.*, 2020 ; Fontana *et al.*, 2015). En effet, par des expériences de « *CLIP-Seq* », il a été montré que l'IN du VIH-1 est capable de lier l'ARNg viral sur des régions préférentielles (Kessl *et al.*, 2016). J'ai ainsi commencé l'étude de la liaison de l'IN à l'ARN TAR, pour lequel l'affinité de l'IN semblait être extrêmement importante (Kessl *et al.*, 2016). Cependant, à la fin de l'année 2020, l'implication de l'IN dans la transcription provirale du VIH-1 a été révélée par des expériences de biologie cellulaire (Winans et Goff, 2020). Cette découverte a donné un regard nouveau à mon projet de thèse car, de manière intéressante, l'ARN TAR est un élément essentiel pour la transcription provirale en complexe avec la protéine virale Tat. Par conséquent, j'ai intégré à mon étude la protéine virale Tat et les facteurs cellulaires d'élongation de la transcription (P-TEFb) pour essayer de comprendre le possible rôle de l'IN dans ce processus.

#### Objectifs spécifiques

Les objectifs spécifiques sont : (1) de trouver le domaine de l'IN capable d'interagir avec TAR et de déterminer le site de liaison sur l'ARN ; (2) de décortiquer les mécanismes moléculaires de formation du complexe entre l'IN, TAR et Tat ; (3) d'obtenir la structure tridimensionnelle du complexe formé entre l'IN et TAR par cristallographie aux rayons X ; et (4) de comprendre si le complexe formé entre l'IN et TAR est capable d'interagir avec le complexe d'élongation de la transcription.

J'ai conduit cette étude en combinant des approches de biologie moléculaire, de biochimie et de biophysique. En effet, j'ai produit et purifié différentes constructions protéiques et d'ARN *in vitro*. J'ai étudié et optimisé les conditions de liaison des complexes ribonucléoprotéiques et je les ai analysés par différentes expériences d'interaction tels que le retard sur gel, l'interférométrie de bio-couche (BLI) et la co-précipitation.

Les travaux sur cette partie nous ont permis de publier un article, pour lequel je suis co-première auteure, portant sur le rôle du CTD de l'IN dans l'interaction avec l'ARN TAR et la protéine Tat, en novembre 2022 :

**« *The HIV-1 Integrase C-Terminal Domain Induces TAR RNA Structural Changes Promoting Tat Binding* »**

Rocchi C<sup>§</sup>, Louvat C<sup>§</sup>, Miele AE, Batische J, Guillon C, Ballut L, Lener D, Negroni M, Ruff M, Gouet P, Fiorini F. *Int J Mol Sci.* 2022 Nov 8;23(22):13742. doi: 10.3390/ijms232213742.

§ : contribution égale dans ce travail

La suite de ce projet est décrite dans le chapitre Résultats à la **page 64**.

## **PARTIE II : Caractérisation biochimique de la liaison et de l'activité enzymatique du facteur de restriction viral ISG20 vis-à-vis d'un substrat ARN**

### Contexte

Pendant la pandémie de COVID-19 en 2020 nous avons commencé des réunions d'équipe partagées avec le laboratoire du Dr. Andrea Cimarelli au Centre International de Recherche en Infectiologie (CIRI) de Lyon. Nous avons donc commencé un travail collaboratif sur la protéine ISG20 : une exonucléase qui digère l'ARNsb de 3' en 5'. Cette protéine est induite par l'IFN en réponse à une infection virale comme décrit précédemment (voir **chapitre II.1**). Les mécanismes d'action et de spécificité d'ISG20 sont très peu connus. La question qui se posait était de comprendre les éléments qui rendaient certains ARN viraux et les ARN cellulaires résistants à l'action de cette enzyme (Wu *et al.*, 2019).

### Objectifs spécifiques

Les objectifs sont de comprendre si l'action d'ISG20 est inhibée par (1) des modifications épitranscriptomiques de l'ARN, (2) des structures spécifiques, (3) la présence d'une extrémité polyA en 3' des ARNm. Nous avons aussi caractérisé un mutant catalytique très utilisé dans les expériences de biologie cellulaire (Wu *et al.*, 2019).

Dans un premier temps, j'ai optimisé l'expression et la purification de cette protéine toxique pour *Escherichia coli* et ensuite j'ai mis au point des tests de digestion afin de caractériser son activité enzymatique. J'ai établi son mécanisme d'action et j'ai étudié sa liaison à l'ARNdb et l'ARNsb par des expériences de retard sur gel. Ensuite, j'ai analysé l'effet des méthylations de l'ARNsb sur la liaison et l'activité d'ISG20. Cependant au cours de ces travaux, un article plus avancé sur le sujet a été publié (El Kazzi *et al.*, 2022) et je me suis plutôt concentrée sur les objectifs 2 et 3.

Les travaux sur cette partie nous ont permis de publier une revue, portant sur les connaissances actuelles des mécanismes moléculaires de l'action antivirale d'ISG20, en février 2022 :

#### **« ISG20: an enigmatic antiviral RNase targeting multiple viruses »**

Deymier S, Louvat C, Fiorini F, Cimarelli A. FEBS Open Bio. 2022 Jun;12(6):1096-1111. doi: 10.1002/2211-5463.13382.

De plus, un article de recherche où je serais première auteure est en cours de préparation :

#### **« Stable Structures or poly(A)-binding protein loading protect cellular and viral RNA against ISG20-mediated decay »**

Louvat C, Deymier S, Nguyen X-N, Labaronne E, Noy K, Corbin A, Mateo M, Ricci EP, Fiorini F, Cimarelli A.

# **PARTIE I : Etude fonctionnelle et structurale de l'interaction entre l'IN du VIH-1 et l'ARNg viral**

## **Chapitre 2 : Résultats**

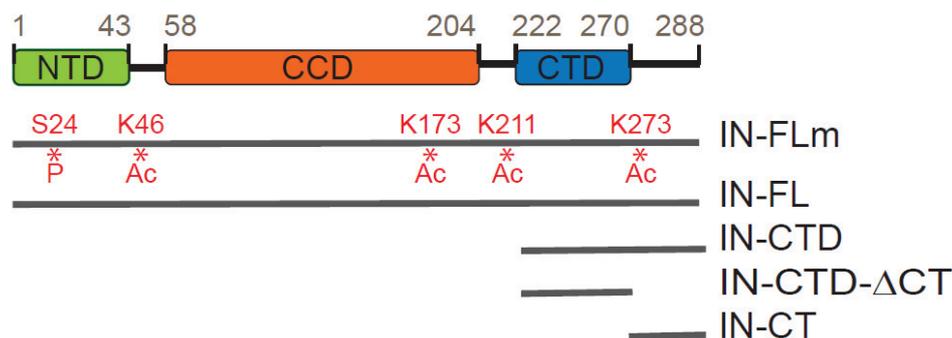
### VI. Partie fonctionnelle

#### VI.1. Etude des mécanismes d'interaction entre IN-CTD, Tat et l'ARN TAR

##### **Partie introductive de l'article de recherche « *The HIV-1 Integrase C-Terminal Domain Induces TAR RNA Structural Changes Promoting Tat Binding* » :**

Des expériences de « *cross-linking* » ont identifiés la séquence de l'ARN TAR comme un site majeur de liaison de l'IN du VIH-1 sur l'ARNg viral (Kessl *et al.*, 2016 ; Elliott *et al.*, 2020). Cette interaction joue un rôle crucial dans la morphogenèse de la particule virale et la stabilité de l'ARNg dans les cellules hôtes. Cette observation nous a encouragé à étudier l'interaction de l'IN avec l'ARN TAR en la comparant à celle entre la protéine rétrovirale Tat et l'ARN TAR déjà étudiée dans la communauté scientifique. Au début de nos travaux, nous avons utilisée l'IN entière produite dans les cellules de hamsters par notre collaborateur le Dr. Marc Ruff (IGBMC, Strasbourg) (**figure 54**). L'analyse par spectrométrie de masse a révélé des modifications post-traductionnelles de cette protéine, et a soulevé le problème de l'homogénéité de la préparation pour notre étude. De plus, nous savons que les trois domaines de l'IN lient l'ADN et nous avons pensé que cette interaction pourrait biaiser notre étude sur l'ARN. Différents travaux ont montré l'importance de résidus présents dans le CTD pour l'interaction à l'ARN, donc nous nous sommes concentrés sur ce domaine (Kessl *et al.*, 2016 ; Elliott *et al.*, 2020).

Au cours de ma thèse, d'autres découvertes sur l'implication du CTD de l'IN du VIH-1 dans la transcription provirale ont été publiés (Winans et Goff, 2020). Comme nous travaillions dans un contexte proche de celui de la transcription provirale, nous avons donc poursuivi nos travaux en étudiant plus en détail les mécanismes d'interaction du CTD de l'IN avec l'ARN TAR et la protéine Tat au cours de la transcription provirale. Pour notre étude, nous avons combinés des approches biophysiques et biochimiques afin de décortiquer le mécanisme d'interaction du CTD de l'IN du VIH-1 avec l'ARN TAR.



**Figure 54.** Schéma illustrant les domaines des protéines étudiées dans l'article de recherche de Rocchi, Louvat *et al.* (2022). IN-FLm : IN entière exprimée en cellules de mammifères ; IN-FL : IN entière exprimée dans *E. coli* ; IN-CTD s'étend des résidus 222 à 288, IN-CTD-ΔCT de 222 à 270, et IN-CT de 270 à 288, ces trois protéines sont exprimées dans *E. coli*. Les sites de phosphorylation (P) et d'acétylation (Ac) sont représentés en rouge.

## Les résultats obtenus :

Nous avons étudié l'interaction de l'IN du VIH-1 produite dans des cellules de mammifères avec l'ARN TAR structuré et un ARN linéaire par la technique de retard sur gel (**figure 1.C** de l'article). Nous avons montré que l'IN lie l'ARN TAR mais pas l'ARN linéaire. Par la suite, nous avons montré par retard sur gel qu'IN-CTD lie sans différence significative différents ARN structurés du 5'UTR de l'ARNg du VIH-1 : TAR, polyA, DIS et SD/ $\Psi$  (**figure 1.D** de l'article). Ainsi, nous avons montré que l'IN lie l'ARN TAR sans spécificité structurale apparente. Nous avons réalisé des expériences de BLI. Cette technique d'analyse optique permet de suivre en temps réel la liaison d'un interactant en solution avec un interactant immobilisé, et de déterminer les paramètres cinétiques. Nous avons déterminé l'affinité d'IN-CTD et IN-CTD- $\Delta$ CT avec l'ARN TAR, qui est de l'ordre de  $0,32 \pm 0,14 \mu\text{M}$  et  $0,77 \pm 0,20 \mu\text{M}$  respectivement (**figure 1.F** de l'article).

Par la suite, nous avons montré que le CT d'IN-CTD permet de reconnaître la forme de l'ARN TAR. En effet, lorsque nous avons étudié l'interaction par retard sur gel de mutants de la tige apicale de l'ARN TAR avec IN-CTD et IN-CTD- $\Delta$ CT : nous voyons qu'IN-CTD lie spécifiquement l'ARN TAR<sub>wt</sub> (« *wild-type* ») tandis qu'IN-CTD- $\Delta$ CT lie sans différence significative l'ARN TAR<sub>wt</sub> et ses mutants (**figure 2.B** de l'article).

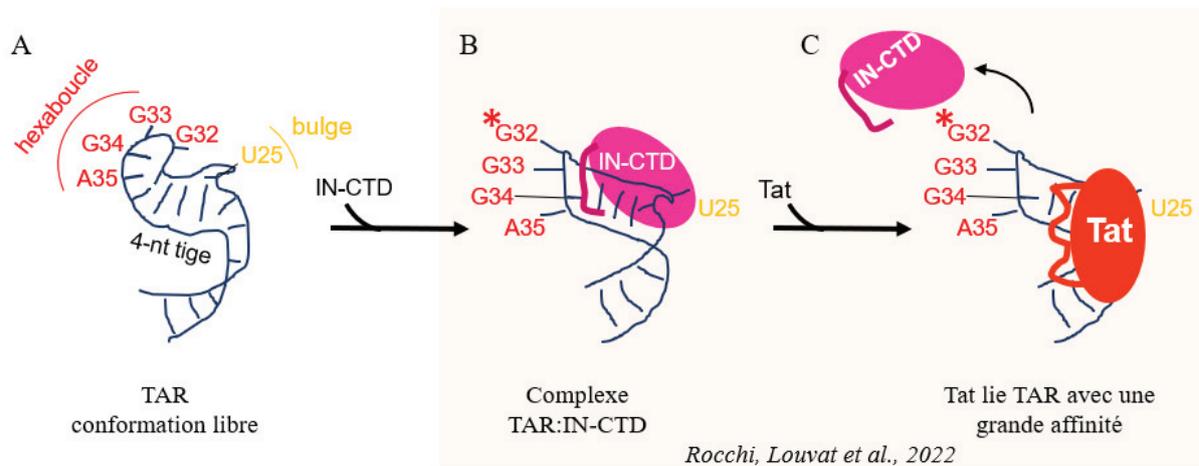
Ensuite, nous avons montré que la liaison d'IN-CTD à l'ARN TAR modifie sa structure et facilite la liaison de Tat à l'ARN TAR. L'effet de la liaison d'IN-CTD sur la structure de TAR a été mis en évidence par la technique de l'empreinte à la RNase en incubant les protéines avec l'ARN en présence de la RNase T1, qui clive l'ARNsb en 3' des G. Nous avons ainsi pu mettre en évidence que la liaison d'IN-CTD permet d'exposer les nucléotides G28 et G32 (**figure 3.A** de l'article). Tandis que les nucléotides G34 et G36 sont protégés de la digestion par la RNase T1 en présence d'IN-CTD, ce qui montre que ce domaine lie la tige-boucle apicale de l'ARN TAR et permet d'exposer certains nucléotides de TAR, modifiant ainsi sa structure.

Mon implication dans ses travaux concerne les expériences de BLI. J'ai tout d'abord lié IN-CTD et IN-CTD- $\Delta$ CT à TAR, puis j'ai incubé le biocapteur dans une solution contenant Tat et j'ai mesuré la variation du signal dans le temps (**figures 3.B et 3.D** de l'article). J'ai répété cette opération en faisant varier la concentration de Tat dans la solution de 0,2 à 2  $\mu\text{M}$ . J'ai observé que l'affinité de Tat pour l'ARN TAR est améliorée d'un facteur 30 à 100 en présence d'IN-CTD et IN-CTD- $\Delta$ CT respectivement (**figure 3.C** de l'article). Cette amélioration apparente de l'affinité est spécifique à l'ARN TAR puisqu'en répétant ces expériences de BLI avec un ARN contrôle faiblement structuré, l'affinité de Tat pour l'ARN n'est pas modifiée en présence d'IN-CTD- $\Delta$ CT ou IN-CTD (**figures 3.D et 3.E** de l'article). De plus, en inversant l'ordre d'incubation : l'ARN TAR est d'abord incubé avec Tat puis avec IN-CTD- $\Delta$ CT ou IN-CTD (**figure supplémentaire 5.A** de l'article), j'ai montré qu'IN-CTD- $\Delta$ CT ou IN-CTD ne peut pas lier l'ARN TAR lorsqu'il est lié à Tat. Ainsi l'interaction est séquentielle : IN-CTD- $\Delta$ CT ou IN-CTD lie l'ARN TAR ce qui facilite ensuite la liaison de Tat, possiblement grâce aux changements structuraux de l'ARN induit par IN-CTD (**figure supplémentaire 5.B** de l'article). D'autre part, le CT d'IN-CTD ne semble pas être impliqué dans ce mécanisme.

Nous avons observé grâce à une expérience de retard sur gel que Tat interfère avec IN-CTD pour la liaison à l'ARN TAR (**figure 4.A** de l'article). J'ai donc fait des expériences de précipitation de l'ARN biotinyllé avec des billes streptavidine incubées en présence d'IN-CTD- $\Delta$ CT ou IN-CTD et de concentration croissante de Tat (**figure 4.B** de l'article). J'ai observé que Tat entraîne la disparition dans la fraction éluée d'IN-CTD- $\Delta$ CT ou IN-CTD de manière dose-dépendante, ce qui signifie une dissociation d'IN-CTD- $\Delta$ CT ou IN-CTD du complexe Tat:TAR.

L'ensemble de nos résultats sont représentés dans la **figure 55** ci-dessous : le CT de l'IN du VIH-1 permet de reconnaître l'ARN TAR<sub>wt</sub>, puis IN-CTD lie l'ARN TAR et le modifie structurellement en

exposant la G32, ce qui améliore la liaison à la protéine Tat. À une certaine concentration de Tat, IN-CTD est dissocié du complexe.



**Figure 55.** Schéma bilan des résultats présentés dans l'article de recherche Rocchi, Louvat et al. (2022).

La suite est connue : Tat recrute le complexe P-TEFb composée de la cycline T1 et de la CDK9 pour déclencher des cascades de phosphorylation et activer l'élongation de la transcription (Muniz *et al.*, 2010).



Article

# The HIV-1 Integrase C-Terminal Domain Induces TAR RNA Structural Changes Promoting Tat Binding

Cecilia Rocchi <sup>1,†</sup>, Camille Louvat <sup>1,†</sup>, Adriana Erica Miele <sup>2,3</sup>, Julien Batisse <sup>4</sup>, Christophe Guillon <sup>1</sup>, Lionel Ballut <sup>1</sup>, Daniela Lener <sup>5</sup>, Matteo Negroni <sup>5</sup>, Marc Ruff <sup>4</sup>, Patrice Gouet <sup>1</sup> and Francesca Fiorini <sup>1,\*</sup>

<sup>1</sup> Molecular Microbiology and Structural Biochemistry, MMSB-IBCP, UMR 5086, CNRS, University of Lyon, 7 Passage du Vercors, CEDEX 07, 69367 Lyon, France

<sup>2</sup> Institute of Analytical Sciences, UMR 5280 CNRS UCBL University of Lyon, 5 Rue de la Doua, 69100 Villeurbanne, France

<sup>3</sup> Department of Biochemical Sciences, Sapienza University of Rome, P.le Aldo Moro 5, 00185 Rome, Italy

<sup>4</sup> Chromatin Stability and DNA Mobility, Department of Integrated Structural Biology, IGBMC, CNRS, UMR 7104—Inserm U 158, University of Strasbourg, 1 rue Laurent Fries, 67404 Illkirch, France

<sup>5</sup> RNA Architecture and Reactivity, IBMC, CNRS, UPR 9002, University of Strasbourg, 2, Allée Konrad Roentgen, 67084 Strasbourg, France

\* Correspondence: francesca.fiorini@ibcp.fr; Tel.: +33-4-72722624; Fax: +33-4-72722616

† These authors contributed equally to this work.

**Abstract:** Recent evidence indicates that the HIV-1 Integrase (IN) binds the viral genomic RNA (gRNA), playing a critical role in the morphogenesis of the viral particle and in the stability of the gRNA once in the host cell. By combining biophysical, molecular biology, and biochemical approaches, we found that the 18-residues flexible C-terminal tail of IN acts as a sensor of the peculiar apical structure of the trans-activation response element RNA (TAR), interacting with its hexaloop. We show that the binding of the whole IN C-terminal domain modifies TAR structure, exposing critical nucleotides. These modifications favour the subsequent binding of the HIV transcriptional trans-activator Tat to TAR, finally displacing IN from TAR. Based on these results, we propose that IN assists the binding of Tat to TAR RNA. This working model provides a mechanistic sketch accounting for the emerging role of IN in the early stages of proviral transcription and could help in the design of anti-HIV-1 therapeutics against this new target of the viral infectious cycle.

**Keywords:** RNA-protein interaction; HIV-1 Integrase; TAR RNA; Tat; proviral transcription; C-terminal tail



**Citation:** Rocchi, C.; Louvat, C.; Miele, A.E.; Batisse, J.; Guillon, C.; Ballut, L.; Lener, D.; Negroni, M.; Ruff, M.; Gouet, P.; et al. The HIV-1 Integrase C-Terminal Domain Induces TAR RNA Structural Changes Promoting Tat Binding. *Int. J. Mol. Sci.* **2022**, *23*, 13742. <https://doi.org/10.3390/ijms232213742>

Academic Editor: Elena A Stolboushkina

Received: 11 October 2022

Accepted: 3 November 2022

Published: 8 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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

Protein–nucleic acid interactions can occur through different types of protein binding domains and are responsible for a variety of essential molecular and cellular mechanisms and their regulation. This binding diversity has been well described by recent RNA interactome screenings revealing that the term RNA-Binding Domain (RBD) is no longer synonymous of a well-structured domain, but it also includes intrinsically disordered regions (IDR) endowed with non-canonical RNA-binding properties [1–3]. Moreover, an increasing level of complexity has been reported for certain transcription factors, which are able to bind both DNA and RNA through separate structured or unstructured regions. This results in a complex pattern of specific and non-specific interactions [4,5]. This protein moonlighting is particularly true for RNA viruses that, being limited in the size of their genome, frequently ensure multiple functions with a single protein [6,7].

A good example of moonlighting is represented by Human Immunodeficiency Virus type 1 (HIV-1) integrase (IN) that binds both DNA and RNA [8,9]. As with all retroviruses, IN is responsible for the integration of viral DNA (vDNA), which is produced by retro-transcription of genomic RNA (gRNA) into the host genome, generating a provirus. Multimeric IN binds the vDNA ends forming the intasome complex able to catalyze

the processing of the 3'-end di- or trinucleotides. After activation, the 3' extremities of vDNA are then used by the intasome to attack the host DNA in order to integrate it and generate the provirus (reviewed in [9]). This step is essential for HIV-1 infection [10]. Retroviral integrases are modular proteins that contain three structured domains: the N-terminal domain (NTD), the catalytic core domain (CCD), and the C-terminal domain (CTD), connected by linkers, either unstructured or helical [11]. All three domains present protein–protein and protein–DNA interaction properties and are essential for enzymatic activity (reviewed in [9,12]). The CCD harbors the essential catalytic triad Asp, Asp, Glu (D,D,E) that coordinates two  $Mg^{2+}$  cofactors and folds similarly to nucleotidyl-transferases and nucleases [13]. The NTD is involved in enzyme multimerization and catalytic activity and shows a His<sub>2</sub>Cys<sub>2</sub> motif coordinating one  $Zn^{2+}$  ion [14–16]. The CTD (reviewed in [11]) is also involved in DNA interaction and multimerization, and it possesses an SH3-like fold followed by a flexible 18-residues tail (CT) [17–20]. The SH3-like domain of IN-CTD shows non-specific DNA-binding properties and contains many basic residues [18,20–22]. In the HIV-1 life cycle, this hub domain mediates the interaction with the reverse transcriptase (RT) with the cellular nuclear import complex TRN-SR2 and with the histone 4 tail, likely anchoring the intasome to the chromatin, thereby promoting efficient integration [21,23–25]. The mutational study of flexible CT revealed a moderate implication in IN enzymatic activity [22,23,26] but significant effect on the incorporation of IN into virions and on HIV-1 infectivity [23,26]. Nevertheless, our understanding of the multiple functions exerted by this domain during infection is still far from complete.

As mentioned before, recent works revealed that in HIV-1, IN is also an RBP with an essential role in virion morphogenesis related to its ability to bind gRNA [8,27,28]. In fact, IN interacts with specific sites within the gRNA, ensuring the correct localization of the viral ribonucleoprotein complex (vRNP) inside the capsid (reviewed in [27]). Aberrant virions are obtained when the IN–gRNA interaction is abolished, highlighting the importance of the proper formation of IN-containing RNPs for HIV-1 infection [8,26,28,29]. In addition, the IN–gRNA interaction also dictates the fate of the gRNA within the host cell in the early steps of infection. Indeed, in virions defective for these interactions, the gRNA is located outside the capsid and it is rapidly degraded in the host cell, blocking the infectious cycle at early stages of reverse transcription [28,29]. Even if the RBD within IN has not yet been structurally identified, it has been shown that most of the Lysine residues interacting with the RNA are located within the CTD and overlap with those subjected to post-translational modification [8,30–32]. Among these, Lysine 273, belonging to the flexible CT, seems to be the only Lysine specifically involved in RNA interaction, which appears essential for viral infectivity [8,33]. Several compensatory IN mutant viruses have been recently isolated and highlighted the essential role of CTD in RNA binding [34].

Yet another role has been recently proposed for HIV-1 IN during proviral transcription at early times after integration. In fact, after strand transfer occurring during the integration process, IN remains bound to DNA and plays a role in proviral transcription, depending on post-translational modifications of specific residues within the CTD [35].

The HIV-1 provirus is transcribed by the cellular RNA polymerase II (Pol II), which pauses shortly after the initiation of transcription due to the presence of negative elongation factors as well as nucleosomes downstream from the transcription start site [36,37]. HIV-1 removes this block by encoding the transcriptional trans-activator Tat protein, which binds the nascent transcript on a structured RNA sequence, named the TAR (trans-activation response) element, using a non-canonical RBD. This allows the recruitment of the human super elongation complex (hSEC) [38–41]. In particular, Tat binds to pTEFb, a complex composed of CDK9 kinase and its regulatory partner cyclin T1 (CycT1), and consequently drives hSEC to TAR RNA. This complex triggers a cascade of phosphorylation of several transcription factors, which activate Pol II and recruit positive chromatin remodellers. Moreover, processivity of Pol II is also enhanced by pTEFb-mediated phosphorylation of its C-terminal domain [42] and is reviewed in [43]. To the best of our knowledge, molecular

details about the interplay between IN and this cellular transcription initiation machinery are still largely unknown.

Crosslinking-immuno-precipitation sequencing (CLIP-seq) experiments have identified the TAR RNA sequence as a major binding site of HIV-1 IN on the gRNA [8]. This observation, together with the recent finding of HIV-1 IN involvement in proviral transcription [35], prompted us to study the interaction of IN with TAR RNA and its interplay with the Tat protein. Our results revealed that, despite the apparent lack of structural specificity of IN *in vitro*, the CT flexible tail discriminates for a proper TAR apical stem-loop. We describe the consequences of IN binding on the structure of TAR and on the subsequent Tat/TAR interaction and propose a working model, which foresees a possible involvement of IN in proviral transcription elongation before the arrival of Tat.

## 2. Results

### 2.1. IN Binds TAR RNA with No Apparent Structural Specificity

We first addressed whether full-length HIV-1 IN (IN-FL) was able to specifically bind TAR RNA. One hindering aspect of this study is the well-known low solubility of HIV-1 IN as well as its flexibility between N-terminal (NTD), C-terminal (CTD), and catalytic core (CCD) domains that for a long time frustrated structural studies (Figure 1A [44]). The poor solubility *in vitro* is usually overcome with mutations of the hydrophobic residues; however, this results in a replication-defective virus with mislocalized viral RNP phenotype, analogous to that observed in IN mutants defective for the IN-gRNA interactions [8,19,28,29]. We have chosen to express recombinant wild-type N-Terminal Flag-tagged IN-FL in a mammalian expression system as previously described ([45] Supplementary Figure S1A, left panel) and called it IN-FLm (1-288 amino acids [aa]). Mass spectrometry analysis of this protein revealed post-translational modification at several residues; Serine 24 was phosphorylated, and Lysine 46, 173, 211, and 273 residues were acetylated. The protein produced in mammalian cells was shown to have an increased solubility compared to that produced in *Escherichia coli* and also an enhanced enzymatic activity *in vitro* [45]. In this study we used an electrophoretic mobility shift assay (EMSA) in order to qualitatively assess the binding of IN to a synthetic TAR RNA, whose secondary structure is depicted in Figure 1B, to a weakly-structured RNA(30)-mer (Supplementary Figure S1B, left panel) and to an unstructured AG(50)-mer RNA (Figure 1C). We have incubated radiolabeled RNA with increasing concentrations of purified IN-FLm. A band shift was observed with TAR RNA upon gel electrophoresis under a nearly physiological salt concentration, reflecting the formation of an IN:TAR complex (Figure 1C). IN-FLm also bound weakly-structured RNA(30)-mer (Supplementary Figure S1B, right panel) while it had a markedly reduced affinity for unstructured AG(50)-mer RNA (Figure 1C). IN-FL produced in the prokaryotic expression system (Supplementary Figure S1A, right panel) is also able to bind TAR (Supplementary Figure S1C). Therefore, the post-translational modifications are not key to binding.

As RNA-interacting Lysine residues are contained within the C-terminal domain of IN [8], and given the premises stated in the introduction, we focused on this domain. We expressed and purified the whole C-Terminal Domain (IN-CTD; aa 220–288) and determined whether IN-CTD was able to interact with structurally distinct viral genomic RNA elements of similar nucleotide length derived from HIV-1 5' UTR. In addition to TAR, we synthesized three other RNA hairpins: the polyadenylation (polyA) signal, the dimerization initiation sequence (DIS), and the major splice-donor (SD) together with a  $\Psi$  packaging element (Psi) (Figure 1D, top panel). Although different  $K_D$  values were measured using IN-FL [8], IN-CTD bound all those RNAs with no measurable difference in our conditions (Figure 1D, bottom panel), indirectly proving that IN can bind structured RNA through its C-terminal domain.

Subsequently, we wanted to address whether the C-terminal flexible tail (CT) spanning the last 18 residues of CTD affected the RNA binding properties of IN. We expressed the CTD without its terminal tail (IN-CTD- $\Delta$ CT; aa 220–270, Figure 1A and Supplementary Figure S1A) and used a chemically synthesized IN-CT peptide (aa 270–288; Figure 1A)

as a control. The boundaries between CTD and CT were defined according to sequence alignment and previous structural studies [17,33,46,47]. The IN-CT is not conserved among lentiviruses (Supplementary Figure S2A); however, multiple alignments of IN-CT from human HIV-1 subtypes and simian viruses (Figure 1E) showed 58% sequence identity and 87% similarity. EMSA assays did not show apparent differences between the affinities of IN-CTD and IN-CTD-ΔCT for TAR under physiological salt conditions (Supplementary Figure S2B). To measure the rate constants of the IN-CTD and IN-CTD-ΔCT interactions with TAR, we used Bio-layer interferometry (BLI). In these experiments, a 3' biotinylated TAR RNA was immobilized on a streptavidin-coated biosensor to serve as a bait molecule (Figure 1F, top panel). Subsequently, the binding of IN subdomains to TAR was monitored in real time by soaking the probe in solutions containing different concentrations of protein. The resulting sensorgrams are shown in Supplementary Figure S2C and represent the binding and dissociation of the protein on the RNA-coated probe. The data analysis resulted in equilibrium dissociation constant ( $K_D$ ) values of 770 and 320 nM for IN-CTD-ΔCT and IN-CTD, respectively (Figure 1F, bottom panel). Thus, the presence of the 18 aa C-terminal tail only slightly affects the RNA binding affinity of IN-CTD.

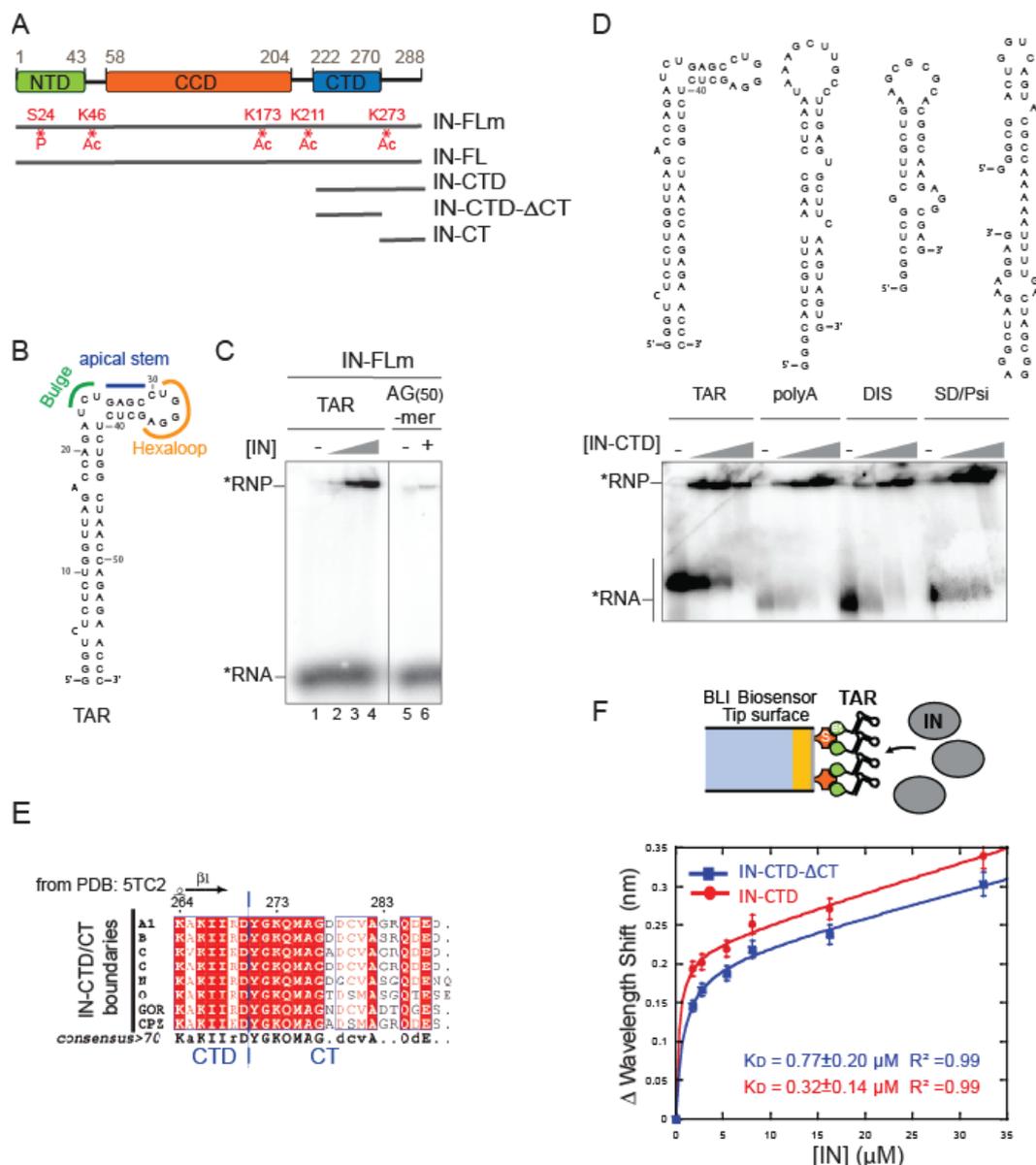


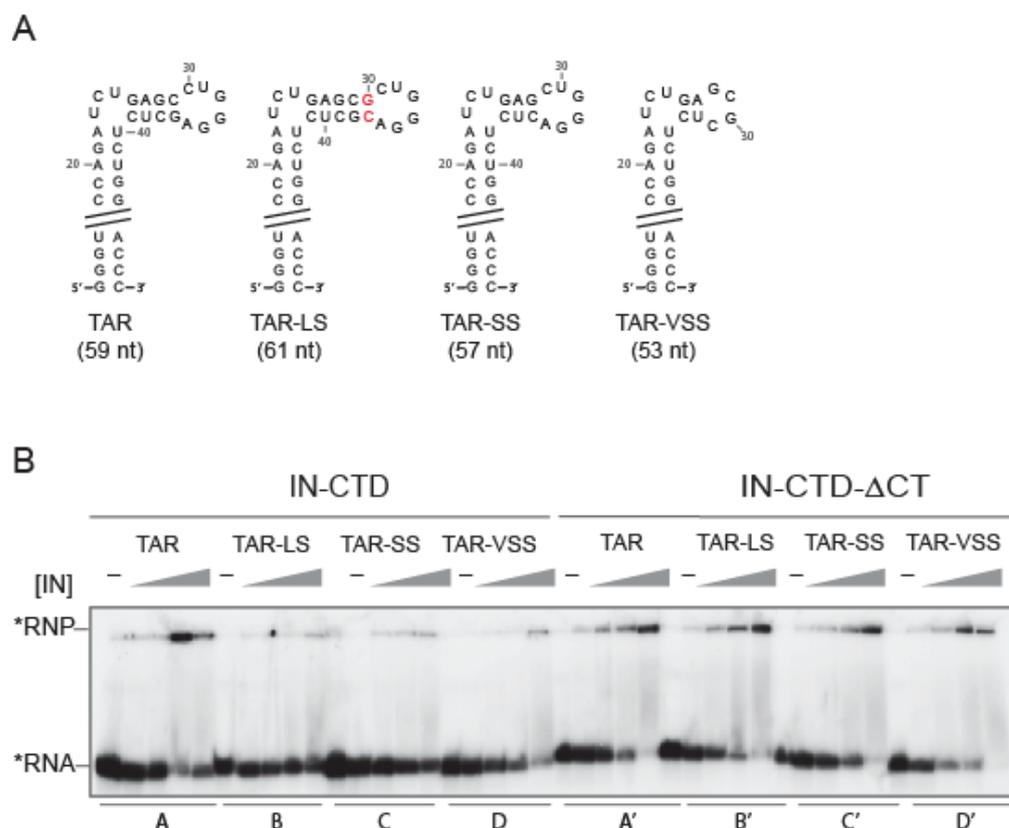
Figure 1. IN binds to structured RNA. (A) Schematic diagram showing the domain organization of

IN: the N-terminal domain (NTD), the catalytic core domain (CCD) and the C-terminal domain (CTD) are indicated respectively in green, orange and blue rectangles. Recombinant protein versions used in this study are represented by grey lines: IN full-length from mammalian expression system (IN-FLm); IN full-length from *E. coli* (IN-FL); IN-CTD (residues 222 to 288); IN-CTD- $\Delta$ CT (222 to 270); IN-CT (270 to 288). Sites of phosphorylation (P) and acetylation (Ac) are shown in red. (B) Structural model of TAR RNA obtained with RNA Fold WebServer [48,49] predicting a minimum free energy of  $-29.60$  kcal/mol. (C) Representative native 6% polyacrylamide gel illustrating the Electrophoretic mobility shift assay (EMSA) showing the interaction of IN-FLm with TAR RNA or an unstructured RNA 50-mer (AG(50)-mer) labelled with  $^{32}$ P (black star). The RNA substrates (50 nM) were incubated with various concentrations of IN-FLm or without protein (TAR RNA: 0; 100, 200, 400 nM of IN-FLm; AG(50)-mer RNA: 0 or 400 nM of IN-FLm). (D) Models of secondary structures of four RNA elements belonging to 5' untranslated region of the HIV-1 RNA genome: TAR, Poly-A, DIS and SD/Psi (upper panel). EMSA assay showing the binding of IN-CTD to these RNA elements (lower panel). Increasing amounts of IN-CTD (0; 100; 200; 400 nM) were incubated with 5'-end radiolabeled RNA (50 nM). (E) Sequence alignment of the C-terminal extremity of IN-CTD from HIV-1 subtypes and simian viruses. All amino acid sequences were obtained from the HIV database compendium (<http://www.hiv.lanl.gov/>, accessed on 28 October 2020) and aligned using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>, accessed on 28 October 2020) in order to have a consensus sequence for each viral subtype. Consensus sequences from subtypes A1, B, C, G, N, O and from GOR and CPZ, where aligned and analyzed by ESPript 3.0 Web server [50]. Secondary structure elements from IN-CTD- $\Delta$ CT structure (PDB code: 5TC2) are presented on top of the alignment (strands with arrows). Red shading indicates sequence identity and boxes indicate sequence similarity, according to physical-chemical properties. (F) top: Schematic representation of the Bio-Layer interferometry (BLI) experiment showing the binding of IN (grey ellipses) to 5'-biotinylated TAR RNA immobilized on streptavidin-coated biosensor; bottom: graph showing the wavelength shifts recorded at 200 s after the start of the protein/RNA binding were plotted against the corresponding IN-CTD- $\Delta$ CT (blue line, squares) and IN-CTD (red line, circles) concentrations, in order to calculate the respective equilibrium dissociation constant ( $K_D$ ) values. Data points were fitted to the equation:  $y = B_{max} \times x / (K_D + x) + NS \times x$  where  $B_{max}$  is the maximum wavelength shift and  $NS$  the slope of the non-linear component as described in [51]. The coefficients of determination ( $R^2$ ) and equilibrium dissociation constant ( $K_D$ ) values obtained are indicated in the graph for each IN fragment. Binding assays were performed in duplicate. Error bars indicate the Standard Error of the Mean.

## 2.2. The C-Terminal Tail Senses the TAR RNA Shape

It has been shown that the presence of the bulge and the loop in TAR RNA, rather than its sequence, is critical for IN binding, as the deletion of one or both markedly decreased IN binding affinity [8]. We wondered whether the shape of the peculiar apical stem-loop of TAR (Figure 2A) could also be important. Therefore, we mutated the 4 nucleotides (nt) stem between the bulge and the loop to change its length, as shortening or lengthening the 4 nt-stem is likely to tighten or loosen the TAR major groove by bringing the bulge and the hexa-loop closer or further, respectively. We produced one TAR with 1 bp longer stem (TAR-LS), two with progressively shorter stem TARs (TAR-SS and TAR-VSS) (Figure 2A), and one without a stem (TAR- $\Delta$ S) (Supplementary Figure S3A). The EMSA assay showed that IN-FLm was able to bind TAR-LS similarly to the wild-type TAR (Supplementary Figure S3A, lanes A and B). On the contrary, shortening of the stem reduced IN-FLm affinity for TAR (Supplementary Figure S3A, lanes C to E). Furthermore, we assessed the binding behavior of IN-CTD and IN-CTD- $\Delta$ CT to TAR mutants with EMSA (Figure 2B). As observed for the full-length protein, the affinity of IN-CTD for TAR mutants decreased as the length of the stem was reduced (Figure 2B, lanes A to D). Surprisingly, the binding ability of IN-CTD- $\Delta$ CT was not affected by the shortening of the TAR 4nt-stem (Figure 2B,

lanes A' to D'), suggesting that IN C-terminal tail senses the shape of the TAR RNA stem in vitro.



**Figure 2.** IN C-terminal tail sensing for TAR RNA apical stem-loop shape. (A) Structural model of TAR RNA and mutated versions used in our study: TAR long stem (TAR-LS, the red font underlines the added bases), TAR short stem (TAR-SS) and TAR very short stem (TAR-VSS). In order to verify that each RNA mutant assumed the expected secondary structure each model was analyzed with Fold WebServer [48,49]. (B) EMSA assay illustrating the interaction of IN-CTD and IN-CTD- $\Delta$ CT with TAR wild type and mutants. The RNA substrates (50 nM) are labelled with  $^{32}$ P (black star) and incubated with increasing concentrations of proteins (0; 100, 200, 400 and 800 nM) under the conditions described in Section 4.

As a control, we studied the interaction of Tat with TAR, the cognate interacting protein. We used a chemically synthesized full-length Tat protein (1–102 aa) to assess the binding of the TAR mutants with an EMSA assay. Similarly to IN-CTD and consistently with other in vitro and ex vivo information [52–55], the binding ability of Tat decreased with stem shortening (Supplementary Figure S3B).

Overall, our results suggest that IN-CTD interacts with the apical stem-loop of TAR, possibly with its major groove. This is consistent with the CLIP-seq data, which had identified the TAR hexaloop as a major binding site for IN, and with a recently published structural model of IN-CTD- $\Delta$ CT bound with TAR [8,56]. This is also consistent with cross-linking SHAPE experiments published while this manuscript was under submission [57].

### 2.3. IN-CTD Deeply Affects the Structure of TAR Favoring Tat Interaction

In order to extend the binding analysis at the molecular level, the IN-TAR interaction was probed by footprinting techniques. The 5' radiolabelled TAR, alone or complexed to a protein, was subjected to partial digestion using RNase T1 that cleaves the 3' to an unpaired guanine. As was consistent with previous results [58], under our conditions Tat-bound TAR RNA formed a large complex and prevented the RNA digestion by the nuclease (Figure 3A, lane 7). Concerning the interaction with IN fragments, the results shown in

Figure 3A indicated that G34 and, to a lesser extent, G36 of TAR were protected in the presence of IN-CTD (Figure 3A, lane 6). However, the binding of IN-CT and IN-CTD- $\Delta$ CT did not protect these nucleotides from nuclease digestion (Figure 3A, lanes 4 and 5). This suggested that the CT region is interacting with G34 and possibly with G36 located in the TAR hexaloop and its junction with the 4-nt stem, but only when CT is part of the whole domain (Figure 3A, lanes 4 to 6). To our surprise, we observed a prominent cleavage following the residues at positions C41, U38, G32, and, to a lesser extent, G28, which was induced by the binding of IN-CTD- $\Delta$ CT and IN-CTD even in the absence of a T1 nuclease (Supplementary Figure S4A). This could reflect the presence of drastic structural constraints on the apical stem-loop of TAR upon the binding of either IN-CTD- $\Delta$ CT or IN-CTD, which leads to a spontaneous mechanical breakage/hydrolysis in the frame of our small TAR RNA construct. No nuclease contamination has been observed in the protein solution (Supplementary Figure S4B), and the data was reproducible.

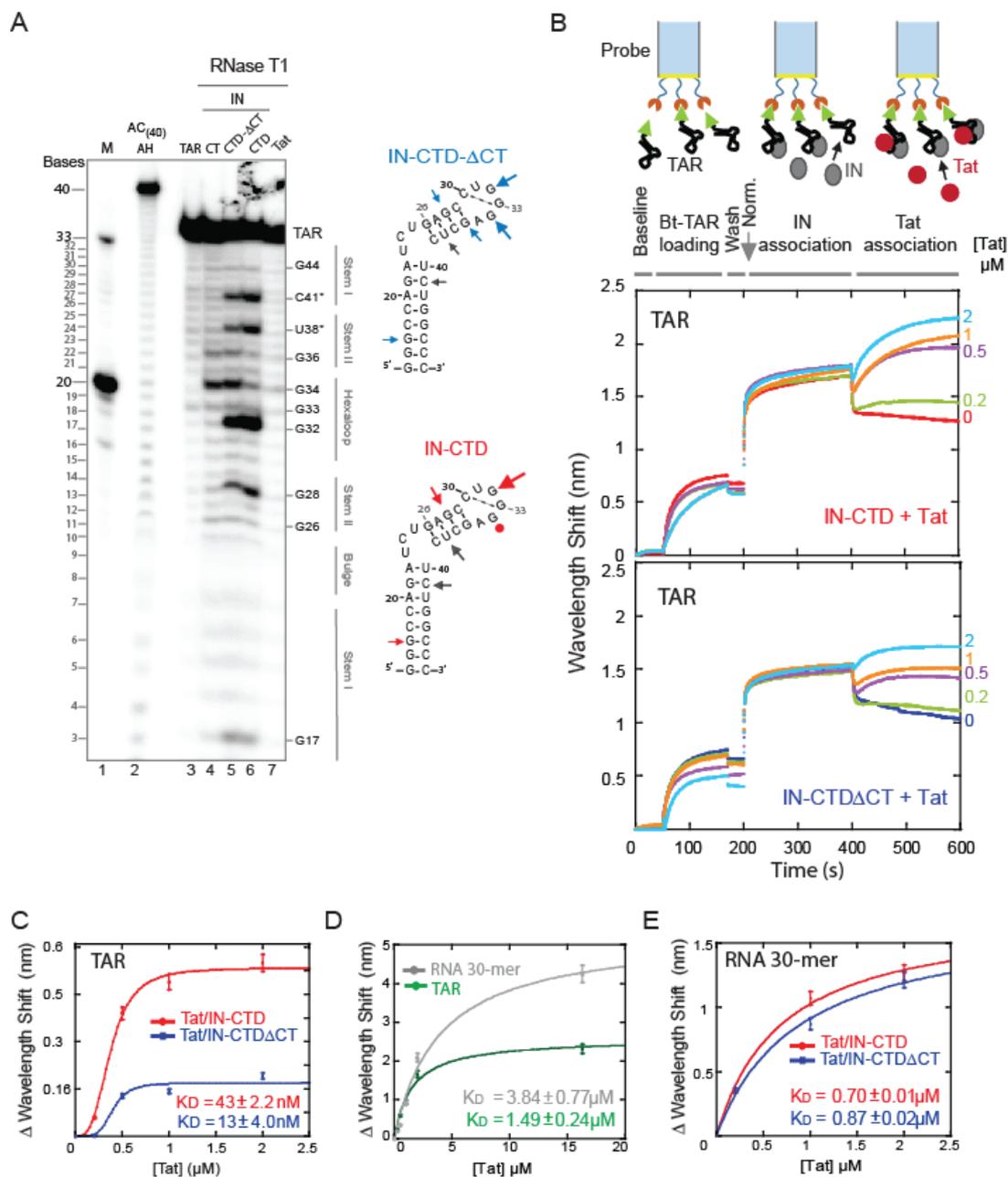


Figure 3. IN induces structural modifications of TAR, promoting Tat binding. (A) Probing the change

in the secondary structure of TAR RNA complexed with different fragments of IN and Tat. 5'-end radiolabeled TAR (34-mer) was incubated in the presence or absence of protein for 30 min at 37 °C prior to RNase T1 treatment, as described in experimental procedures. RNA fragments were separated on a 15% denaturing polyacrylamide sequencing gel. Bands corresponding to certain T1 cleavage (at G bases) products are identified as position markers. Probing gel lanes are as follows: (M) Ladder of two RNA transcripts of 33 and 20 nucleotides in length (lane 1); (AC(40) AH) alkaline ladder of AC(40)-mer RNA (lane 2); TAR native (lane 3); TAR RNA complexes with IN-CT (lane 4); IN-CTD- $\Delta$ CT (lane 5); IN-CTD (lane 6) and Tat (lane 7). Digestion patterns were mapped on TAR secondary structure depicted on the right of the gel: circles identify nucleotides protected from RNase T1 digestion and arrows mark the RNA cleavage sites. Grey arrows indicated nonspecific hydrolysis. The dimensions of arrows are proportional to the intensity of the band. (B) Real-time measurements of protein-RNA interaction by Interferometry (BLItz<sup>®</sup> System instrument, FortéBio). The 3' biotinylated TAR (Bt-TAR) was first loaded on the streptavidin-coated biosensor for 120 s (Bt-association) then the unbound RNA was washed for 30 s (wash). The sensor was absorbed in a solution containing about 90  $\mu$ M of IN protein for 200 s then incubated with different concentrations of Tat (0.2, 0.5, 1 and 2  $\mu$ M) for 200 s. (C) Analysis of interferometry shifts as a function of Tat concentration (from data shown in Panel B). Data points were fitted with Hill equation:  $y = B_{max} \cdot x^n / (K_D + x^n)$ ;  $K_D = 43 \pm 2.2$  nM ( $R^2 = 0.99$ ),  $n = 3$  and  $B_{max} = 0.57 \pm 0.02$  nm for Tat/IN-CTD (red curve) and  $K_D = 13 \pm 4.0$  nM ( $R^2 = 0.96$ )  $n = 5$  and  $B_{max} = 0.18 \pm 0.02$  nm for Tat/IN-CTD- $\Delta$ CT (blue curve). (D) BLI experiment to measure Tat binding to immobilized Bt-TAR RNA (green line) or immobilized Bt-RNA(30)-mer (grey line). Data points were fitted with Michaelis-Menten equation. Tat affinities are:  $K_D = 1.49 \pm 0.24$   $\mu$ M ( $R^2 = 0.99$ ) for TAR;  $K_D = 3.84 \pm 0.77$   $\mu$ M ( $R^2 = 0.99$ ) for RNA(30)-mer. (E) BLI experiments as in B and C, but with immobilised RNA(30)-mer to test Tat ability to displace IN-CTD or IN-CTD- $\Delta$ CT (see methods section for protocol). Data points were fitted with Michaelis-Menten equation.  $K_D = 0.7 \pm 0.1$   $\mu$ M ( $R^2 = 0.99$ ) for Tat/IN-CTD (red) and  $K_D = 0.87 \pm 0.2$   $\mu$ M ( $R^2 = 0.99$ ) for Tat/IN-CTD- $\Delta$ CT (blue).

Altogether, the digestion patterns showing the protection of G34 and G36 by IN-CTD, but not IN-CTD- $\Delta$ CT, as well as the evidence of the structural deformation of TAR induced by the binding of both fragments, indicate that (i) IN-CTD binds to the TAR apical stem-loop, (ii) IN-CTD modifies the structure of the RNA, (iii) and the CT exerts a specific role in the interaction of IN-CTD with the TAR hexaloop and its junction with the stem.

The fact that CT binds TAR only in the IN-CTD context could reflect the necessity for a distortion of TAR by the IN-CTD- $\Delta$ CT moiety to allow for the correct binding of CT to the hexaloop. Interestingly, in a recent work, Liu and co-workers [57] studied the binding of IN-FL to a whole HIV-1 5'UTR with crosslinking followed by SHAPE analysis. They found the crosslinking site for IN-FL near our breakage site (C41, U38) at position C39. Moreover, they also found an increased SHAPE reactivity of G32 indicating its greater exposure upon IN-FL binding [57] as we could show for IN-CTD.

In order to understand whether the structural changes of TAR induced by IN binding were affecting the affinity of Tat to TAR, we performed real-time binding experiments with Bio-Layer Interferometry (BLI). We measured the interactions as follows (Figure 3B, upper scheme). First, we immobilized biotinylated TAR RNA to a streptavidin-coated biosensor and washed out the unbound RNA (Figure 3B, curve between 50 and 170 s and between 170 s and 200 s); then, we soaked the probe in solutions containing a large excess (more than 10 times the  $K_D$ ) of IN-CTD or IN-CTD- $\Delta$ CT (Figure 3B, ascendant curve between 200 and 400 s indicating positive interference and hence association). Afterwards, the IN:TAR bound-biosensor was plunged in a solution containing Tat. Different concentrations of Tat were analyzed each time on different probes prepared as described in the methods. According to the experimental results in Figure 3B, a supplementary association of Tat on the IN:TAR complex occurred as indicated by the second ascendant curve between 400 and 600 s (Figure 3B). Interestingly, this association was dose-dependent as demonstrated by using increasing concentrations of Tat, between 0.2 and 2  $\mu$ M (Figure 3B).

In order to quantify this observation, we calculated the apparent  $K_D$  of Tat upon the binding of IN (Figure 3C) and found that it was about 34.7 and 114.6 times lower in presence of IN-CTD- $\Delta$ CT and IN-CTD, respectively (Figure 3C), compared to the  $K_D$  of Tat measured in the same conditions for the naked RNA (Figure 3D and Supplementary Figure S4C). Moreover, the binding kinetics of Tat on the IN:TAR complex is consistent with a cooperative interaction (Figure 3C). Notably, the IN-CTD:TAR complex can accommodate more Tat than IN-CTD- $\Delta$ CT:TAR, as indicated by the higher BMax (maximum wavelength shift, Figure 3C). We performed the same BLI experiments coating the biosensor with the biotinylated generic RNA(30)-mer used in the EMSA assay (Supplementary Figure S1B) instead of TAR RNA, and we repeated the same protocol (Supplementary Figure S4D). Indeed, there is a displacement, but, comparing the  $K_D$  of Tat binding to naked RNA(30)-mer (Figure 3D) to the apparent  $K_D$  after the binding of IN (Figure 3E), we observed a decrease of only about 6 times. Moreover, the experimental points followed a hyperbolic Michaelis–Menten kinetics (Figure 3E) different from the cooperative sigmoidal Hill kinetics observed for TAR (Figure 3C). Therefore, CT within IN-CTD enhances Tat binding to TAR and makes the binding cooperative.

To prove that these results were not the consequence of a direct interaction between IN-CTD and Tat, we performed a pulldown assay, which did not show protein–protein interaction in these conditions (Supplementary Figure S5D).

We also performed the competition experiment by reversing the order of addition of proteins (Supplementary Figure S5A); after TAR or RNA(30)-mer loading and wash, we first incubated the probe with Tat (Supplementary Figure S5B,C ascendant curve between 200 and 400 s). Then, we added a large amount of IN-CTD or IN-CTD- $\Delta$ CT (Supplementary Figure S5B,C, curve between 400 and 600 s). We did not observe any supplementary association of IN to TAR upon Tat binding. Overall, these results suggest that there is no competition between IN and Tat for TAR once Tat is bound first, nor a facilitating effect of Tat for IN binding to TAR. Therefore, this supports our hypothesis that IN binding to TAR precedes and facilitates Tat interaction.

#### 2.4. Tat Competes with IN-CTD and Displaces It from TAR

To investigate the mechanism involved in Tat binding to the IN:TAR complex, we also performed a dose-dependent competition EMSA (Figure 4A); we added increasing concentrations of Tat to a preformed IN-CTD: or IN-CTD- $\Delta$ CT:TAR complex. Interestingly, we observed a dose-dependent interference between Tat and IN-CTD for TAR binding as demonstrated by the reappearance of free RNA at intermediate concentrations of Tat, suggesting that the presence of Tat displaces IN from the TAR RNA (Figure 4A, lanes 4 and 5). This is further supported by the fact that this interference disappeared at saturating concentrations of Tat, where displacement of IN would be complete and RNA uniquely bound to Tat (Figure 4A, lane 6). Interestingly, no interference was observed at any concentration of IN-CTD- $\Delta$ CT (Figure 4A, lanes 9 to 12), thus implying that the mechanism was dependent on the presence of the CT tail.

Since the migration pattern of the EMSA assay cannot discriminate between IN-bound and Tat-bound RNAs, we assessed whether the presence of Tat would modify the binding of IN to 3' biotinylated TAR with pull-down assays (Figure 4B). IN-CTD and IN-CTD- $\Delta$ CT were efficiently co-precipitated (Figure 4B, lanes 1 and 5, respectively), whereas the addition of increasing concentrations of Tat serially decreased their binding to TAR (Figure 4B, lanes 2 to 4 and 6 to 8, respectively), confirming our hypothesis of a displacement of IN by Tat for TAR binding. Control experiments showed that Tat also interacted with TAR and was not precipitated when biotinylated TAR was absent (Figure 4B, lanes 9 and 10). Very small differences were detected between IN-CTD and IN-CTD- $\Delta$ CT during the TAR-pulldown assay (Figure 4B) compared to the EMSA assay (Figure 4A). This is likely due to the higher sensitivity offered by radioactivity labelling in EMSA, since nanomolar concentrations of proteins and RNA were used. Altogether, these results suggest that IN facilitates the



possible structurally defective TAR conformers, but it is not involved in the discrimination between viral structured RNA regions. Interestingly, this behaviour of IN-CT reminds us of that of p6, the C-terminal domain of HIV-1 Pr55Gag [60]. The Pr55Gag $\Delta$ p6 mutant, deleted from the p6 domain, showed no RNA binding specificity compared to the full-length Pr55Gag, suggesting that the presence of this region is required for the specific binding of Pr55Gag to DIS RNA within the 5'UTR of HIV-1 genome [60].

Previous biochemical and structural studies have demonstrated that the unstructured Tat Arginine-rich motif (ARM) penetrates the TAR major groove, which is made with stem-bulge-stem-loop secondary structures and mostly interacts with the U-rich bulge and nearby double-stranded regions (Supplementary Figure S6A; [39,61–63]). Despite its absence in the CT of a motif comparable to the Arginine stretch of Tat, our results suggest that the unstructured IN-CT region also binds the TAR major groove, but it does so through the interaction with G34 and, to a lesser extent, G36 within the hexaloop (Figure 3A). This could explain the lower affinity of IN-CTD for TAR mutants (Figure 2) where the G34 and G36 positions are altered/affected. Interestingly, upon IN-CTD or IN-CTD- $\Delta$ CT binding to TAR RNA, we observed an increased exposure of G32 of the hexaloop (Figure 3A). This result is also supported by a very recent study published by Liu and co-workers [57], in which a crosslinking-coupled SHAPE assay was used to probe the modifications induced in the 5'UTR upon IN-FL binding [57]. In their experiments, G32 also presented a greater reactivity after the binding of the HIV-1 IN-FL protein [57]. Moreover, SHAPE results indicated that IN-FL binding increased G32 reactivity, increased backbone flexibility of the apical loop of TAR, and increased crosslinking with C39 [57]. This observed flexibility could explain the TAR breakage around C39 that we observed upon IN-CTD and IN-CTD- $\Delta$ CT binding (namely U38 and C42, Figure 3A and Supplementary Figure S4A).

A complex of IN-CTD:TAR has been modelled following the IN-CTD:INI1183-304 structure based on the fact that the same 6 residues were engaged for the binding of IN-CTD to both INI1 and TAR [8,56] (Supplementary Figure S6B). In this model, IN-CTD binds the minor groove of the 4-nt apical stem of TAR. Noteworthy, the last C-terminal modelled residue (D270), which immediately precedes the CT tail, is oriented towards the major groove of TAR, which is consistent with our experimental hypothesis [56] (Supplementary Figure S6B).

The interaction of IN with gRNA has been shown to be critical for a proper localization of vRNP inside the protective capsid core [8,27,28,57]. In this context, TAR-selection is exerted by the IN-CT of the mature protein (alone or together with NC); while in the Pr160Gag-Pol precursor, this could be an additional mechanism to that of the Pr55Gag protein interaction with the packaging signals (reviewed in [64]) to allow the selective recruitment and encapsidation of the HIV-1 gRNA in the viral core.

Our data is also fully supported by recent *in vivo* evidence of the implication of IN in proviral transcription at early times after integration and in a Tat-independent manner [35]. The authors of this work demonstrated that the mutation of four Lysine residues within the IN-CTD was responsible for the reduction of proviral transcription [35]. Interestingly, the same Lysine residues are involved in the binding of IN to the viral RNA [8,27].

Along the same line, our data indicates that IN interacts with the TAR RNA and that this interaction favours the interaction of Tat to TAR. Indeed, the analysis of real-time binding kinetics revealed that the affinity of Tat for TAR increased when the latter is complexed with IN-CTD or IN-CTD- $\Delta$ CT beforehand (Figure 3B–E). This binding behaviour and the cooperative kinetics seem specific for TAR RNA compared to generic RNAs (Figure 3C–E). Moreover, our experiments suggest a temporal directionality of the binding reaction; Tat can bind the IN-CTD/CTD- $\Delta$ CT:RNA complexes but IN-CTD/CTD- $\Delta$ CT cannot bind Tat:RNA (Supplementary Figure S5). Therefore, based on our data, we propose a temporal sequence for the interplay between IN, Tat, and TAR RNA in the early events of proviral transcription from the LTR; first, IN-CTD binds TAR on the nascent RNA through its C-terminal tail by contacting the apical stem-loop and, in particular, the hexaloop. This interaction modifies the structure of TAR (Figure 3A), promoting Tat binding (Figure 3B,C),

which finally displaces IN from the TAR RNA in order to boost transcription elongation (Figure 4). Noteworthy, the competition of the Tat ARM with IN for TAR binding has also been reported elsewhere [8].

## 4. Materials and Methods

### 4.1. Protein Expression and Purification

Construction of plasmid pET15b encoding N-terminal 6XHis tagged IN-CTD (aa 220-270) and IN-CTD- $\Delta$ CT (aa 220-288) were previously reported [33]. The IN-full length (IN-FL) gene from pNL4-3 was cloned in a pPROEX-HTa vector in a frame with a 6XHis tag at the N-terminus. All proteins were expressed in *Escherichia coli* BL21(DE3) Rosetta/pLysS strain (Novagen, Sigma Aldrich, Saint-Quentin-Fallavier, France). Cells were grown in an LB medium supplemented with 10% (*w/v*) glucose, and protein expression was induced at an optical density at 600 nm (OD<sub>600</sub>) of 0.6 with 1 mM IPTG (isopropyl- $\beta$ -D-thiogalactopyranoside). Cells were incubated overnight at 18 °C under continuous shaking then pelleted. The cell pellets were resuspended in a lysis buffer composed of 50 mM Tris-HCl pH 8, 1 M NaCl, 20 mM imidazole, 0.1 mM EDTA, 2 mM  $\beta$ -mercaptoethanol, and 10% (*w/v*) glycerol. Exclusively for bacterial lysis, the buffer was freshly supplemented with 2 M urea, 2 mM of Adenosine triphosphate (ATP), 5mM CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate), and 1 tablet of Protease Inhibitor Cocktail (ROCHE, cOmplete™). The preparation was sonicated for 120 s on ice, then the resulting lysate was subjected to centrifugation at 11,000  $\times$  g for 1 h. The recovered supernatant was then applied to a HisTrap™ Fast Flow Crude column (Cytiva Europe GmbH, France) and purified by an AKTA pure system (Cytiva Europe GmbH, France). The sample was first abundantly washed with a lysis buffer containing 100 mM imidazole and 2 M NaCl, then the protein was eluted using a gradient up to 500 mM imidazole in 10 column volumes. A second step of purification was carried out using a Superdex 75 10/300 GL column (Cytiva Europe GmbH, France) with an isocratic elution carried out with a storing buffer (50 mM HEPES pH 7.5, 1 M NaCl, 5 mM  $\beta$ -mercaptoethanol, and 5% glycerol). Mammalian FlagIN-FL (IN-FLm) was expressed in Baby Hamster Kidney suspension cells (BHK21-C13-2P, Sigma Aldrich, Saint-Quentin-Fallavier, France) using a vaccinia virus expression system as previously described [45]. IN-CT peptide (YGKQ-MAGDDCVASRQDED) and 101-residue long Tat protein of the primary isolate 133 of HIV-1 were chemically synthesized [65,66]. This Tat has been biochemically characterized and its full biological activity had been previously validated [67].

### 4.2. In Vitro RNA Synthesis, Purification, and Radiolabeling

We produced several RNAs as listed in Supplementary Table S1 using partially double-stranded templates formed by the hybridization of the T7 promoter-containing DNA oligonucleotides listed in Supplementary Table S2 and following the protocol detailed in [68]. Templates for polyA, DIS, and SD/Psi RNAs were produced with PCR using a pNL4-3 plasmid as a template and T7 promoter-containing primers. RNA was transcribed with kit MEGAshortscript™ T7 (Thermo Fisher Scientific, Illkirch-Graffenstaden, France) following the manufacturer's instructions, de-phosphorylated using Alkaline Phosphatase (New England Biolabs, Evry, France), and purified with phenol:chloroform:isoamyl alcohol (25:24:1) extraction and ethanol precipitation. The 3' biotinylated TAR and RNA 30-mer (Supplementary Table S1) were chemically synthesized (Sigma Aldrich, Saint-Quentin-Fallavier, France). RNA (50 pmol) was radiolabelled at the 5' end using 10 units of T4 polynucleotide kinase (New England Biolabs) mixed to 3  $\mu$ L of  $\gamma$ -<sup>32</sup>P-ATP (3000 Ci/mmol, 10 mCi/mL, Perkin Elmer, Villebon-sur-Yvette, France) for 1 h at 37 °C then purified on denaturing 10% (*w/v*) polyacrylamide gel (29:1) as previously described [68]. Before use in binding and structural studies, RNA was heated in a refolding buffer (20 mM HEPES pH 7.5, 0.2 M NaCl, 2 mM MgCl<sub>2</sub>, 2 mM DTT) for 3 min at 95 °C followed by 40 min of slow controlled cooling to room temperature and finally placed on ice.

#### 4.3. Electrophoretic Mobility Shift Assays (EMSA)

The electrophoretic mobility shift assays (EMSA) were performed as described in [69]. Samples were prepared by mixing a radiolabelled RNA with increasing concentrations of proteins, as indicated, in a buffer containing 20 mM MES pH 6.0, 150 mM NaCl, 2 mM DTT, 2 mM MgCl<sub>2</sub>, 0.2 µg BSA, and 8% (v/v) PEG8000. The samples were incubated at 37 °C for 30 min before being resolved by native 6% polyacrylamide (19:1) gel electrophoresis in a 0.5× TAE (Tris acetate EDTA) buffer. Results were analysed by phosphorimaging using ImageQuant software. IN-FLm was incubated for 2 h at 4 °C with the RNA substrate. For the dose-dependent competition assay shown in Figure 4A, we used for all the samples a constant saturating IN-to-TAR concentrations with an IN-to-TAR ratio per sample exceeding about ten times their respective K<sub>D</sub>. Then, we added increasing concentrations of Tat.

#### 4.4. RNA Structural Probing

Enzymatic treatments were performed in 10 µL of the reaction mix containing 0.5 pmol of 5' radiolabeled RNA, 0.2 µg of yeast tRNA, 1× Structure buffer (Thermo Fisher Scientific, Illkirch-Graffenstaden, France), and 0.01 U of RNase T1 (Thermo Fisher Scientific, Illkirch-Graffenstaden, France). Incubation was done at 37 °C for 5 min. Reactions were stopped by the addition of 40 µL of a quenching buffer composed of 10 mM HEPES pH 7.5, 1 mM EDTA, and 3% SDS. Partial alkaline hydrolysis was performed as follows; 10 µL of reaction mix containing 0.5 pmol of RNA, 1 µg of yeast tRNA, and 1× Alkaline Hydrolysis buffer were incubated at 95 °C for 12 min then quenched with 2× denaturing loading buffer and placed on ice. For RNA/protein complexes, 0.5 pmol RNA were previously incubated with a 36 µM protein (final concentration) at 37 °C for 30 min then treated with RNase T1. After quenching, samples were phenol extracted and ethanol precipitated. After recovering from precipitation, all samples were run on a 15% sequencing polyacrylamide gel in 0.5× TBE (Tris Borate EDTA). Results were analysed by phosphorimaging.

#### 4.5. Pulldown Assay

The pulldown assays were conducted as described before [68]. Briefly, proteins were mixed in a binding buffer (40 mM HEPES pH 7.5, 2.5 mM MgCl<sub>2</sub>, 2 mM β-mercaptoethanol and 5% glycerol) adjusting the final NaCl concentration to 150 mM. Samples were complemented or not with 600 pmol of 3' end-biotinylated TAR (Supplementary Table S1) in a final volume of 30 µL, and incubated 30 min at 37 °C. To the mix were added 8 µL of magnetic streptavidin beads (Dynabeads MyOne, Thermo Fischer Scientific, Illkirch-Graffenstaden, France), and the mix was further incubated for 1h at 4 °C in a gentle rotation. The resin was washed three times with 500 µL BB-200 (40 mM HEPES pH 7.5, 200 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 2 mM β-mercaptoethanol and 10% glycerol) on ice and proteins were eluted with an SDS loading buffer and analysed on a polyacrylamide 16% (37.5:1) SDS-PAGE. For the Histidine pulldown assay, magnetic streptavidin resin was replaced with HisPur™ Ni-NTA Magnetic Beads (Thermo Fischer Scientific, Illkirch-Graffenstaden, France) and protein elution was done with a 0.5 M Imidazole containing buffer.

#### 4.6. Bio-Layer Interferometry

For the Bio-Layer Interferometry (BLI) analysis, we used the BLItz platform (FortéBio). For the high throughput experiments such as shown in Figure 1F and Supplementary Figure S2C, we used the Octet RED96e System (FortéBio, Avantor, VWR, Fontenay-sous-Bois, France). All measurements were performed in an assay buffer composed of 50 mM HEPES pH 7.5, 200 mM NaCl, 2 mM β-mercaptoethanol, 5% glycerol, 2 mM MgCl<sub>2</sub>, 1 µM ZnSO<sub>4</sub>, and 0.5 M BSA. For the BLItz platform, we used Streptavidin (SA) Biosensors (FortéBio, Avantor, VWR, Fontenay-sous-Bois, France), and for the Octet RED96e, we used the Streptavidin (SAX) Biosensors (FortéBio, Avantor, VWR, Fontenay-sous-Bois, France) that were hydrated for 10 min in an assay buffer then plunged in a solution containing 1 µM 3' end-biotinylated TAR RNA in an assay buffer with a 1× RNAase inhibitor (RNA

Secure, Invitrogen, Thermo Fischer Scientific, Illkirch-Graffenstaden, France) and 0.1  $\mu$ M BSA for the RNA loading step. A wash was performed after RNA loading. In the BLItz experiment, the association of the protein to RNA was monitored in real-time for 200 s; the hydrated biosensor tip was placed in a 1.5 mL black assay tube containing 200  $\mu$ L of the protein solution in an assay buffer as indicated. Afterward, a dissociation step or a second association with Tat was performed for 200 s. Biosensors were discarded after each measurement. All kinetic assays shown in Figure 1F and Supplementary Figure S2C were performed using the Octet RED96e system and carried out using black 96-well plates, and samples were diluted in a freshly prepared assay buffer and incubated at 37 °C with an orbital shake. Association and dissociation steps were as in the BLItz experiments. Each time, reference sensors and negative control sensors were included. Sensorgrams were exported and data analysis was performed with KaleidaGraph software (Synergy Software, Reading, PA, USA).

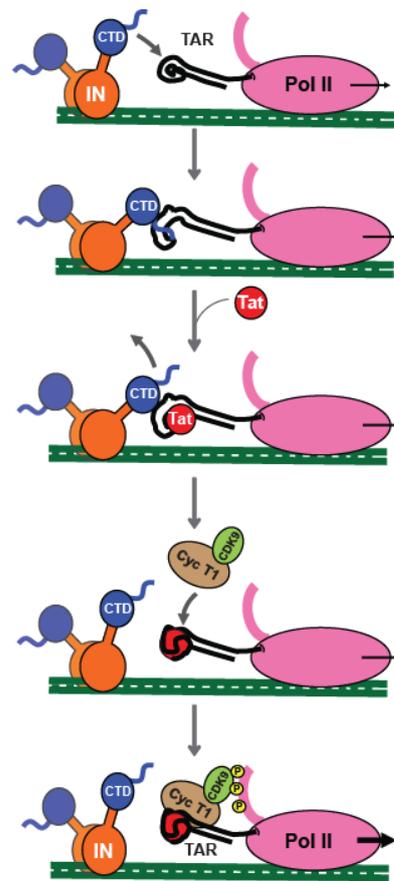
## 5. Conclusions

In this study, we have brought quantitative evidence to the role exerted by HIV-1 Integrase C-terminal domain (IN-CTD) as an RNA binding protein in the very first stages of proviral transcription. In particular, we have shown that full length IN, despite its intrinsic instability and independently of the post-transcriptional modifications (i.e., recombinantly expressed in *E. coli* or in mammalian cells), binds TAR RNA (Figure 1 and Supplementary Figure S1). The C-terminal domain (IN-CTD) also binds TAR RNA and the structured parts of the HIV-1 5'UTR (Figure 1D) as the full length, indirectly proving that IN-CTD is indeed the RNA-binding domain. Moreover, the last 18 residues of IN-CTD, which are predicted to be unstructured, only slightly affect the affinity of IN-CTD for TAR (Figure 1F). On the other hand, this same tail (IN-CT) is able to sense the shape of TAR RNA, in particular the hexaloop, and to modify it by exposing specific nucleotides, namely G32, U38, and C41 (Figure 2) and by protecting G34 and G36, as it was seen in *in vivo* SHAPE experiments [52]. This modification is preparatory to Tat binding to the same TAR RNA. Interestingly, the apparent affinity of Tat for TAR was about 115 times lower in the presence of IN-CTD and about 35 times lower in the presence of IN-CTD- $\Delta$ CT. Moreover, Tat binding to IN/TAR is cooperative (Figure 3). We further proved Tat competition and its ability to displace IN-CTD from TAR by interferometry, EMSA, and pull-down assays.

Overall, our *in vitro* data complements the *in vivo* observations [34] and together converge to the working model schematized in Figure 5, where IN-CTD selectively binds the nascent TAR transcript hexaloop with its CT tail. This binding modifies the TAR structure, facilitating its interaction with Tat, then Tat displaces IN and allows the subsequent trans-activation of provirus transcription (Figure 5 bottom panel). This process is dynamic and might also be modulated by the post-translational modifications of IN, by the interaction with cellular partners, and/or by the chromatin remodeling processes.

Noteworthy, the structure of the HIV-1 intasome during the strand transfer process revealed that the inner IN tetrameric core, which contacts both the viral and host DNA, is surrounded by twelve other subunits [64,65] that might be available for other functions [11]. The structure of IN after strand transfer is not yet available, thus it is not possible to predict whether the CTD is at disposal for transcription.

Finally, our data adds to studies demonstrating the pleiotropic role of IN in viral packaging (reviewed in [27,52]), integration (reviewed in [9]), and early transcription [34] via its protein–RNA interactions. The direct role of CT in TAR binding puts IN back centre stage as a therapeutic target and opens the way to the development of a new generation of therapeutics.



**Figure 5.** Cartoon of the working Model proposed for the interplay of IN-CTD, Tat and TAR during the early stage of HIV-1 proviral transcription. Once the integration has taken place, IN interacts with nascent TAR transcript through its CTD, directly interacting with the major groove and the hexalooop, thanks to its C-terminal tail. This binding induces TAR conformational changes, which promote the interaction of Tat with its substrate. Afterwards, Tat displaces IN-CTD from TAR and recruits the SEC complex in order to boost Pol II transcription.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms232213742/s1>.

**Author Contributions:** C.R., C.L. and F.F. conceived and designed experiments. C.R., C.L., A.E.M., C.G., L.B., D.L. and F.F. performed the experiments and/or analyzed data. J.B. and M.R. prepared protein from mammalian expression system; C.R., C.L., A.E.M., C.G., L.B., D.L., M.N., M.R., P.G. and F.F. discussed the results; F.F. wrote the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the French Agency for Research on AIDS and Viral Hepatitis (ECTZ72240-1 to F.F., M.R. and M.N.; Ph.D. fellowship for C.R.).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All additional data are available upon request.

**Acknowledgments:** We acknowledge: the support and the use of resources of the French Infrastructure for Integrated Structural Biology FRISBI ANR-10-INBS-05 and of Instruct-ERIC; Sylvie Ricard-Blum and the Interaction platform of the UMR 5246 ICBMS for the experiments on the Octet RED96; the Protein Science Facility (PSF) of SFR Biosciences (UAR3444/CNRS, US8/Inserm, ENS de Lyon, UCBL) especially Virginie Gueguen-Chaignon and Eric Diesis for their assistance, and Frédéric Galisson for his help with radioactive sources. We are grateful to the Viral DNA Integra-

tion and Chromatin Dynamics Network (DyNAVIR) for fruitful discussions and to Anne-Catherine Dock-Bregeon for critical reading of the manuscript. We acknowledge Andrea Cimarelli for pNL4-3 plasmid donation.

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

## References

1. Beckmann, B.M.; Castello, A.; Medenbach, J. The expanding universe of ribonucleoproteins: Of novel RNA-binding proteins and unconventional interactions. *Pflugers Arch.-Eur. J. Physiol.* **2016**, *468*, 1029–1040. [[CrossRef](#)] [[PubMed](#)]
2. Castello, A.; Fischer, B.; Eichelbaum, K.; Horos, R.; Beckmann, B.M.; Strein, C.; Davey, N.E.; Humphreys, D.T.; Preiss, T.; Steinmetz, L.M.; et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **2012**, *149*, 1393–1406. [[CrossRef](#)] [[PubMed](#)]
3. Hentze, M.W.; Castello, A.; Schwarzl, T.; Preiss, T. A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 327–341. [[CrossRef](#)] [[PubMed](#)]
4. Brodsky, S.; Jana, T.; Mittelman, K.; Chapal, M.; Kumar, D.K.; Carmi, M.; Barkai, N. Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Mol. Cell* **2020**, *79*, 459–471.e4. [[CrossRef](#)]
5. Cassidy, L.A.; Maher, L.J., 3rd. Having it both ways: Transcription factors that bind DNA and RNA. *Nucleic Acids Res.* **2002**, *30*, 4118–4126. [[CrossRef](#)]
6. Basu, S.; Bahadur, R.P. A structural perspective of RNA recognition by intrinsically disordered proteins. *Cell. Mol. Life Sci.* **2016**, *73*, 4075–4084. [[CrossRef](#)]
7. Garcia-Moreno, M.; Jarvelin, A.I.; Castello, A. Unconventional RNA-binding proteins step into the virus-host battlefield. *Wiley Interdiscip. Rev. RNA* **2018**, *9*, e1498. [[CrossRef](#)]
8. Kessl, J.J.; Kutluay, S.B.; Townsend, D.; Rebersburg, S.; Slaughter, A.; Larue, R.C.; Shkriabai, N.; Bakouche, N.; Fuchs, J.R.; Bieniasz, P.D.; et al. HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. *Cell* **2016**, *166*, 1257–1268.e12. [[CrossRef](#)]
9. Maertens, G.N.; Engelman, A.N.; Cherepanov, P. Structure and function of retroviral integrase. *Nat. Rev. Microbiol.* **2022**, *20*, 22–34. [[CrossRef](#)]
10. Sakai, H.; Kawamura, M.; Sakuragi, J.; Sakuragi, S.; Shibata, R.; Ishimoto, A.; Ono, N.; Ueda, S.; Adachi, A. Integration is essential for efficient gene expression of human immunodeficiency virus type 1. *J. Virol.* **1993**, *67*, 1169–1174. [[CrossRef](#)]
11. Rocchi, C.; Gouet, P.; Parissi, V.; Fiorini, F. The C-Terminal Domain of HIV-1 Integrase: A Swiss Army Knife for the Virus? *Viruses* **2022**, *14*, 1397. [[CrossRef](#)] [[PubMed](#)]
12. Esposito, D.; Craigie, R. HIV integrase structure and function. *Adv. Virus Res.* **1999**, *52*, 319–333. [[PubMed](#)]
13. Dyda, F.; Hickman, A.B.; Jenkins, T.M.; Engelman, A.; Craigie, R.; Davies, D.R. Crystal structure of the catalytic domain of HIV-1 integrase: Similarity to other polynucleotidyl transferases. *Science* **1994**, *266*, 1981–1986. [[CrossRef](#)] [[PubMed](#)]
14. Cai, M.; Zheng, R.; Caffrey, M.; Craigie, R.; Clore, G.M.; Gronenborn, A.M. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* **1997**, *4*, 567–577. [[CrossRef](#)] [[PubMed](#)]
15. Lee, S.P.; Xiao, J.; Knutson, J.R.; Lewis, M.S.; Han, M.K. Zn<sup>2+</sup> promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. *Biochemistry* **1997**, *36*, 173–180. [[CrossRef](#)] [[PubMed](#)]
16. Zheng, R.; Jenkins, T.M.; Craigie, R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13659–13664. [[CrossRef](#)]
17. Eijkelenboom, A.P.; Lutzke, R.A.; Boelens, R.; Plasterk, R.H.; Kaptein, R.; Hard, K. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nat. Struct. Biol.* **1995**, *2*, 807–810. [[CrossRef](#)]
18. Engelman, A.; Hickman, A.B.; Craigie, R. The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J. Virol.* **1994**, *68*, 5911–5917. [[CrossRef](#)]
19. Jenkins, T.M.; Engelman, A.; Ghirlando, R.; Craigie, R. A soluble active mutant of HIV-1 integrase: Involvement of both the core and carboxyl-terminal domains in multimerization. *J. Biol. Chem.* **1996**, *271*, 7712–7718. [[CrossRef](#)]
20. Woerner, A.M.; Marcus-Sekura, C.J. Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. *Nucleic Acids Res.* **1993**, *21*, 3507–3511. [[CrossRef](#)]
21. Lutzke, R.A.; Vink, C.; Plasterk, R.H. Characterization of the minimal DNA-binding domain of the HIV integrase protein. *Nucleic Acids Res.* **1994**, *22*, 4125–4131. [[CrossRef](#)] [[PubMed](#)]
22. Vink, C.; Oude Groeneger, A.M.; Plasterk, R.H. Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucleic Acids Res.* **1993**, *21*, 1419–1425. [[CrossRef](#)] [[PubMed](#)]
23. Dar, M.J.; Monel, B.; Krishnan, L.; Shun, M.C.; Di Nunzio, F.; Helland, D.E.; Engelman, A. Biochemical and virological analysis of the 18-residue C-terminal tail of HIV-1 integrase. *Retrovirology* **2009**, *6*, 94. [[CrossRef](#)] [[PubMed](#)]
24. De Houwer, S.; Demeulemeester, J.; Thys, W.; Taltynov, O.; Zmajkovicova, K.; Christ, F.; Debyser, Z. Identification of residues in the C-terminal domain of HIV-1 integrase that mediate binding to the transportin-SR2 protein. *J. Biol. Chem.* **2012**, *287*, 34059–34068. [[CrossRef](#)]

25. Mauro, E.; Lesbats, P.; Lapailierie, D.; Chaignepain, S.; Maillot, B.; Oladosu, O.; Robert, X.; Fiorini, F.; Kieffer, B.; Bouaziz, S.; et al. Human H4 tail stimulates HIV-1 integration through binding to the carboxy-terminal domain of integrase. *Nucleic Acids Res.* **2019**, *47*, 3607–3618. [[CrossRef](#)]
26. Mohammed, K.D.; Topper, M.B.; Muesing, M.A. Sequential deletion of the integrase (Gag-Pol) carboxyl terminus reveals distinct phenotypic classes of defective HIV-1. *J. Virol.* **2011**, *85*, 4654–4666. [[CrossRef](#)]
27. Elliott, J.L.; Kutluay, S.B. Going beyond Integration: The Emerging Role of HIV-1 Integrase in Virion Morphogenesis. *Viruses* **2020**, *12*, 1005. [[CrossRef](#)]
28. Madison, M.K.; Lawson, D.Q.; Elliott, J.; Ozanturk, A.N.; Koneru, P.C.; Townsend, D.; Errando, M.; Kvaratskhelia, M.; Kutluay, S.B. Allosteric HIV-1 Integrase Inhibitors Lead to Premature Degradation of the Viral RNA Genome and Integrase in Target Cells. *J. Virol.* **2017**, *91*, e00821-17. [[CrossRef](#)]
29. Fontana, J.; Jurado, K.A.; Cheng, N.; Ly, N.L.; Fuchs, J.R.; Gorelick, R.J.; Engelman, A.N.; Steven, A.C. Distribution and Redistribution of HIV-1 Nucleocapsid Protein in Immature, Mature, and Integrase-Inhibited Virions: A Role for Integrase in Maturation. *J. Virol.* **2015**, *89*, 9765–9780. [[CrossRef](#)]
30. Cereseto, A.; Manganaro, L.; Gutierrez, M.I.; Terreni, M.; Fittipaldi, A.; Lusic, M.; Marcello, A.; Giacca, M. Acetylation of HIV-1 integrase by p300 regulates viral integration. *EMBO J.* **2005**, *24*, 3070–3081. [[CrossRef](#)]
31. Manganaro, L.; Lusic, M.; Gutierrez, M.I.; Cereseto, A.; Del Sal, G.; Giacca, M. Concerted action of cellular JNK and Pin1 restricts HIV-1 genome integration to activated CD4+ T lymphocytes. *Nat. Med.* **2010**, *16*, 329–333. [[CrossRef](#)] [[PubMed](#)]
32. Zamborlini, A.; Coiffic, A.; Beauclair, G.; Delelis, O.; Paris, J.; Koh, Y.; Magne, F.; Giron, M.L.; Tobaly-Tapiero, J.; Deprez, E.; et al. Impairment of human immunodeficiency virus type-1 integrase SUMOylation correlates with an early replication defect. *J. Biol. Chem.* **2011**, *286*, 21013–21022. [[CrossRef](#)] [[PubMed](#)]
33. Kanja, M.; Cappy, P.; Levy, N.; Oladosu, O.; Schmidt, S.; Rossolillo, P.; Winter, F.; Gasser, R.; Moog, C.; Ruff, M.; et al. NKNK: A New Essential Motif in the C-Terminal Domain of HIV-1 Group M Integrases. *J. Virol.* **2020**, *94*, e01035-20. [[CrossRef](#)] [[PubMed](#)]
34. Shema Mugisha, C.; Dinh, T.; Kumar, A.; Tanneti, K.; Eschbach, J.E.; Davis, K.; Gifford, R.; Kvaratskhelia, M.; Kutluay, S.B. Emergence of Compensatory Mutations Reveals the Importance of Electrostatic Interactions between HIV-1 Integrase and Genomic RNA. *mBio* **2022**, *13*, e00431-22. [[CrossRef](#)]
35. Winans, S.; Goff, S.P. Mutations altering acetylated residues in the CTD of HIV-1 integrase cause defects in proviral transcription at early times after integration of viral DNA. *PLoS Pathog.* **2020**, *16*, e1009147. [[CrossRef](#)]
36. Jonkers, I.; Lis, J.T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 167–177. [[CrossRef](#)]
37. Mousseau, G.; Valente, S.T. Role of Host Factors on the Regulation of Tat-Mediated HIV-1 Transcription. *Curr. Pharm. Des.* **2017**, *23*, 4079–4090. [[CrossRef](#)]
38. He, N.; Liu, M.; Hsu, J.; Xue, Y.; Chou, S.; Burlingame, A.; Krogan, N.J.; Alber, T.; Zhou, Q. HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. *Mol. Cell* **2010**, *38*, 428–438. [[CrossRef](#)]
39. Pham, V.V.; Salguero, C.; Khan, S.N.; Meagher, J.L.; Brown, W.C.; Humbert, N.; de Rocquigny, H.; Smith, J.L.; D'Souza, V.M. HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. *Nat. Commun.* **2018**, *9*, 4266. [[CrossRef](#)]
40. Schulze-Gahmen, U.; Hurley, J.H. Structural mechanism for HIV-1 TAR loop recognition by Tat and the super elongation complex. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 12973–12978. [[CrossRef](#)]
41. Sobhian, B.; Laguette, N.; Yatim, A.; Nakamura, M.; Levy, Y.; Kiernan, R.; Benkirane, M. HIV-1 Tat assembles a multifunctional transcription elongation complex and stably associates with the 7SK snRNP. *Mol. Cell* **2010**, *38*, 439–451. [[CrossRef](#)] [[PubMed](#)]
42. Wei, P.; Garber, M.E.; Fang, S.M.; Fischer, W.H.; Jones, K.A. 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* **1998**, *92*, 451–462. [[CrossRef](#)]
43. Ne, E.; Palstra, R.J.; Mahmoudi, T. Transcription: Insights From the HIV-1 Promoter. *Int. Rev. Cell Mol. Biol.* **2018**, *335*, 191–243. [[PubMed](#)]
44. Craigie, R. The molecular biology of HIV integrase. *Future Virol.* **2012**, *7*, 679–686. [[CrossRef](#)]
45. Levy, N.; Eiler, S.; Pradeau-Aubret, K.; Maillot, B.; Stricher, F.; Ruff, M. Production of unstable proteins through the formation of stable core complexes. *Nat. Commun.* **2016**, *7*, 10932. [[CrossRef](#)]
46. Eijkelenboom, A.P.; Sprangers, R.; Hard, K.; Puras Lutzke, R.A.; Plasterk, R.H.; Boelens, R.; Kaptein, R. Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase. *Proteins* **1999**, *36*, 556–564. [[CrossRef](#)]
47. Lodi, P.J.; Ernst, J.A.; Kuszewski, J.; Hickman, A.B.; Engelman, A.; Craigie, R.; Clore, G.M.; Gronenborn, A.M. Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* **1995**, *34*, 9826–9833. [[CrossRef](#)]
48. Gruber, A.R.; Lorenz, R.; Bernhart, S.H.; Neubock, R.; Hofacker, I.L. The Vienna RNA websuite. *Nucleic Acids Res.* **2008**, *36*, W70–W74. [[CrossRef](#)]
49. Lorenz, R.; Bernhart, S.H.; Honer Zu Siederdisen, C.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **2011**, *6*, 26. [[CrossRef](#)]
50. Robert, X.; Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014**, *42*, W320–W324. [[CrossRef](#)]
51. Muller-Esparza, H.; Osorio-Valeriano, M.; Steube, N.; Thanbichler, M.; Randau, L. Bio-Layer Interferometry Analysis of the Target Binding Activity of CRISPR-Cas Effector Complexes. *Front. Mol. Biosci.* **2020**, *7*, 98. [[CrossRef](#)] [[PubMed](#)]

52. Berkhout, B.; Jeang, K.T. Detailed mutational analysis of TAR RNA: Critical spacing between the bulge and loop recognition domains. *Nucleic Acids Res.* **1991**, *19*, 6169–6176. [[CrossRef](#)] [[PubMed](#)]
53. Churcher, M.J.; Lamont, C.; Hamy, F.; Dingwall, C.; Green, S.M.; Lowe, A.D.; Butler, J.G.; Gait, M.J.; Karn, J. High affinity binding of TAR RNA by the human immunodeficiency virus type-1 tat protein requires base-pairs in the RNA stem and amino acid residues flanking the basic region. *J. Mol. Biol.* **1993**, *230*, 90–110. [[CrossRef](#)] [[PubMed](#)]
54. Delling, U.; Reid, L.S.; Barnett, R.W.; Ma, M.Y.; Climie, S.; Sumner-Smith, M.; Sonenberg, N. 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.* **1992**, *66*, 3018–3025. [[CrossRef](#)]
55. Roy, S.; Delling, U.; Chen, C.H.; Rosen, C.A.; Sonenberg, N. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes Dev.* **1990**, *4*, 1365–1373. [[CrossRef](#)]
56. Dixit, U.; Bhutoria, S.; Wu, X.; Qiu, L.; Spira, M.; Mathew, S.; Harris, R.; Adams, L.J.; Cahill, S.; Pathak, R.; et al. INI1/SMARCB1 Rpt1 domain mimics TAR RNA in binding to integrase to facilitate HIV-1 replication. *Nat. Commun.* **2021**, *12*, 2743. [[CrossRef](#)]
57. Liu, S.; Koneru, P.C.; Li, W.; Pathirage, C.; Engelman, A.N.; Kvaratskhelia, M.; Musier-Forsyth, K. HIV-1 integrase binding to genomic RNA 5'-UTR induces local structural changes in vitro and in virio. *Retrovirology* **2021**, *18*, 37. [[CrossRef](#)]
58. Dingwall, C.; Ernberg, I.; Gait, M.J.; Green, S.M.; Heaphy, S.; Karn, J.; Lowe, A.D.; Singh, M.; Skinner, M.A.; Valerio, R. Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 6925–6929. [[CrossRef](#)]
59. Lu, R.; Ghory, H.Z.; Engelman, A. Genetic analyses of conserved residues in the carboxyl-terminal domain of human immunodeficiency virus type 1 integrase. *J. Virol.* **2005**, *79*, 10356–10368. [[CrossRef](#)]
60. Dubois, N.; Khoo, K.K.; Ghossein, S.; Seissler, T.; Wolff, P.; McKinstry, W.J.; Mak, J.; Paillart, J.C.; Marquet, R.; Bernacchi, S. The C-terminal p6 domain of the HIV-1 Pr55(Gag) precursor is required for specific binding to the genomic RNA. *RNA Biol.* **2018**, *15*, 923–936. [[CrossRef](#)]
61. Cordingley, M.G.; LaFemina, R.L.; Callahan, P.L.; Condra, J.H.; Sardana, V.V.; Graham, D.J.; Nguyen, T.M.; LeGrow, K.; Gotlib, L.; Schlabach, A.J.; et al. Sequence-specific interaction of Tat protein and Tat peptides with the transactivation-responsive sequence element of human immunodeficiency virus type 1 in vitro. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8985–8989. [[CrossRef](#)] [[PubMed](#)]
62. Puglisi, J.D.; Tan, R.; Calnan, B.J.; Frankel, A.D.; Williamson, J.R. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **1992**, *257*, 76–80. [[CrossRef](#)] [[PubMed](#)]
63. Weeks, K.M.; Ampe, C.; Schultz, S.C.; Steitz, T.A.; Crothers, D.M. Fragments of the HIV-1 Tat protein specifically bind TAR RNA. *Science* **1990**, *249*, 1281–1285. [[CrossRef](#)] [[PubMed](#)]
64. Comas-Garcia, M.; Davis, S.R.; Rein, A. On the Selective Packaging of Genomic RNA by HIV-1. *Viruses* **2016**, *8*, 246. [[CrossRef](#)]
65. Mayol, K.; Munier, S.; Beck, A.; Verrier, B.; Guillon, C. Design and characterization of an HIV-1 Tat mutant: Inactivation of viral and cellular functions but not antigenicity. *Vaccine* **2007**, *25*, 6047–6060. [[CrossRef](#)]
66. Guillon, C.; Stankovic, K.; Ataman-Onal, Y.; Biron, F.; Verrier, B. Evidence for CTL-mediated selection of Tat and Rev mutants after the onset of the asymptomatic period during HIV type 1 infection. *AIDS Res. Hum. Retrovir.* **2006**, *22*, 1283–1292. [[CrossRef](#)]
67. Foucault, M.; Mayol, K.; Receveur-Brechot, V.; Bussat, M.C.; Klinguer-Hamour, C.; Verrier, B.; Beck, A.; Haser, R.; Gouet, P.; Guillon, C. UV and X-ray structural studies of a 101-residue long Tat protein from a HIV-1 primary isolate and of its mutated, detoxified, vaccine candidate. *Proteins* **2010**, *78*, 1441–1456. [[CrossRef](#)]
68. Fiorini, F.; Bonneau, F.; Le Hir, H. Biochemical characterization of the RNA helicase UPF1 involved in nonsense-mediated mRNA decay. *Methods Enzymol.* **2012**, *511*, 255–274.
69. Fiorini, F.; Boudvillain, M.; Le Hir, H. Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. *Nucleic Acids Res.* **2013**, *41*, 2404–2415. [[CrossRef](#)]

## **The HIV-1 Integrase C-Terminal domain induces TAR RNA structural changes promoting Tat binding.**

Cecilia Rocchi<sup>1§</sup>, Camille Louvat<sup>1§</sup>, Adriana Erica Miele<sup>2,3</sup>, Julien Batisse<sup>4</sup>, Christophe Guillon<sup>1</sup>, Lionel Ballut<sup>1</sup>, Daniela Lener<sup>5</sup>, Matteo Negroni<sup>5</sup>, Marc Ruff<sup>4</sup>, Patrice Gouet<sup>1</sup> and Francesca Fiorini<sup>1\*</sup>.

<sup>1</sup> Molecular Microbiology and Structural Biochemistry, MMSB-IBCP, UMR 5086 CNRS University of Lyon, 7 passage du Vercors, 69367, Lyon Cedex 07, France.

<sup>2</sup> Institute of Analytical Sciences, UMR 5280 CNRS UCBL University of Lyon, 5 Rue de la Doua, 69100 Villeurbanne, France. ORCID 0000-0002-4637-2606

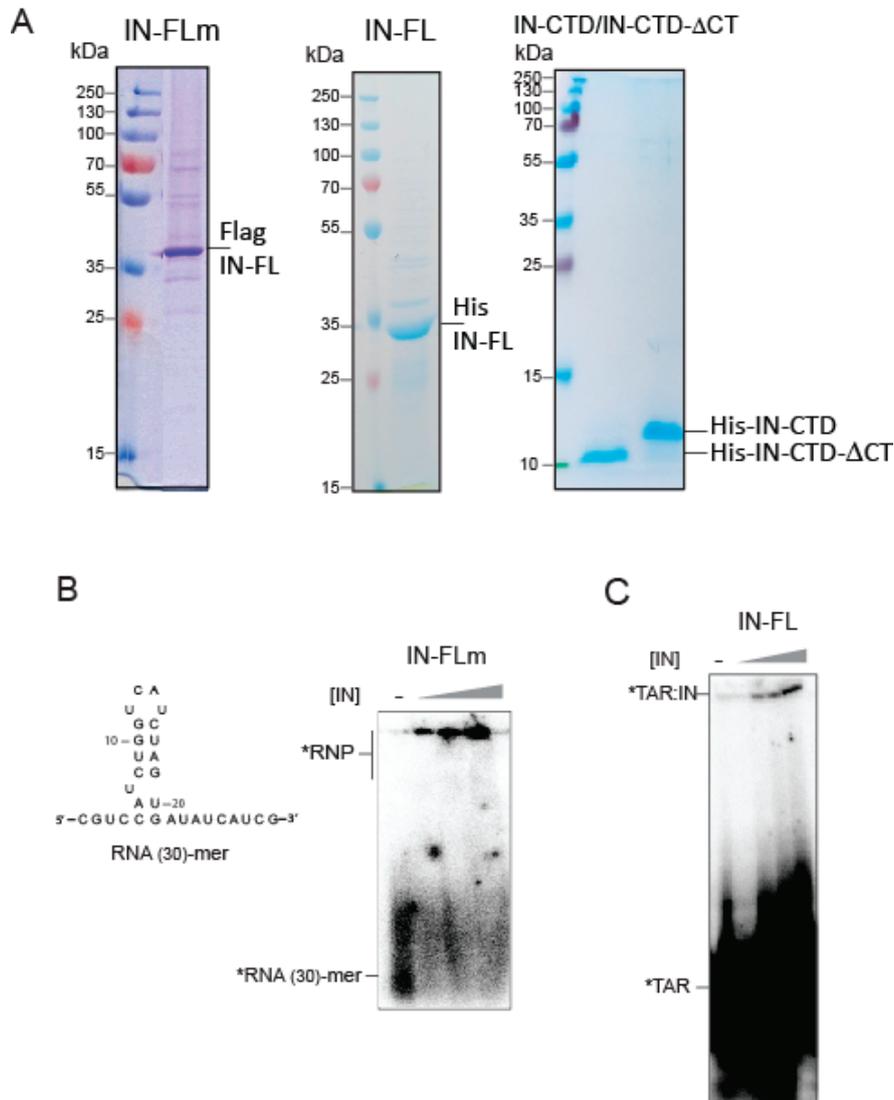
<sup>3</sup> Department of Biochemical Sciences, Sapienza University of Rome, P.le Aldo Moro 5, 00185 Rome, Italy.

<sup>4</sup> Chromatin Stability and DNA Mobility, Department of Integrated Structural Biology, IGBMC, CNRS UMR 7104 – Inserm U 158 University of Strasbourg, 1 rue Laurent Fries, 67404, Illkirch Cedex, France.

<sup>5</sup> RNA architecture and reactivity, IBMC, CNRS UPR 9002 University of Strasbourg, 2, Allée Konrad Roentgen, 67084 Strasbourg Cedex, France.

### **Supplementary Material**

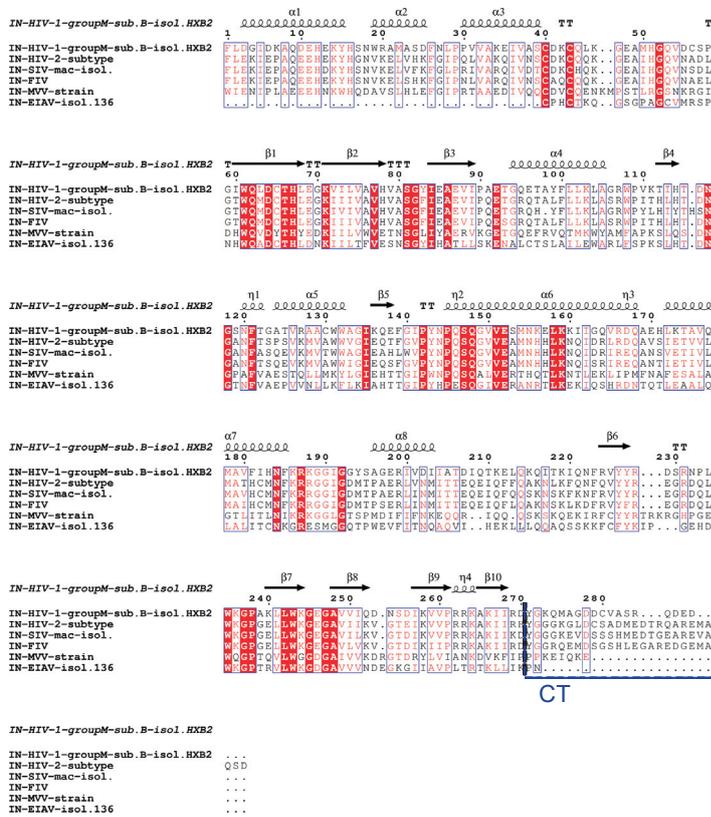
## Supplementary Figure 1



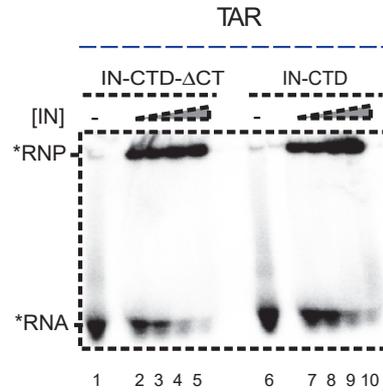
**Supplementary Figure S1. A)** SDS-PAGE illustrating flag-tagged IN-FL expressed from mammalian expression system, IN-FL, IN-CTD and IN-CTD- $\Delta$ CT from *E. coli* used in this study. **B)** Structural model (left panel) of a weakly structured RNA<sub>(30)</sub>-mer used for EMSA assay (right panel) illustrating the interaction of IN-FLm. The RNA substrate was incubated with increasing concentrations of proteins (0; 100; 200; 400  $\mu$ M) for 30 minutes at 37°C in binding buffer as indicated in experimental procedures. **C)** EMSA assay indicating the binding of IN-FL produced in *E.coli* to TAR substrate.

## Supplementary Figure 2

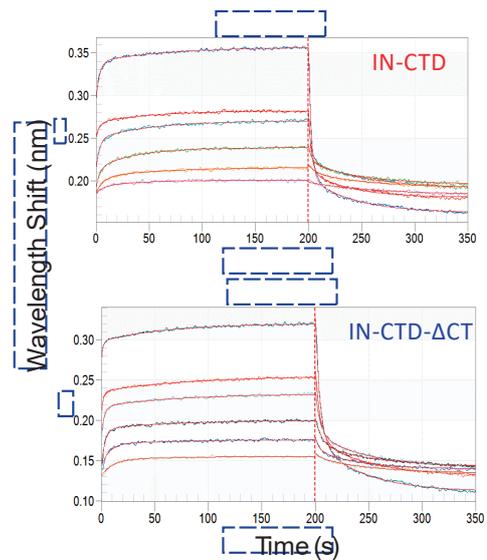
**A**



**B**



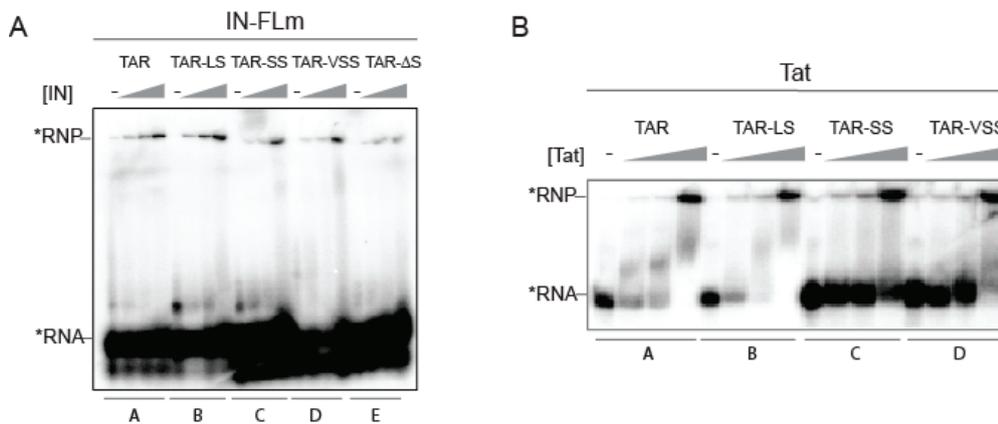
**C**



**Supplementary Figure S2. A)** Sequence alignment of Lentivirus IN proteins. Amino acid sequences were obtained from the HIV database compendium (<http://www.hiv.lanl.gov/>), aligned using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) and analysed by ESPript 3.0 Web server (65). Secondary structure elements were presented on top of the alignment (helices with squiggles, strands with arrows and turns with TT letters). Red shading indicates sequence identity and boxes indicate sequence similarity, according to physico-chemical properties. HIV - Human Immunodeficiency virus, SIV - *Simian* immunodeficiency virus, FIV - feline immunodeficiency virus, MVV - Maedi visna virus, EIAV - equine infectious anemia virus. **B)** EMSA assay illustrating the interaction of IN-CTD and IN-CTD- $\Delta$ CT with TAR

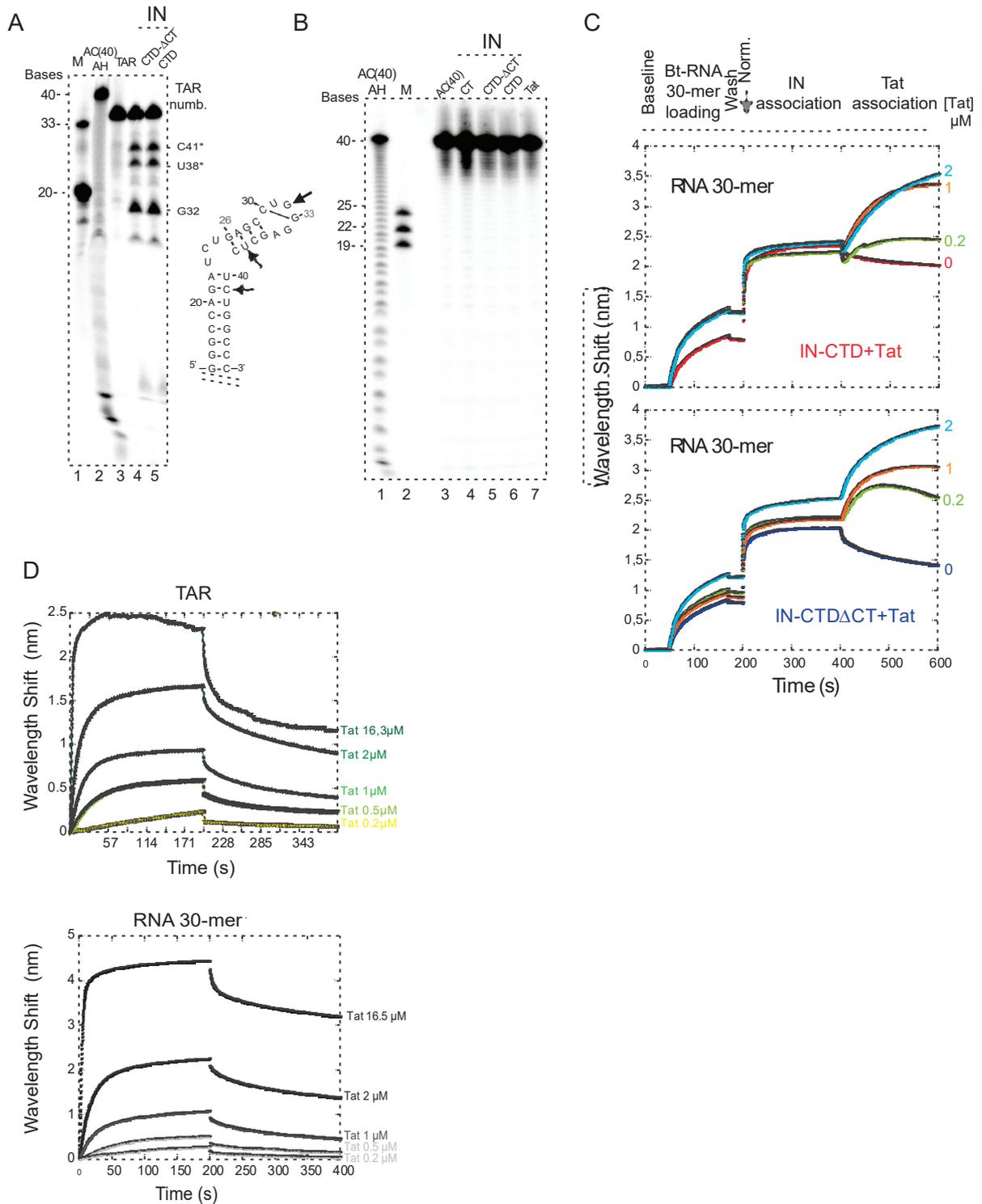
RNA. The RNA substrates (50 nM) are labelled with  $^{32}\text{P}$  (black star) and incubated with increasing concentrations of proteins (0; 100, 200, 400 and 800 nM) under the conditions described in 'Materials and Methods' section. **C)** Real-time sensorgrams of the BLI experiment relative to experiment shown in Figure 1F, showing the interaction of IN-CTD and IN-CTD- $\Delta\text{CT}$  with 5'-biotinylated TAR RNA. Association and dissociation curves are obtained for different protein concentrations 1.8, 2.7, 5.4, 8.1, 16.2 and 32.5  $\mu\text{M}$ , at 37°C in a buffer containing 200 mM NaCl.

### Supplementary Figure 3



**Supplementary Figure S3. A)** EMSA assay illustrating the binding of IN-FLm to TAR wild type and mutants. Increasing concentrations of IN-FLm were added to 5'-end radiolabelled TAR RNA and treated as indicated in 'Materials and Methods' paragraph. **B)** EMSA assay showing the interaction of Tat protein (0, 100, 200, 400 nM) with TAR mutants (50 nM) showed in Figure 2A. **C)** Graph showing the fractions of RNA bound by Tat as a function of Tat concentration. Detection, quantification, and data analysis were carried out as described in Figure 2C.

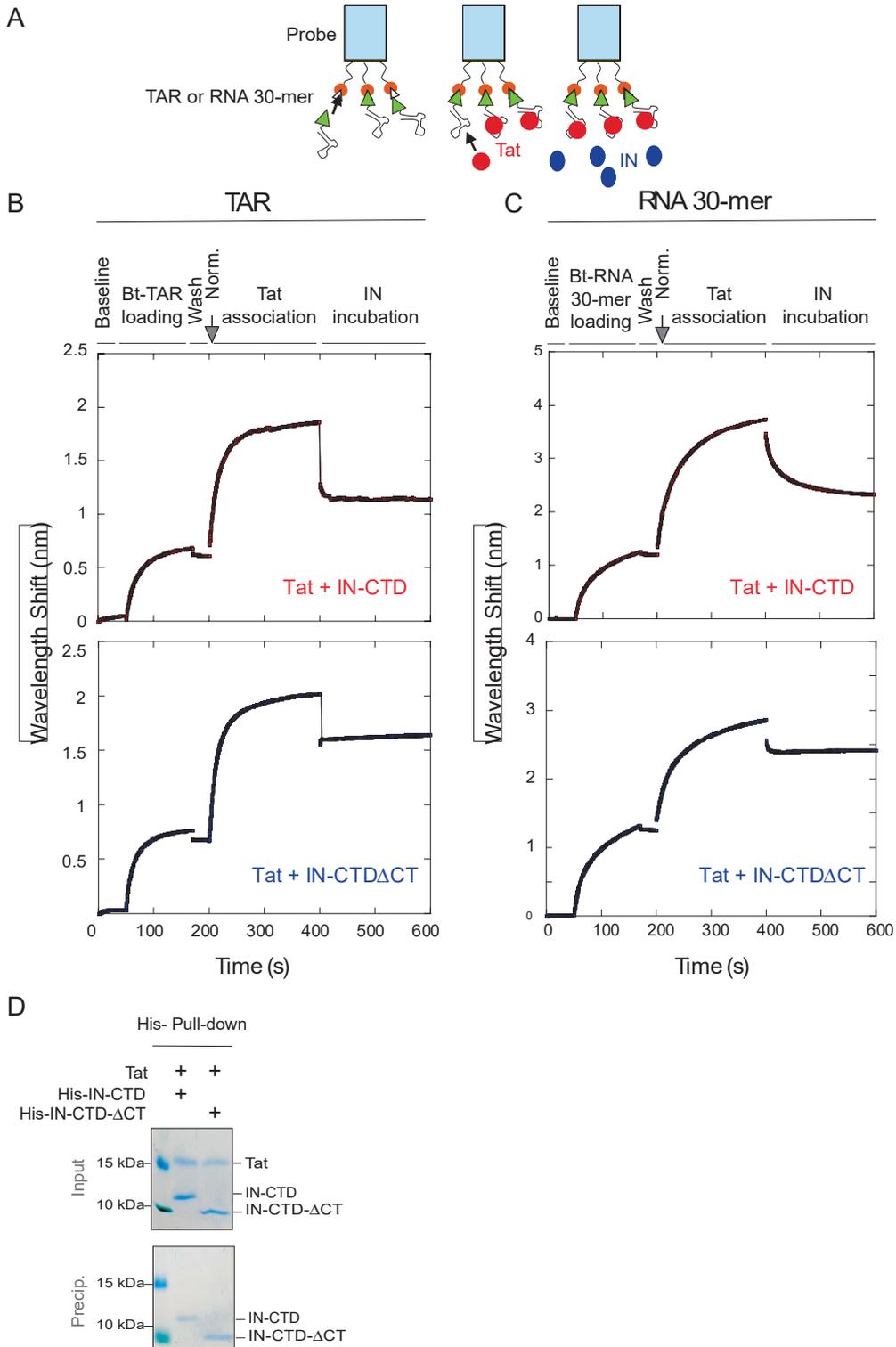
## Supplementary Figure 4



**Supplementary Figure S4. A)** RNA-denaturing gel showing the effect of IN-CTD and IN-CTD- $\Delta$ CT on TAR structure in the absence of RNase T1. Radiolabelled TAR RNA was incubated with IN truncations in the

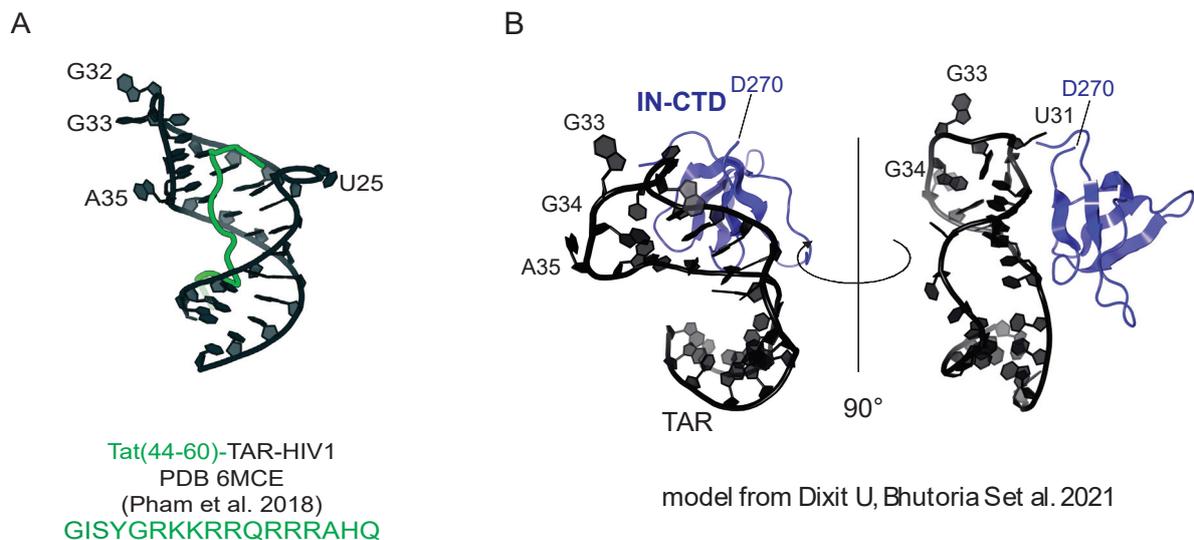
same conditions used in Figure 3A. Bands corresponding to spontaneous cleavage in the presence IN-CTD and IN-CTD- $\Delta$ CT are identified as position markers. Gel lanes are as follows: (M) Ladder of two RNA transcripts of 33 and 20 nucleotides in length (lane 1); (AC<sub>(40)</sub>-AH) alkaline ladder of AC<sub>(40)</sub>-mer RNA (lane 2); TAR RNA; TAR RNA incubated with IN-CTD- $\Delta$ CT (lane 4) and IN-CTD (lane 5). Digestion patterns were mapped on TAR secondary structure depicted on the right of the gel by arrows. **B**) Denaturing RNA gel showing the absence of nucleases in protein samples used for experiments of Figure 3A and Supplementary Figure 4A. The linear AC<sub>(40)</sub>-mer RNA was incubated with indicated protein preparations in the same conditions used for probing experiments: (AC<sub>(40)</sub>-AH) alkaline ladder of AC<sub>(40)</sub>-mer RNA (lane 1); (M) Ladder of three RNA transcripts of 25, 22 and 19 nucleotides in length (lane 2); (AC<sub>(40)</sub>) native AC<sub>(40)</sub>-mer RNA; TAR RNA incubated with IN-CT (lane 4), IN-CTD- $\Delta$ CT (lane 5); IN-CTD (lane 5) and Tat (lane 6). **C**) Interferometry sensorgrams of protein-RNA interaction used to calculate the equilibrium dissociation constants of Figure 3E. Bt-RNA<sub>(30)</sub>-mer was first loaded on the streptavidin-coated biosensor for 120 s (Bt-RNA<sub>(30)</sub>-mer loading) then the unbound RNA was washed for 30 s (wash). The sensor was dipped in a solution containing 90  $\mu$ M of IN protein for 200 s then incubated with different concentrations of Tat (0.2, 1 and 2  $\mu$ M) for 200 s. **D**) Interferometry sensorgrams to measure the affinity of Tat for immobilized Bt-TAR RNA (upper panel) and Bt-RNA<sub>(30)</sub>-mer (lower panel). The graph shows the real-time association and dissociation kinetics at different Tat concentrations.

## Supplementary Figure 5



**Supplementary Figure S5.** **A)** Schematic illustration of the steps performed during real-time binding experiments by BLI. First, we immobilized Bt-TAR or Bt-RNA<sub>(30)</sub>-mer to streptavidin-coated biosensor and washed out the unbound RNA. Then we soaked the probe in a solution containing Tat. Afterwards, the Tat:RNA bound-biosensor was dipped in a solution containing either IN-CTD or IN-CTD- $\Delta$ CT. **B)** Sensorgram showing the association of Tat (32.2  $\mu$ M) to Bt-TAR (first rising curve between 200 and 400 s). Then the sensor was dipped in a solution containing 32.2  $\mu$ M of either IN-CTD or IN-CTD- $\Delta$ CT for 200 s (curve between 400 and 600 s). **C)** Sensorgrams to measure the association of Tat (2  $\mu$ M) to Bt-RNA<sub>(30)</sub>-mer (first rising curve between 200 and 400 s). Then the sensor was dipped in a solution containing 90  $\mu$ M of either IN-CTD or IN-CTD- $\Delta$ CT for 200 s (curve between 400 and 600 s). **D)** Histidine pull-down assay using His-tagged IN-CT and IN-CTD- $\Delta$ CT. Tagged IN truncations were mixed with Tat and incubated in a buffer containing 200 mM NaCl before co-precipitation. Input (20% of total) and pull-down fractions were analysed by 15% SDS-PAGE followed by Coomassie blue staining.

## Supplementary Figure 6



**Supplementary Figure S6.** **A)** Solution NMR structure of TAR (black) bound to Tat RBD (green) deposited with PDB code: 6MCE from (Pham et al. 2018). **B)** Structural model of IN-CTD- $\Delta$ CT/TAR complex from (Dixit et al. 2021).

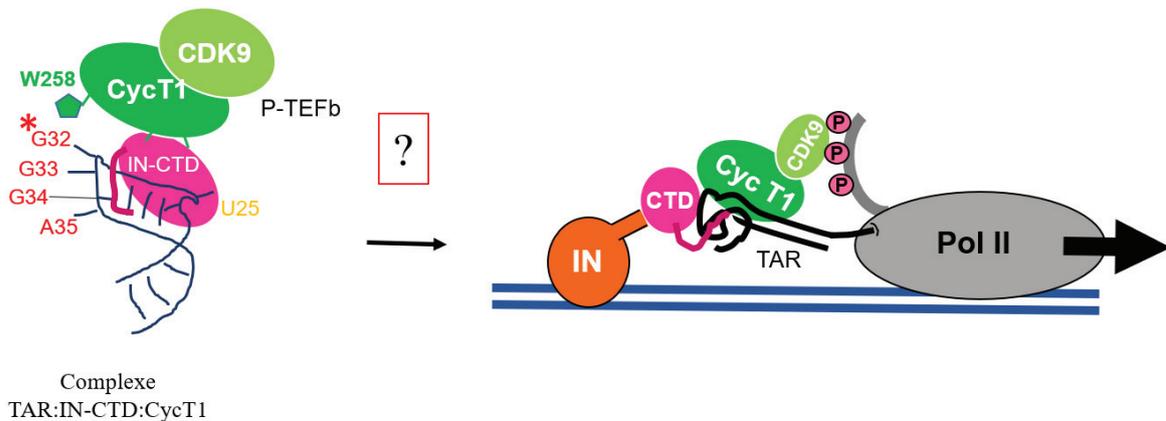




## VI.1. Etude de l'interaction entre IN-CTD, le complexe P-TEFb et l'ARN TAR

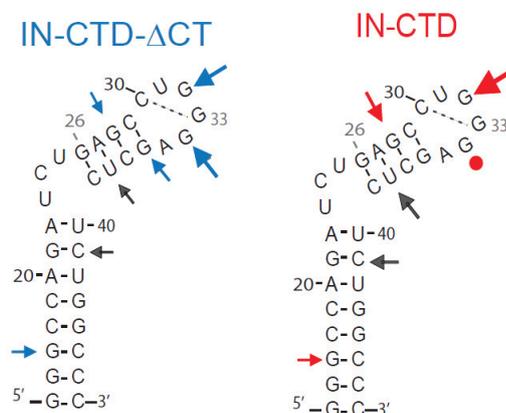
Une question historique est encore sans réponse aujourd'hui : comment Tat, qui n'est ni internalisée dans les virions ni exprimée avant la transcription provirale, peut recruter le complexe P-TEFb et induire l'élongation ? La théorie la plus raisonnable est l'existence d'une transcription basale qui permet de synthétiser Tat (Rampalli *et al.*, 2003). Deux questions se posent aujourd'hui suite à la découverte de l'implication d'IN-CTD dans la transcription provirale (Winans et Goff, 2020) et suite à nos travaux : est-ce qu'en absence de Tat, le CTD est capable de recruter le complexe d'élongation de la transcription ? Et dans ce cas, est-ce que le complexe recruté est actif et permet d'activer l'élongation de la transcription provirale ? (**Figure 56**)

J'ai essayé de répondre à ces questions, et pour cela j'ai produit les protéines recombinantes du complexe d'élongation, la cycline T1 et la CDK9, pour étudier leur mécanisme d'interaction avec l'ARN TAR et IN-CTD.



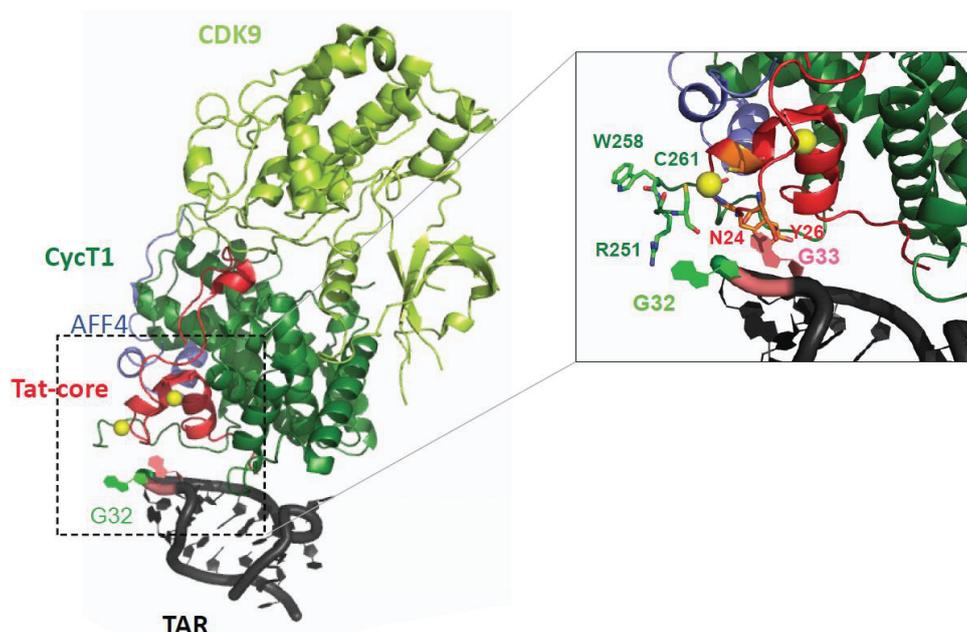
**Figure 56.** Représentations des problématiques soulevées par l'article de recherche Rocchi, Louvat *et al.*, 2022.

Nos précédents travaux sur le mécanisme d'interaction d'IN-CTD avec l'ARN TAR ont montré que la liaison d'IN-CTD modifie la structure de l'ARN TAR en exposant la G32 (Rocchi, Louvat *et al.*, 2022) (**figure 57**). Or, ce nucléotide est impliqué dans l'interaction entre l'ARN TAR et le TRM de la cycline T1 (Schulze-Gahmen *et al.*, 2018) (**figure 58**). La base du nucléotide G32 établit des contacts avec le résidu Trp258 ; les résidus Arg251, Arg254 et Arg259 interagissent avec le squelette phosphate-ribose des nucléotides G32 et G33 ; et le résidu Cys261 coordonne l'ion  $Zn^{2+}$  requis pour l'interaction avec Tat.



**Figure 57.** Schéma illustrant les modifications de la structure de l'ARN TAR suite à la liaison d'IN-CTD- $\Delta$ CT ou IN-CTD par une expérience d'empreinte à l'ARN (Rocchi, Louvat *et al.*, 2022). Le point

correspond au nucléotide protégé de la digestion par la RNase T1 et les flèches représentent les sites de cassure de l'ARN.

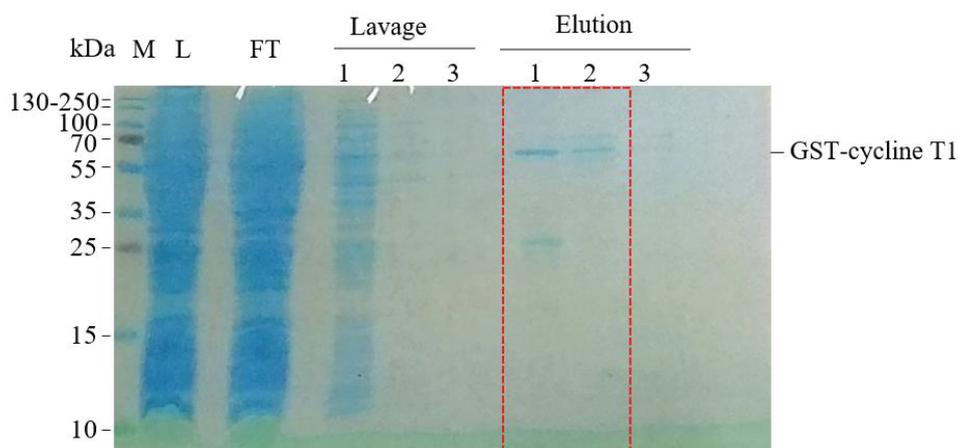


**Figure 58.** Structure du complexe TAR:Tat:Cycline T1:CDK9 obtenu par Schulze-Gahmen et al., 2018 (code PDB : 6CYT).

Dans le contexte de la transcription provirale, nous avons essayé de comprendre les mécanismes d'interaction entre le complexe d'élongation de la transcription comprenant la cycline T1 et la CDK9 avec IN-CTD et l'ARN TAR. Dans ce but, j'ai produit les protéines recombinantes comme décrit dans le chapitre Matériel et Méthodes (voir **chapitres XII.1.2, XII.1.3 et XII.1.4**).

#### VI.1.1. Purification de la protéine recombinante GST-cycline T1

La purification de GST-cycline T1 se fait par une étape de chromatographie d'affinité en « batch ». Les fractions d'élué sont contrôlées par une électrophorèse sur gel de polyacrylamide en condition dénaturante. La protéine GST-cycline T1 a une masse moléculaire de 61 kDa. Ainsi, en regardant le gel de polyacrylamide où ont été chargés les fractions d'élué (**figure 59**), on peut voir que la protéine est éluée dans les deux premières fractions d'élué. Ces fractions d'élué ont été conservées pour des expériences fonctionnelles.



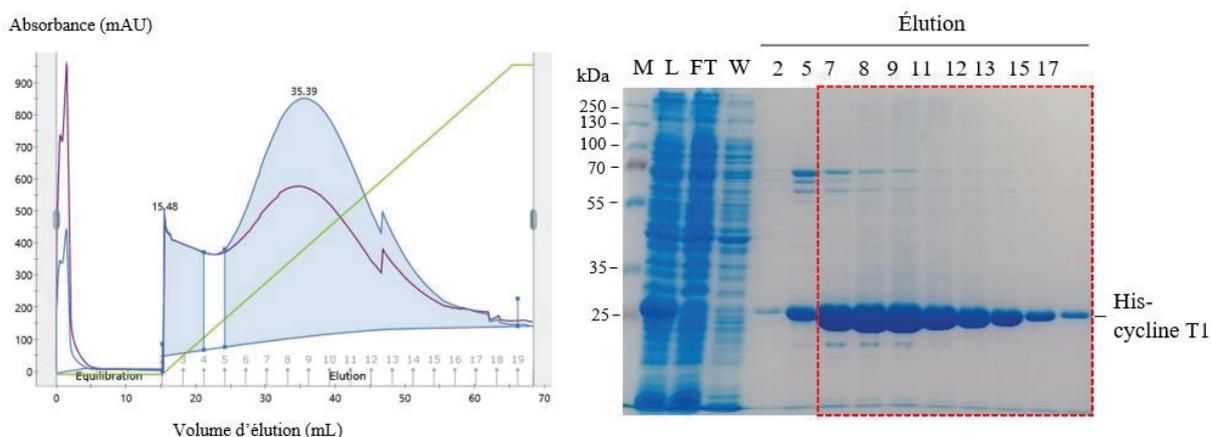
**Figure 59.** SDS-PAGE (« sodium dodecyl sulfate-polyacrylamide gel electrophoresis ») 15% sur les fractions de la purification de GST-cycline T1 avec une résine Gluthatione Sepharose 4B (M : marqueur de masses moléculaires ; L : lysat ; FT : fraction non retenue). Les fractions encadrées en rouge sont celles qui ont été conservées pour la suite des manipulations. Les numéros des fractions de lavage et d'élution contrôlés sont précisés sur le gel.

### VI.1.1. Purification de la protéine recombinante his-cycline T1

Afin d'étudier plus en détail les mécanismes d'interaction impliquant IN-CTD pendant les premières étapes de la transcription provirale, la cycline T1 a été clonée dans un plasmide avec une étiquette histidine pour avoir une protéine plus proche du contexte physiologique et pouvoir étudier son interaction avec IN-CTD en présence de CDK9.

La purification de his-cycline T1 se fait par une étape de chromatographie IMAC (« immobilized metal ion affinity chromatography »). Un pic d'élution est observé à 220 mM imidazole sur le chromatogramme (**figure 60**, à gauche). Une analyse par SDS-PAGE est réalisée pour confirmer l'identité des espèces séparées lors de la chromatographie. La protéine recombinante his-cycline T1 a une masse moléculaire de 32 kDa. Ainsi, en regardant le gel de polyacrylamide où ont été chargés les échantillons, on peut déterminer que la protéine d'intérêt a été surproduite (fraction du lysat).

La protéine surexprimée présente une migration aberrante d'après le marqueur de masse moléculaire (**figure 60**, à droite). En effet, au lieu de migrer comme une protéine de 32 kDa, elle migre comme une protéine de 25 kDa. Des contaminants sont également élués avec la protéine d'intérêt au début du pic d'élution. Les fractions encadrées en rouge, correspondant au milieu et à la fin du pic, ont été conservées pour des expériences fonctionnelles.

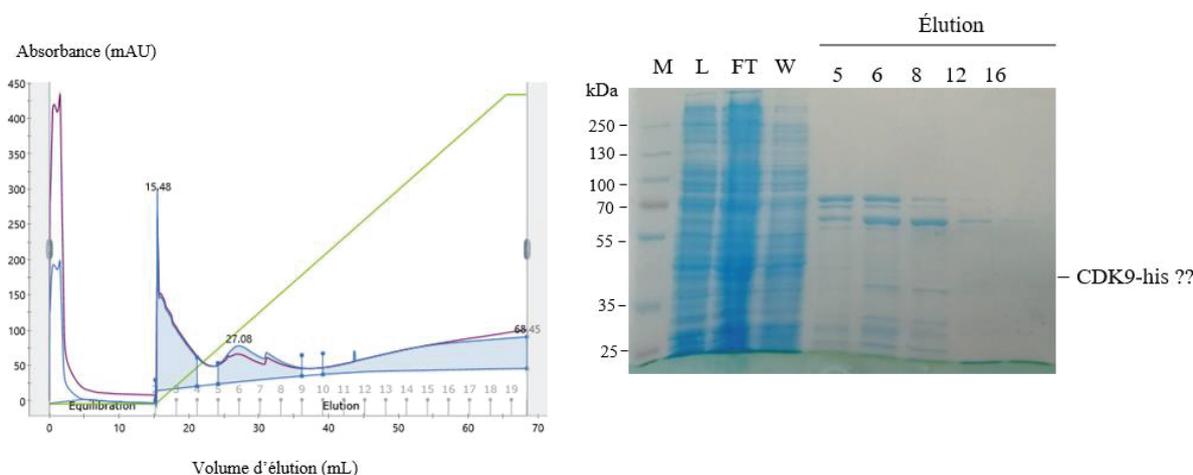


**Figure 60.** *A gauche*, chromatogramme de la purification de his-cycline T1 sur une colonne HisTrap crude 5 mL. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'élution en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et celle en violet à la mesure de l'absorbance à 260 nm. Le tracé vert correspond au pourcentage de tampon d'élution. *A droite*, SDS-PAGE 12% sur les fractions de la purification de his-cycline T1 sur colonne HisTrap (M : marqueur de masses moléculaires ; L : lysat ; FT : fraction non retenue ; W : fraction de lavage). Les fractions encadrées en rouge sont celles qui ont été conservées pour la suite de la purification. Les numéros des fractions d'élution contrôlées sont précisés sur le gel.

### VI.1.2. Expression et purification de la protéine recombinante CDK9-his

En parallèle à la production de la protéine recombinante his-cycline T1, j'ai essayé de produire la protéine recombinante CDK9-his. La purification de CDK9-his se fait de la même manière que pour la

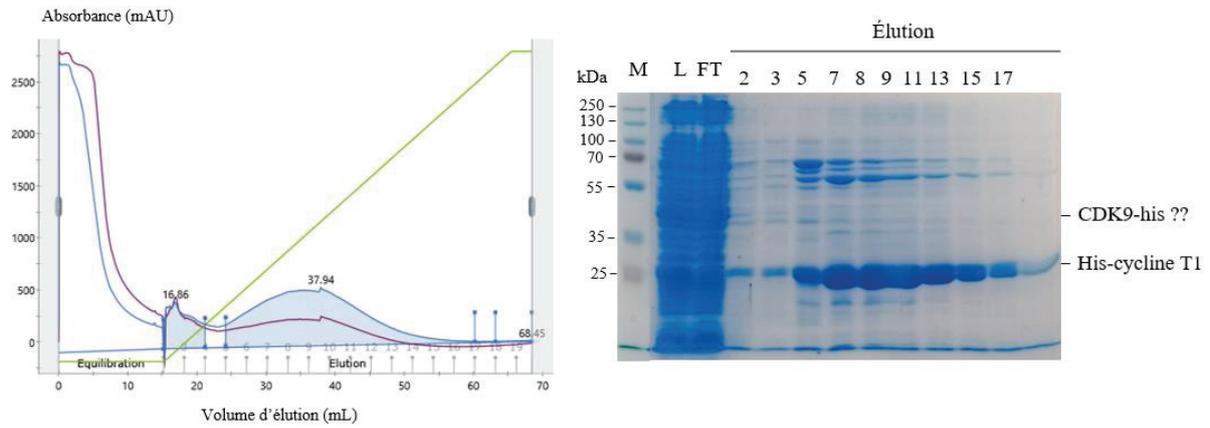
protéine his-cycline T1. Plusieurs pics d'élution sont observés sur le chromatogramme (**figure 61**, à gauche). Un premier pic est observé au commencement du gradient d'élution puis un second pic est observé à 130 mM imidazole. CDK9-his a une masse moléculaire de 41 kDa. Ainsi, en regardant le gel de polyacrylamide où ont été chargés les échantillons (**figure 61**, à droite), notamment le lysat et les fractions d'élution, on peut voir qu'aucune bande ne correspond à une protéine surexprimée à cette taille.



**Figure 61.** *A gauche*, chromatogramme de la purification de CDK9-his sur une colonne HisTrap crude 5 mL. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'élution en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et celle en violet à la mesure de l'absorbance à 260 nm. Le tracé vert correspond au pourcentage de tampon d'élution. *A droite*, SDS-PAGE 12% sur les fractions de la purification de CDK9-his sur colonne HisTrap (M : marqueur de masses moléculaires ; L : lysat ; FT : fraction non retenue ; W : fraction de lavage). Les numéros des fractions d'élution contrôlées sont précisés sur le gel.

Afin de réussir à purifier la protéine d'intérêt, une co-lyse de culots bactériens ayant surexprimés séparément his-cycline et CDK9-his est réalisée en condition native. Par la suite une purification IMAC est réalisée.

Sur le chromatogramme (**figure 62**, à gauche), un pic majoritaire d'élution est observé à 220 mM imidazole, ce qui correspond à la concentration d'imidazole à laquelle est éluée his-cycline T1 d'après les résultats précédents. En effet, en regardant le gel de polyacrylamide où ont été chargés les fractions d'élution (**figure 62**, à droite), on peut voir lorsqu'on contrôle ce pic que la protéine éluee correspond à his-cycline T1. Cependant, on n'observe pas de bande correspondante à la protéine CDK9-his dans les fractions d'élution. Une bande intense en comparaison aux autres bandes dans la fraction du lysat à une masse moléculaire de 41 kDa pourrait correspondre à CDK9-his surexprimée. Ainsi, une hypothèse serait que CDK9-his est présent dans les corps d'inclusion de la bactérie et n'est pas solubilisée dans nos conditions natives.

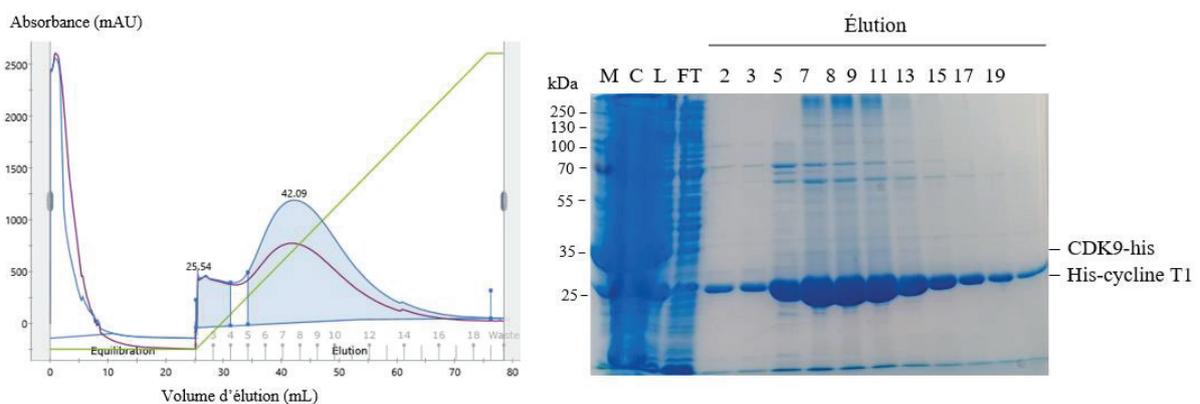


**Figure 62.** *A gauche*, chromatogramme de la purification d'un co-lysate en condition native de culots bactériens exprimant his-cycline T1 et CDK9-his sur une colonne HisTrap crude 5 mL. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'élution en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et celle en violet à la mesure de l'absorbance à 260 nm. Le tracé vert correspond au pourcentage de tampon d'élution. *A droite*, SDS-PAGE 12% sur les fractions de la purification d'un co-lysate en condition native de culots bactériens exprimant his-cycline T1 et CDK9-his sur colonne HisTrap (M : marqueur de masses moléculaires ; L : lysate ; FT : fraction non retenue). Les numéros des fractions d'élution contrôlées sont précisés sur le gel.

Une expérience de co-lyse à partir de culots bactériens ayant surexprimé séparément his-cycline et CDK9-his est ensuite répétée mais en condition dénaturante.

Sur le chromatogramme de purification (**figure 63**, à gauche), un pic majoritaire d'élution est observé à 200 mM imidazole, ce qui correspond à une concentration d'imidazole légèrement plus faible que celle pour laquelle est éluée his-cycline T1 d'après les résultats précédents. En effet, en regardant le gel de polyacrylamide où ont été chargés les fractions d'élution (**figure 63**, à droite), on peut voir lorsqu'on contrôle ce pic que la protéine éluée est bien his-cycline T1. Cependant, à nouveau on n'observe pas de bande correspondante à la protéine CDK9-his dans les fractions d'élution. Une bande plus intense à une masse moléculaire de 41 kDa est de nouveau observé.

Ainsi, une partie de CDK9-his est présent dans le lysate bactérien dans des conditions de co-lyse en condition dénaturante, mais la protéine n'est pas retrouvée dans les fractions d'élution ce qui était également le cas lors de la co-lyse en condition native (**figure 62**, à droite). Une autre partie de CDK9-his est présente dans les corps d'inclusion de la bactérie, car observée dans le culot, et n'est pas solubilisée dans nos conditions dénaturantes. Pour la suite, puisque je n'arrive pas à extraire complètement CDK9-his des corps d'inclusion et à la purifier, j'ai décidé d'optimiser les conditions d'expression comme décrit par Leopoldino *et al.* (2006) en rallongeant les temps d'incubation avec des températures plus basses et une concentration d'inducteur moindre.



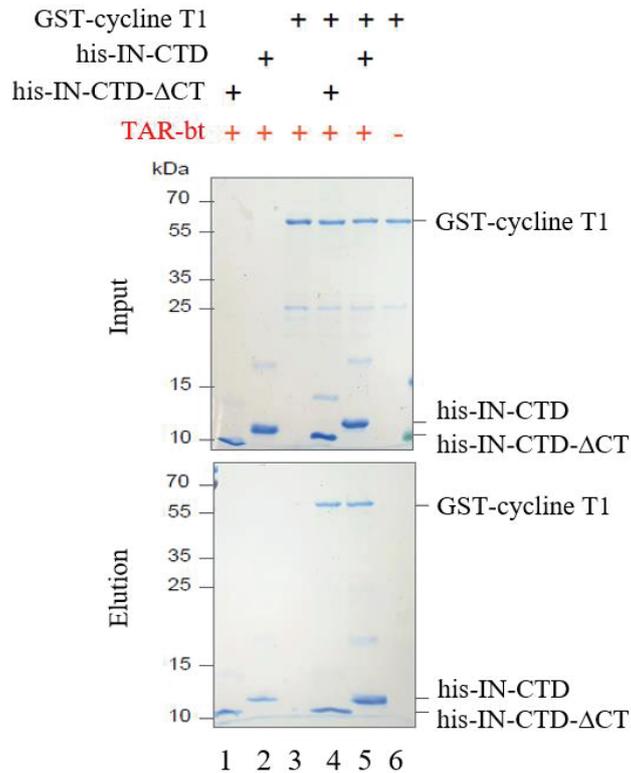
**Figure 63.** *A gauche*, chromatogramme de la purification d'un co-lysate en condition dénaturante de culots bactériens exprimant his-cycline T1 et CDK9-his sur une colonne HisTrap crude 5 mL. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'élution en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et celle en violet à la mesure de l'absorbance à 260 nm. Le tracé vert correspond au pourcentage de tampon d'élution. *A droite*, SDS-PAGE 12% sur les fractions de la purification d'un co-lysate en condition dénaturante de culots bactériens exprimant his-cycline T1 et CDK9-his sur colonne HisTrap (M : marqueur de masses moléculaires ; C : culot ; L : lysat ; FT : fraction non retenue). Les numéros des fractions d'élution contrôlées sont précisés sur le gel.

### VI.1.3. Etude de l'interaction entre IN-CTD, la cycline T1 et l'ARN TAR-bt par co-précipitation

Dans le contexte de la transcription provirale, nous avons essayé de comprendre les mécanismes d'interaction entre le complexe d'élongation de la transcription comprenant la cycline T1 et la CDK9 avec IN-CTD et l'ARN TAR. Dans ce but, j'ai réalisé une co-précipitation de l'ARN TAR avec ces protéines en utilisant des billes streptavidine comme décrit dans le chapitre Matériel et Méthodes (voir **chapitre XII.5.2**).

Le gel « *input* » permet de vérifier que toutes les protéines sont présentes dans les bonnes concentrations (**figure 64**). Ensuite, en comparant les bandes entre le gel « *input* » et le gel d'élution, nous voyons que his-IN-CTD- $\Delta$ CT ou his-IN-CTD interagit bien avec l'ARN TAR-bt (puits 1 et 2). Ce n'est pas le cas de GST-cycline T1 qui n'est pas présente dans la fraction éluee (puits 3). Cependant, en présence de his-IN-CTD- $\Delta$ CT ou his-IN-CTD, GST-cycline T1 est présente sur le gel d'élution et interagit avec l'ARN TAR (puits 4 et 5). Ainsi, la cycline T1 ne lie pas l'ARN TAR mais interagit avec les complexes IN-CTD- $\Delta$ CT:TAR ou IN-CTD:TAR. Le CT d'IN-CTD ne semble pas être nécessaire pour cette interaction, puisqu'il n'y a pas de différence significative observée entre IN-CTD- $\Delta$ CT et IN-CTD (puits 4 et 5).

Le dernier puits est un contrôle négatif, qui permet de s'assurer que les protéines n'interagissent pas spécifiquement avec les billes streptavidine (puits 6). Pour éviter les interactions aspécifiques, des tubes spécifiques Protein LoBind® (Eppendorf) sont utilisés pour les réactions et les étapes de lavage sont réalisées avec un tampon de lavage contenant 10 mM d'imidazole.

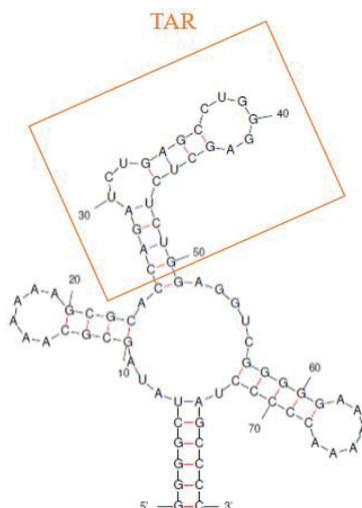


**Figure 64.** Expérience de co-précipitation avec l'ARN TAR-bt. **Gel « Input ».** L'ARN biotinyllé est incubé avec les protéines seules : his-IN-CTD-ΔCT (puits 1), his-IN-CTD (puits 2), GST-cycline T1 (puits 3) ; ou ensemble : his-IN-CTD-ΔCT + GST-cycline T1 (puits 4) et his-IN-CTD + GST-cycline T1 (puits 5). Un contrôle négatif sans ARN est réalisé en incubant les trois protéines étudiées (puits 6). Ces échantillons sont contrôlés par SDS-PAGE 15%. **Gel Elution.** Des billes streptavidine sont ajoutées et après plusieurs lavages, les protéines interagissant avec TAR-bt sont éluées et contrôlées par SDS-PAGE 15%. Les numéros des puits sont précisés en bas du gel d'éluion.

## VII. Partie structurale

Aucune structure d'IN avec de l'ARN n'a été résolue à ce jour. Afin de mieux caractériser la formation de ce complexe au cours de la morphogénèse et de la transcription provirale, des études structurales sur IN:ARN sont menées dans l'équipe.

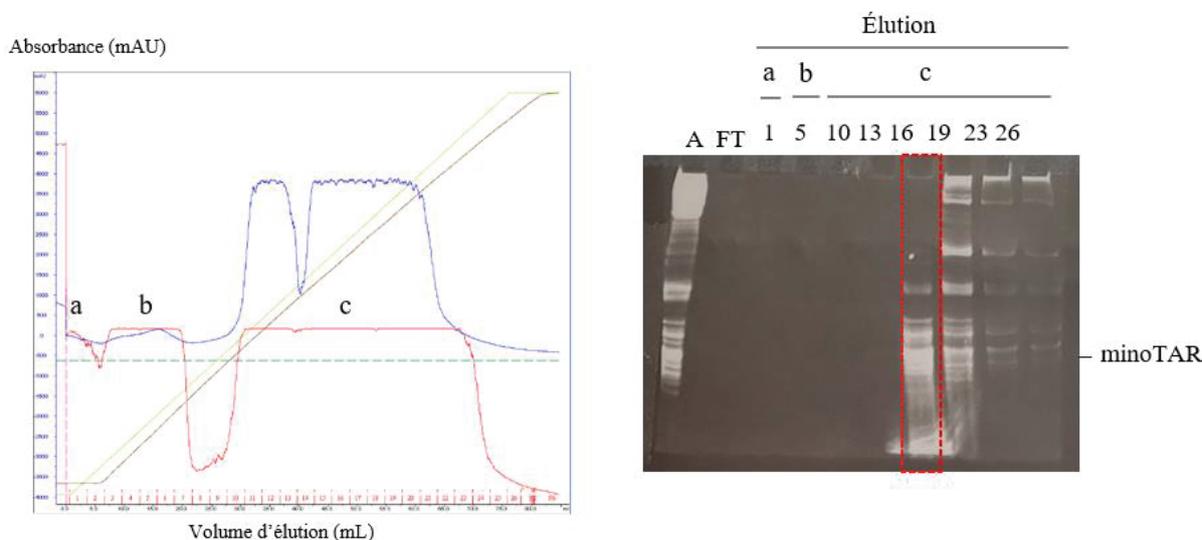
Comme expliqué précédemment, l'IN est une protéine modulaire ce qui rend difficile la cristallisation de la protéine entière. Pour cela, j'ai travaillé uniquement sur le domaine CTD impliqué dans l'interaction avec l'ARN. Toutefois, des études précédentes menées au laboratoire avec un fragment de l'ARN TAR n'avaient pas permis d'obtenir des cristaux. J'ai donc construit une chimère de l'ARN TAR sur un échafaudage d'ARNt (**figure 65**), que j'ai ensuite produite en système bactérien comme décrit dans le chapitre Matériel et Méthodes (voir **chapitres XI.6.3** et **XII.3.2**). Cette méthode permet d'obtenir une grande quantité homogène d'ARN (Ponchon et Dardel, 2007). De plus, l'échafaudage d'ARNt permet de protéger l'ARN de l'action des nucléases.



**Figure 65.** Modèle des structures secondaires de l'ARN chimérique minoTAR obtenu avec RNAfold web server avec  $\Delta G = -37,74$  kcal/mol.

### VII.1. Purification de l'ARN chimérique minoTAR

La purification de cet ARN chimérique, que nous avons appelé minoTAR, se fait par une première étape de chromatographie sur une colonne échangeuse d'anions. Sur le chromatogramme (**figure 66**, à gauche), trois pics d'éluion sont observés à 260 nm : le premier est au commencement du gradient d'éluion, le deuxième à 220 mM NaCl, et le troisième, plus large, est à 530 mM NaCl. Pour confirmer l'identité des espèces séparées lors de la chromatographie, une électrophorèse d'ARN sur gel de polyacrylamide en condition dénaturante est réalisée. L'ARN minoTAR a une taille de 79 nucléotides. En regardant le gel de polyacrylamide où ont été chargés les fractions d'éluion (**figure 66**, à droite), on peut voir que les deux premiers pics correspondent à l'éluion de nucléotides ou d'ARN trop petits qui ont migré hors du gel. Les fractions d'éluion 14 à 18 du troisième pic correspondent à l'éluion de l'ARN minoTAR avec le moins de contaminants, tandis que les fractions d'éluion de la fin du troisième pic (fractions 19 à 26) contiennent essentiellement des contaminants. Les fractions d'éluion de l'ARN minoTAR avec le moins de contaminants sont donc conservées et purifiées par la suite par une chromatographie SEC (« *size-exclusion chromatography* »).



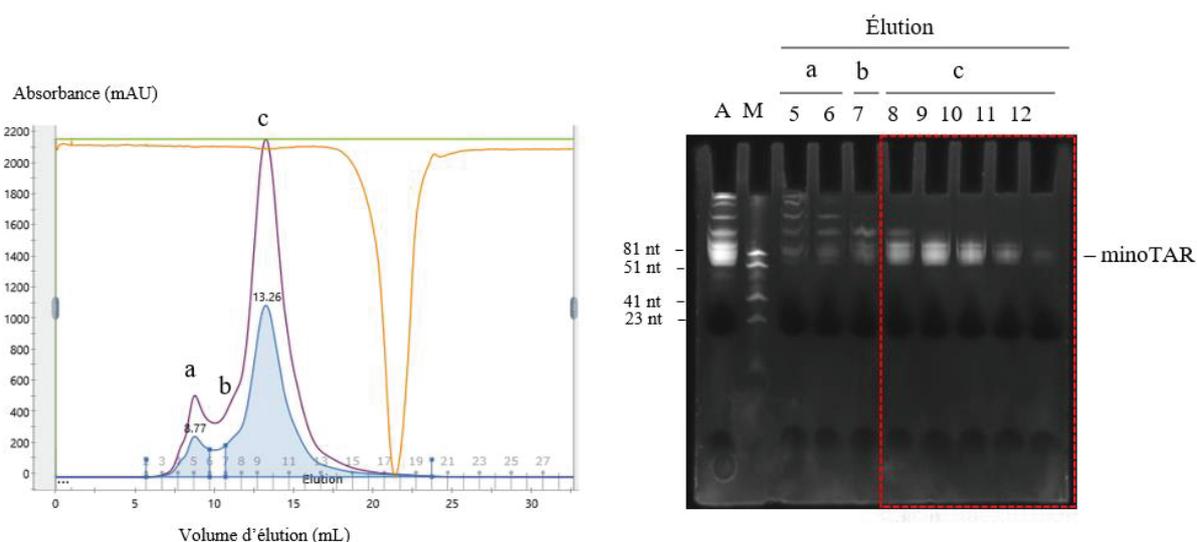
**Figure 66.** A gauche, chromatogramme de la purification de l'ARN chimérique minoTAR sur une colonne HiTrapQ-XL 5 mL. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du

volume d'éluion en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et celle en rouge à la mesure de l'absorbance à 260 nm. Le tracé vert correspond au pourcentage de tampon d'éluion et le tracé marron à la conductivité. **A droite**, électrophorèse en condition dénaturante sur un gel de polyacrylamide 10% à partir des fractions de la purification de l'ARN minoTAR sur colonne HiTrapQ-XL (A : échantillon avant l'injection sur la colonne de purification ; FT : fraction non retenue). Les numéros des fractions d'éluion contrôlées sont précisés sur le gel.

La seconde étape de purification de l'ARN minoTAR est réalisée sur une colonne Superdex 75 Increase 10/300 GL. Trois pics sont observés sur le chromatogramme (**figure 67**, à gauche) à 8,77 mL, 11,63 mL (un léger épaulement) et 13,26 mL. Ces volumes d'éluion correspondent à des masses moléculaires respectives de 108 kDa, 49 kDa et 31 kDa d'après la calibration de la colonne Superdex 75 Increase 10/300 GL (**Annexes, figure 93**).

Afin de mieux caractériser ces trois pics, une électrophorèse d'ARN dans un gel de polyacrylamide en condition dénaturante est réalisée (**figure 67**, à droite). Le puits 1 contenant la fraction avant la chromatographie SEC est un témoin positif qui confirme la présence de l'ARN minoTAR avant la purification. Le premier pic d'éluion et le deuxième correspondent à l'éluion majoritaire de contaminants avec un peu d'ARN d'intérêt, tandis que l'ARN d'intérêt est élué majoritairement dans le troisième pic d'éluion avec quelques contaminants.

Ce résultat est cohérent avec la masse moléculaire de 26 kDa de l'ARN minoTAR, et le troisième pic d'éluion est conservé pour des expériences structurales.

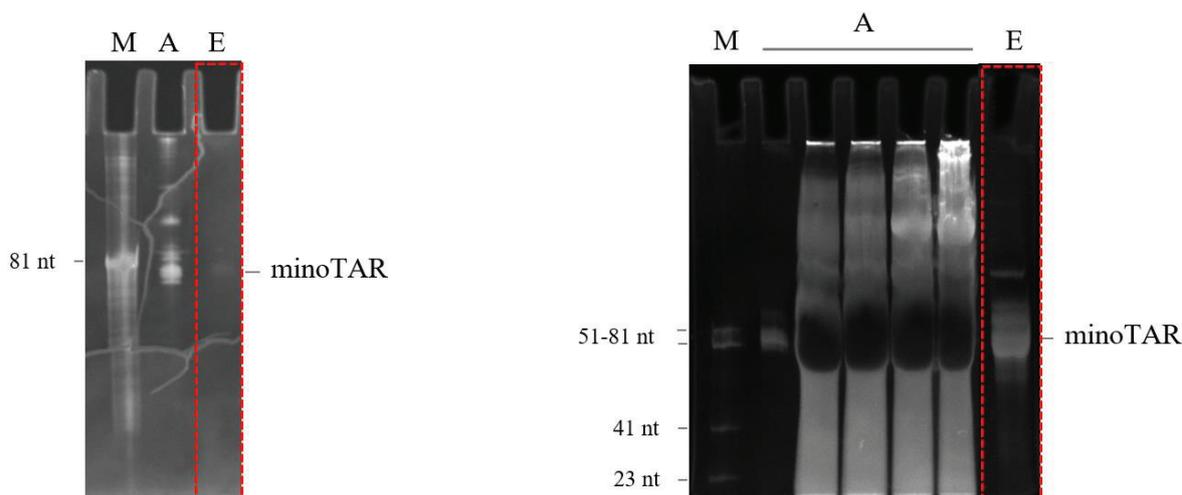


**Figure 67.** **A gauche**, chromatogramme de la purification de l'ARN minoTAR sur une colonne Superdex 75 10/300 GL. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'éluion en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et celle en violet à la mesure de l'absorbance à 260 nm. Le tracé vert correspond au pourcentage de tampon d'éluion et le tracé orange à la conductivité. **A droite**, électrophorèse en condition dénaturante sur un gel de polyacrylamide 15% à partir des fractions de la purification de l'ARN minoTAR sur une colonne Superdex 75 10/300 GL ; M : marqueur de taille). Les numéros des fractions d'éluion contrôlées sont précisés sur le gel.

Avant de réaliser des essais de cristallo-genèse avec l'ARN minoTAR, il est indispensable que l'ARN soit le plus pur possible. Pour cela, une purification de l'ARN sur gel est ensuite réalisée selon les conditions décrites dans le chapitre Matériel et Méthodes (voir **chapitre XII.3.3**). Deux modes d'éluion ont été comparés pour optimiser la qualité et la quantité d'ARN récupéré après purification. Il s'agit de l'éluion par diffusion (**figure 68**, à gauche) et de l'électro-éluion (**figure 68**, à droite).

Dans le cas de l'élution par diffusion, l'échantillon avant la purification sur gel contient des contaminants puisqu'il y a la présence de bandes supplémentaires. Après l'élution, une seule bande à la bonne taille est visible. Cependant cette bande étant faiblement intense, il n'est pas possible de savoir si des contaminants sont présents dans la limite de résolution due à la concentration de l'ARN.

Dans le cas de l'électro-élution, l'échantillon avant la purification sur gel contient également des contaminants. Après l'élution, quelques bandes contaminantes sont encore visibles, mais l'ARN d'intérêt récupéré est en grande quantité. Le rendement semble supérieur à celui de l'élution par diffusion. Cependant, les quantités d'ARN de départ avant la purification sur gel ne sont pas identiques donc il n'est pas possible de comparer correctement ces deux modes d'élution. Il sera nécessaire de répéter ces éluions avec une même quantité d'ARN de départ. Les deux préparations d'ARN ont été conservées et utilisées pour des essais de cristallisation.



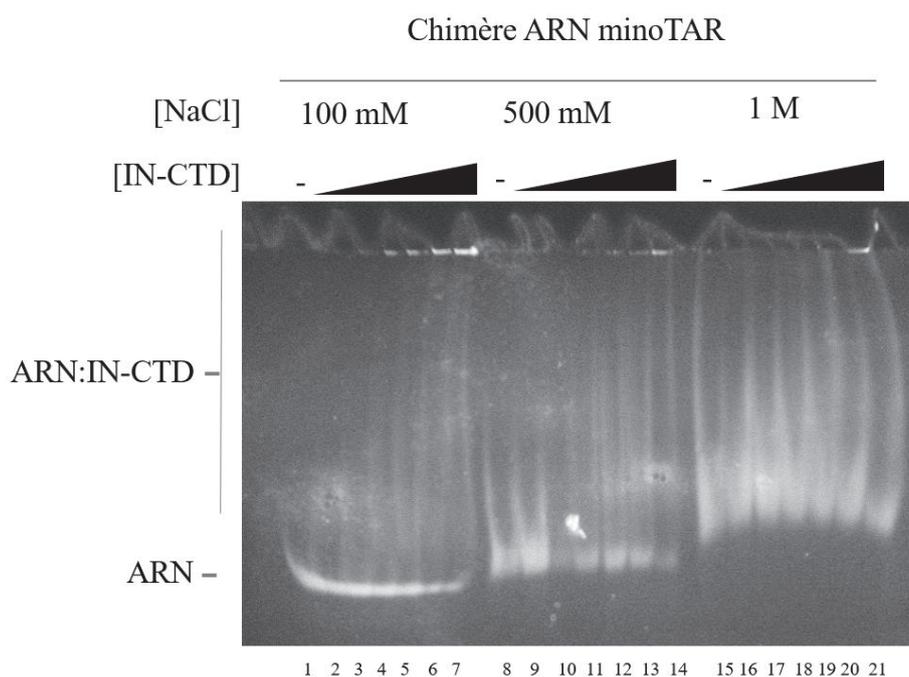
**Figure 68.** Electrophorèse en condition dénaturante sur un gel de polyacrylamide 10% pour contrôler la purification sur gel de l'ARN minoTAR avant et après élution par diffusion (à gauche) ou électro-élution (à droite). M : marqueur de taille ; A : échantillon avant purification sur gel ; E : échantillon après purification sur gel et élution par diffusion/ électro-élution. Les fractions encadrées en rouge sont celles qui ont été conservées pour les essais de cristallogénèse.

## VII.2. Optimisation des conditions de formation du complexe IN-CTD:ARN minoTAR

Des expériences de retard sur gel sont ensuite réalisées afin de s'assurer de l'interaction entre IN-CTD et l'ARN minoTAR et d'optimiser les conditions d'interaction. Lors de ces expériences, différents ratio ARN:protéine sont testés pour trois concentrations différentes de NaCl : 100 mM, 500 mM et 1 M, selon les conditions décrites dans le chapitre Matériel et Méthodes (voir **chapitre XII.5.1.1**). Les échantillons sont analysés par une électrophorèse en condition native.

Les puits 1, 8 et 15 sont des contrôles et permettent de déterminer la distance de migration de l'ARN minoTAR libre (**figure 69**). Dans le cas d'une concentration de NaCl finale de l'ordre de 100 mM ou 500 mM, plus la concentration d'IN-CTD augmente et plus le retard de migration est important. Une bande de très faible migration est observée dans certains puits et correspond au complexe IN-CTD:minoTAR puisque l'intensité de la bande de l'ARN libre dans le puits correspondant diminue (puits 4 à 7, puits 10 à 14 et puits 20-21). Ainsi plusieurs conditions de liaison permettent d'obtenir le complexe IN-CTD:minoTAR. Les conditions les plus intéressantes sont celles qui sont à plus basse concentration de NaCl (100 mM et 500 mM) car elles se rapprochent des conditions physiologiques et limitent le risque d'apparition de cristaux de sel lors des expériences de cristallogénèse. Ainsi, à 100 mM de NaCl, les quatre ratios ARN:protéine suivants permettent la formation du complexe IN-CTD:minoTAR : 1:4 ; 1:6 ; 1:8 et 1:16 ; à 500 mM NaCl, il s'agit des cinq ratios suivants : 1:2 ; 1:4 ; 1:6 ; 1:8 et 1:16.

De plus, des complexes de plus faibles masses moléculaires sont observés pour des concentrations en NaCl de 100 et 500 mM aux ratios molaires ARN:protéine de 1:8 et 1:16 (puits 6, 7, 13 et 14).



**Figure 69.** Expérience de retard sur gel montrant l'interaction de his-IN-CTD avec l'ARN minoTAR pour différents ratios molaires ARN:protéine (1:0, 1:1, 1:2, 1:4, 1:6, 1:8 et 1:16) à différentes concentrations de NaCl (100, 500 et 1 000 mM). Les numéros des puits sont précisés en bas du gel d'élution.

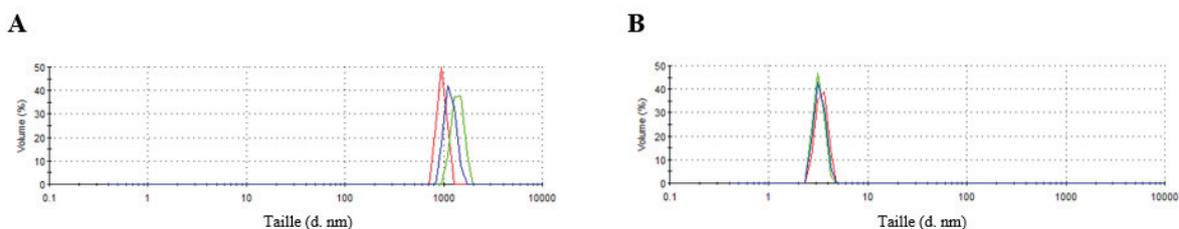
Des expériences de DLS (« *dynamic light scattering* ») ont ensuite été réalisées à des concentrations de NaCl de 100 mM et 250 mM afin de déterminer si le complexe en solution était monodisperse.

### VII.3. Analyse de l'état oligomérique du complexe IN-CTD:ARN minoTAR

Le complexe IN-CTD:ARN minoTAR à un ratio molaire ARN:protéine de 1:8, a été analysé par DLS à deux concentrations de NaCl, 100 et 250 mM, comme décrit dans le chapitre Matériel et Méthodes (voir **chapitre XIII.1**). Ce ratio 1:8 est choisi car il permet d'avoir une quantité importante de complexes IN-CTD:minoTAR en comparaison aux autres ratios testés dans l'expérience précédente (**figure 69**). Trois séries de mesures consécutives du diamètre hydrodynamique ( $D_H$ ) sont réalisées pour chaque échantillon.

Dans le cas du complexe formé à 100 mM NaCl, la courbe de distribution en volume ne présente qu'un seul pic pour chacune des trois séries de mesures (**figure 70.A**). Ces trois pics ne sont pas superposés et correspondent à un  $D_H$  aux alentours de 1 000 nm, ce qui signifie que les biomolécules sont agrégées. Ainsi, cet échantillon n'est pas utilisable pour des essais de cristallogénèse.

Dans le cas du complexe formé à 250 mM NaCl, la courbe de distribution en volume ne présente également qu'un seul pic pour chacune des trois séries de mesures (**figure 70.B**). Cependant dans ce cas, les trois pics se superposent avec un  $D_H$  de 3,3 nm et un bon PDI (« *polydispersity index* ») de 0,3. Cette valeur de  $D_H$  correspond à une masse moléculaire d'environ 18 kDa, ce qui pourrait correspondre à un dimère d'IN-CTD mais pas au complexe puisque l'ARN minoTAR seul a une masse moléculaire de 26 kDa.



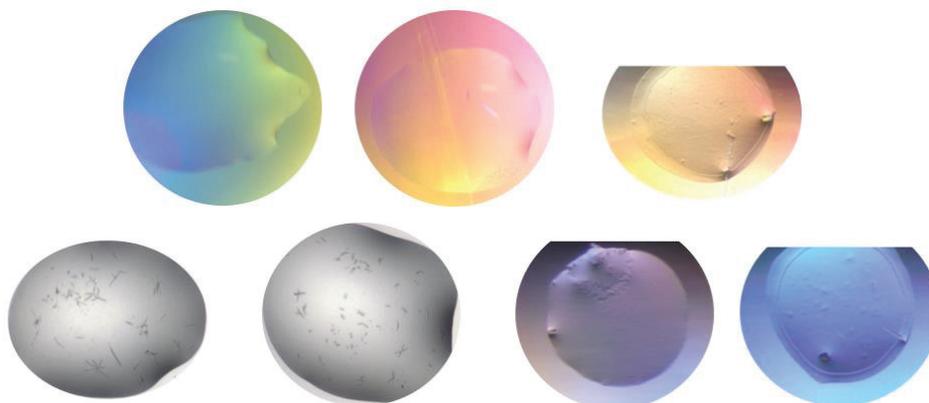
**Figure 70.** Analyse par DLS des complexes his-IN-CTD:minoTAR formés à deux concentrations de NaCl : 100 mM (A) et 250 mM (B). Les trois courbes (rouge, bleu et verte) sur chaque graphe correspondent aux trois jeux de données générés par l'appareil Zetasizer Nano S90.

#### VII.4. Essais de cristallogénèse sur le complexe IN-CTD:ARN minoTAR

Bien que les résultats de DLS ne corroborent pas la formation du complexe observée sur gel, nous avons profité du restant de protéine purifiée pour faire des essais de cristallisation selon la méthode de la diffusion de vapeur d'eau (voir **chapitres XIII.2.3.2** et **XIII.2.3.3**) (**tableau 1**). Sept cristaux ont été analysés au synchrotron SOLEIL et à l'ESRF au total pour ces deux essais (**figure 71**) mais lors de leur caractérisation, il est apparu qu'il s'agissait de cristaux de sel (paramètres de maille trop petits pour des cristaux de protéines).

		Ratio molaire ARN minoTAR:IN-CTD	Kits commerciaux			
			MIDAS plus	Natrix HT	PEG I	Mini Screen
Température	19°C	1:2				
		1:4				
		1:16				
	37°C	1:2				
		1:4				

**Tableau 1.** Représentation des différentes conditions de cristallisation testées (en jaune) et de celles ayant données des cristaux (en vert).



**Figure 71.** Photos de cristaux apparus dans des gouttes lors des essais de cristallogénèse.

### Chapitre 3 : Discussion et perspectives (première partie)

Les objectifs étaient : (1) de trouver le domaine de l'IN capable d'interagir avec l'ARN TAR et de déterminer le site de liaison sur l'ARN ; (2) de décortiquer les mécanismes moléculaires de formation du complexe entre l'IN, TAR et Tat ; (3) d'obtenir la structure tridimensionnelle du complexe formé entre l'IN et TAR par cristallographie aux rayons X ; et (4) de comprendre si le complexe formé entre l'IN et TAR est capable d'interagir avec le complexe d'élongation de la transcription.

Les travaux présentés dans l'article de recherche (Rocchi, Louvat *et al.*, 2022) répondent aux objectifs (1) et (2), ils sont discutés dans la partie Discussion de l'article.

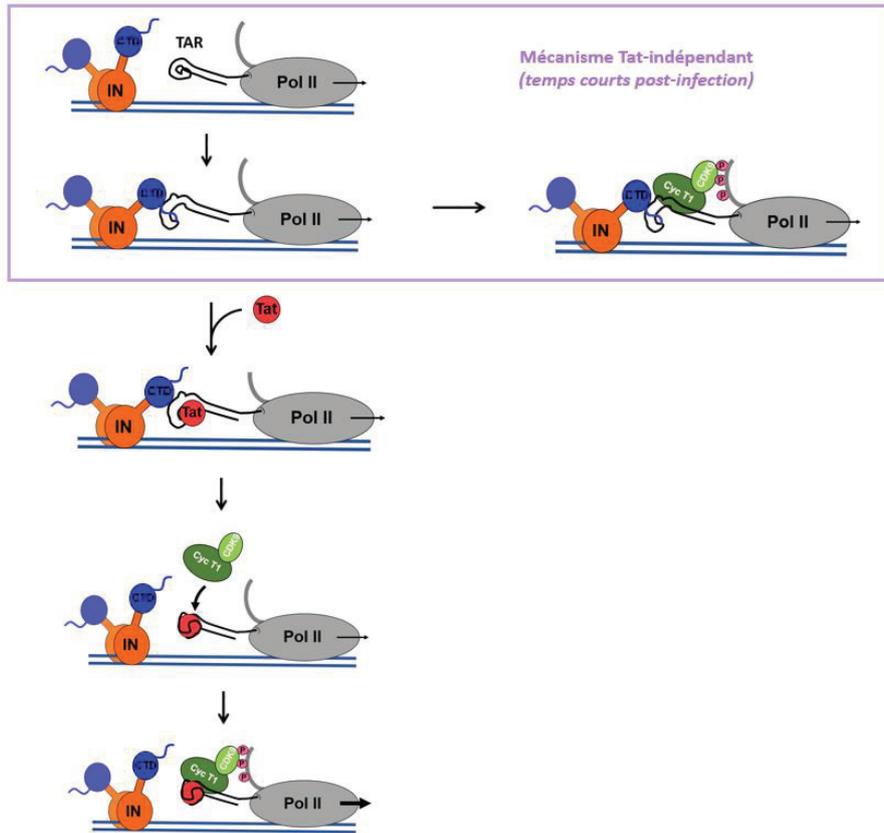
Concernant l'objectif (3), je n'ai pas réussi à obtenir des cristaux du complexe CTD:minoTAR. Plusieurs optimisations peuvent être envisagées pour augmenter les chances de cristalliser le complexe IN:ARN ; par exemple utiliser une construction du CTD sans le CT pour diminuer la flexibilité, et réaliser le complexe avec une IN sous forme tétramérique pour obtenir un complexe plus stable comme décrit par Kessl *et al.* (2016) et Elliott *et al.* (2020) qui montrent que l'IN tétramérique a une meilleure affinité pour l'ARN. De plus, l'IN au cours de la morphogenèse est présente dans la polyprotéine Gag-Pol, nous sommes donc intéressés par la suite à la structure tridimensionnelle de la polyprotéine RT-IN en complexe avec l'ARN. Ces travaux sont en cours dans l'équipe en utilisant la microscopie électronique.

J'ai mis en évidence *in vitro* que la cycline T1 interagit avec l'ARN TAR en présence du CTD de l'IN du VIH-1. Ce résultat fait écho à des études *in vitro* et *in vivo* qui ont démontré que l'interaction de la cycline T1 et du complexe P-TEFb avec l'ARN TAR était possible uniquement en présence de Tat (Richter *et al.*, 2002 ; Wei *et al.*, 1998 ; Bieniasz *et al.*, 1999). La liaison de Tat à la cycline T1 permet d'exposer une boucle de la cycline T1 qui comprend la C261 requise pour l'interaction avec l'ARN TAR (Fujinaga *et al.*, 1999 ; Garber *et al.*, 1998). Ainsi, la liaison de la cycline T1 à l'ARN TAR en présence du CTD suggère que le CTD modifie la conformation de la cycline T1 pour permettre cette interaction. Il serait intéressant de vérifier cette hypothèse par la suite.

De plus, il est intéressant de noter que l'interaction de Tat avec la cycline T1 entraîne un changement de conformation de la protéine Tat qui permet au complexe d'être plus affiné pour lier l'ARN TAR en mettant en jeu six cystéines et deux ions  $Zn^{2+}$  au niveau du domaine ARM de la protéine Tat (Garber *et al.*, 1998). L'interaction de la cycline T1 et de Tat avec l'ARN TAR se fait au niveau de la tige-boucle apicale, en mettant en jeu les nucléotides G32 et G33 (Schulze-Gahmen et Hurley, 2018). Nous avons montré que le CTD lie la tige-boucle apicale de l'ARN TAR et permet d'exposer les nucléotides G28 et G32 (Rocchi, Louvat *et al.*, 2022). Ce résultat a été conforté par des expériences *in vitro* et *in vivo* (Liu *et al.*, 2021), cela suggère que le CTD pourrait recruter le complexe P-TEFb et jouer le rôle d'un adaptateur moléculaire qui permettrait la liaison de la cycline T1 à TAR via l'exposition du G32. Ce mécanisme pourrait expliquer l'origine de la protéine Tat et pourrait avoir lieu à des temps courts après l'infection, ce qui serait en accord avec le rôle observé de l'IN du VIH-1 par Winans et Goff (2020). Nos résultats *in vitro* montrent que l'affinité de la protéine Tat pour l'ARN TAR est supérieure à celle du CTD, donc à terme, lorsque la concentration de Tat est suffisante, Tat entraîne la dissociation du CTD de l'ARN TAR et joue son rôle de transactivateur de la transcription (**figure 72**).

Il serait intéressant d'étudier ce mécanisme dans un contexte cellulaire par des expériences de transcription *in vivo* et des expériences de colocalisation *in vivo*. De plus, l'étude de mutants de la cycline T1, notamment pour les résidus W258 et R251 qui sont critiques pour l'interaction avec le complexe TAR:Tat (Schulze-Gahmen et Hurley, 2018), permettrait d'étudier les sites d'interaction de la cycline T1 avec l'IN. Troisièmement, l'étude structurale du complexe formé entre le CTD de l'IN, l'ARN TAR et le complexe P-TEFb permettrait de confirmer les résultats fonctionnels obtenus *in vitro* et d'obtenir des détails structuraux sur les résidus impliqués dans les interactions entre ces différents partenaires.

J'ai essayé de produire la CDK9 dans un système d'expression bactérien afin d'étudier l'interaction entre TAR, l'IN et P-TEFb. Cependant la protéine se retrouve dans les corps d'inclusion. Dans la littérature scientifique, la production de la CDK9 est essentiellement décrite en système mammifère ou baculovirus à l'exception d'une étude réalisée en système bactérien (Leopoldino *et al.*, 2006). J'ai donc exprimé la CDK9 dans ces conditions et je prévois de réaliser la purification dans les conditions décrites par la suite.



**Figure 72.** Schéma du mécanisme d'activation de l'élongation de la transcription provirale proposé suite à mes travaux de thèse.

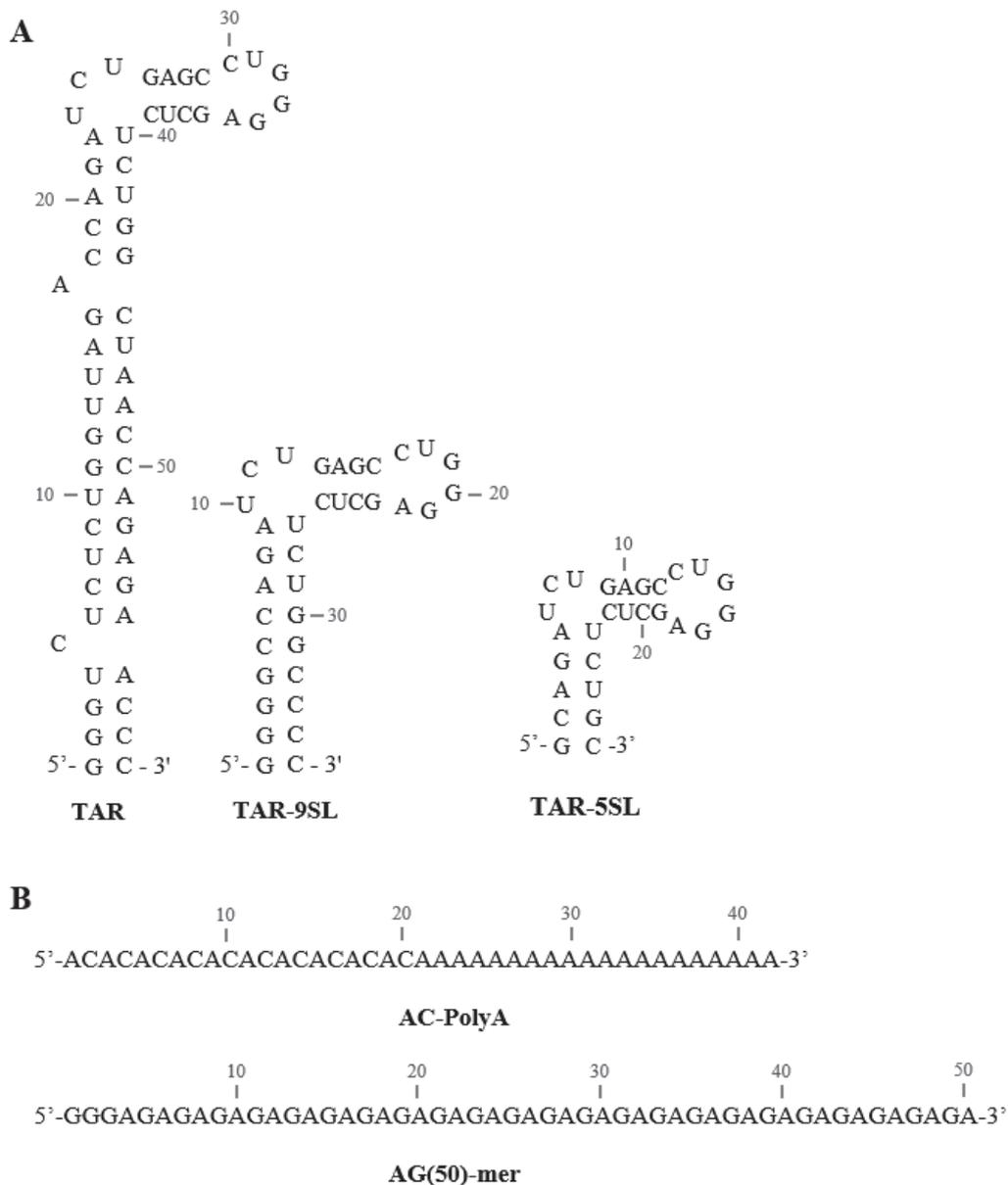
## **PARTIE II : Caractérisation biochimique de la liaison et de l'activité enzymatique du facteur de restriction viral ISG20 vis-à-vis d'un substrat ARN**

Le but de cette étude est de comprendre les mécanismes d'action et de spécificité d'ISG20 vis-à-vis d'un substrat ARN dans le contexte d'une infection virale. Pour cela, j'ai purifié ISG20<sub>wt</sub> et mutante ainsi que la PABPC1 et différents substrats ARN, puis j'ai effectué des tests de liaison et de digestion enzymatique de l'ARN.

### **Chapitre 4 : Résultats**

#### VIII. Préparation de l'ARN

Les ARN utilisés pour les études avec ISG20 sont les suivants : TAR-5SL, TAR-9SL, TAR, AC-PolyA et AG(50)-mer (**figure 73**). Les ARN TAR-5SL, TAR-9SL et TAR sont des ARN structurés (**figure 73.A**), tandis que les ARN AC-PolyA et AG(50)-mer sont des ARN linéaires (**figure 73.B**).



**Figure 73. (A)** Modèles des structures secondaires des ARN obtenus grâce à RNAfold web server avec  $\Delta G(\text{TAR}) = -29.60 \text{ kcal/mol}$ ,  $\Delta G(\text{TAR-9SL}) = -19,74 \text{ kcal/mol}$ , et  $\Delta G(\text{TAR-5SL}) = -9,80 \text{ kcal/mol}$ . **(B)** Séquences des ARN linéaires AC-PolyA et AG(50)-mer.

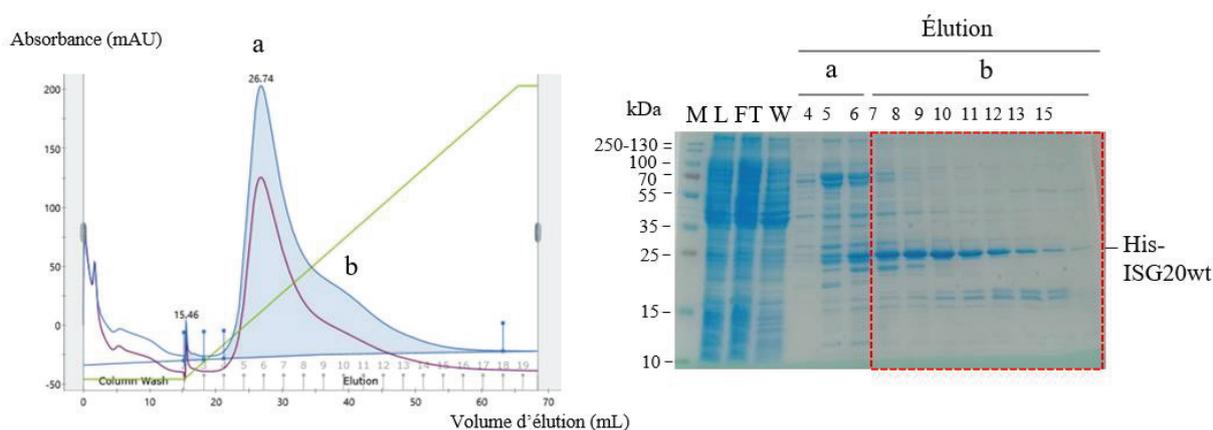
Les ARN TAR-9SL, TAR et AG(50)-mer sont produits par une réaction de transcription *in vitro*, tandis que les ARN TAR-5SL et AC-PolyA sont synthétisés chimiquement et déphosphorylés en 5', comme détaillé dans le chapitre Matériel et Méthodes (voir **chapitre XII.3**). Ils sont ensuite marqués radioactivement à l'extrémité 5' par l'ajout d'un  $^{32}\text{P}$  et purifiés d'abord avec une colonne G-25 et ensuite sur un gel dénaturant (voir **chapitre XII.4**, Matériel et Méthodes). Les ARNs purifiés sont dosés et utilisés dans les tests présentés ci-dessous.

## IX. Caractérisation de l'activité enzymatique d'ISG20

### IX.1. Purification et stockage de la protéine recombinante his-ISG20*wt*

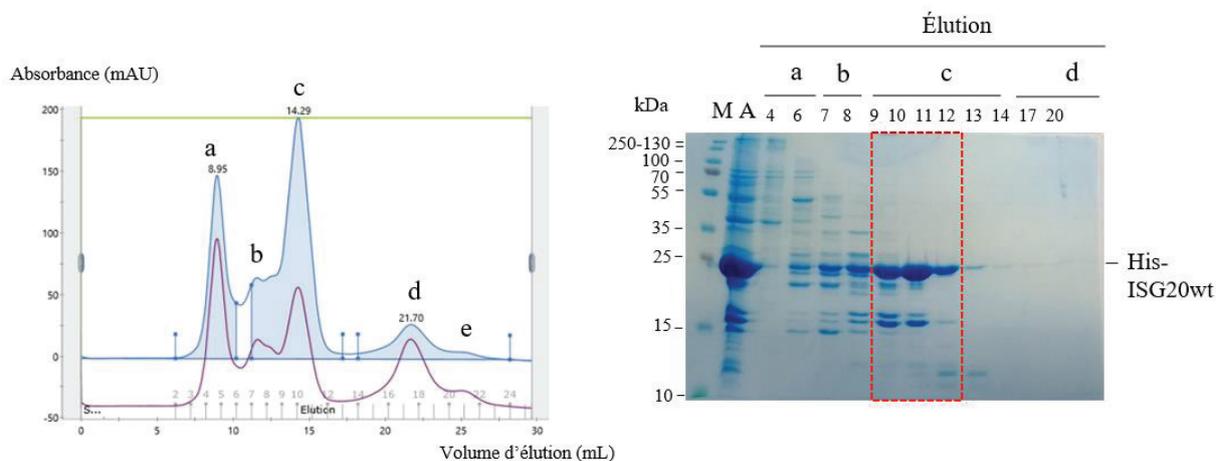
Les conditions de production de la protéine his-ISG20*wt* ont été scrupuleusement optimisées car cette protéine s'est révélée toxique pour *E. coli*. Nous avons donc trouvé la température optimale pour la croissance bactérienne ainsi que celle pour l'expression (voir **chapitre XII.1.5**, Matériel et Méthodes). La concentration d'IPTG a été réduite et les temps de croissance et d'expression ont été augmentés.

La première étape de purification de his-ISG20*wt* se fait par une chromatographie IMAC. Sur le chromatogramme présenté dans la **figure 74** (à gauche), deux pics d'éluion sont observés (pics a et b). Pour confirmer l'identité des espèces séparées lors de la chromatographie, une analyse SDS-PAGE est réalisée. La protéine recombinante his-ISG20*wt* a une masse moléculaire de 23 kDa. Ainsi, en regardant le gel de polyacrylamide où ont été chargés les fractions d'éluion sur la **figure 74** (à droite), on peut voir que le premier pic correspond à l'éluion des protéines d'intérêt avec des contaminants, tandis que le second pic correspond à l'éluion des protéines d'intérêt avec moins de contaminants. Le second pic est donc conservé et purifié par la suite par une seconde étape de chromatographie.



**Figure 74. A gauche**, chromatogramme de la purification de his-ISG20*wt* sur une colonne HisTrap HP 5 mL. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'éluion en mL. La courbe en bleu correspond à la mesure de l'absorbance à 280 nm et celle en violet à la mesure de l'absorbance à 260 nm. Le tracé vert correspond au pourcentage de tampon d'éluion. **A droite**, SDS-PAGE 15% sur les fractions de la purification de his-ISG20*wt* sur colonne HisTrap (M : marqueur de masses moléculaires ; L : lysat ; FT : fraction non retenue ; W : fraction de lavage). Les fractions encadrées en rouge sont celles qui ont été conservées pour la suite de la purification. Les numéros des fractions d'éluion contrôlées sont précisés sur le gel.

La seconde étape de purification de his-ISG20<sub>wt</sub> se fait par une chromatographie SEC sur une colonne Superdex 75 Increase 10/300 GL. Plusieurs pics sont observés sur le chromatogramme (**figure 75**, à gauche), ils sont contrôlés par SDS-PAGE (**figure 75**, à droite). Le puits 2 contenant la fraction avant la chromatographie SEC confirme la présence de la protéine avant la purification. His-ISG20<sub>wt</sub> est éluée dans les fractions 9 à 12, correspondant au pic d'éluion à 14,29 mL. D'après la calibration de la colonne Superdex 75 Increase 10/300 GL, il semblerait que his-ISG20<sub>wt</sub> soit sous forme monomérique en solution. En effet, la masse moléculaire d'un monomère est de 23 kDa et le pic obtenu à 14,29 mL de volume d'éluion correspond à une masse moléculaire de 23,5 kDa (**Annexes, figure 93**). Ce pic d'éluion est conservé pour des expériences fonctionnelles.

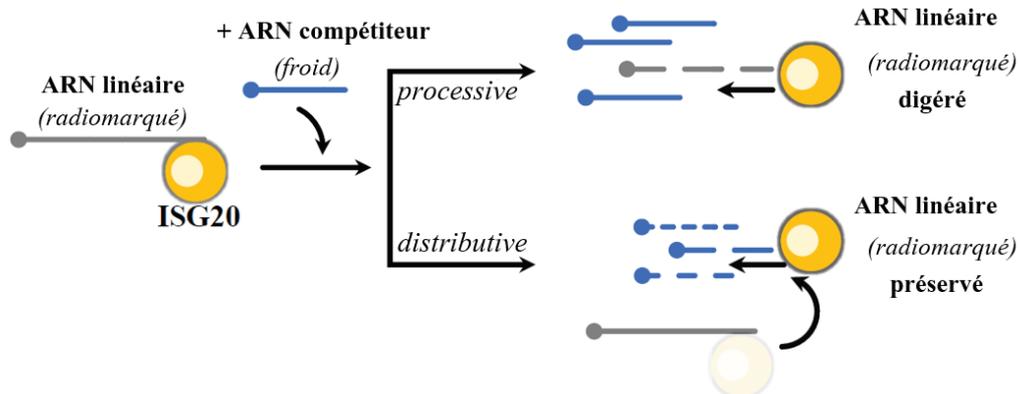


**Figure 75.** *A gauche*, chromatogramme de la purification de his-ISG20<sub>wt</sub> sur une colonne Superdex 75 10/300 GL. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'éluion en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et celle en violet à la mesure de l'absorbance à 260 nm. Le tracé vert correspond au pourcentage de tampon SEC. *A droite*, SDS-PAGE 15% sur les fractions de la purification SEC de his-ISG20<sub>wt</sub> (M : marqueur de masses moléculaires ; A : échantillon avant l'injection sur la colonne Superdex 75 10/300 GL). Les fractions encadrées en rouge sont celles qui ont été conservées pour la suite des expériences. Les numéros des fractions d'éluion contrôlés sont précisés sur le gel.

Au fil des expériences, nous nous sommes rendu compte qu'ISG20 perdait rapidement son activité au cours du temps. Nguyen *et al.* (2001) ont décrit qu'une conservation à -20°C dans un tampon contenant 50% de glycérol permet de conserver l'activité enzymatique pour environ trois mois. Nous avons ainsi décidé de conserver ISG20 dans ces conditions.

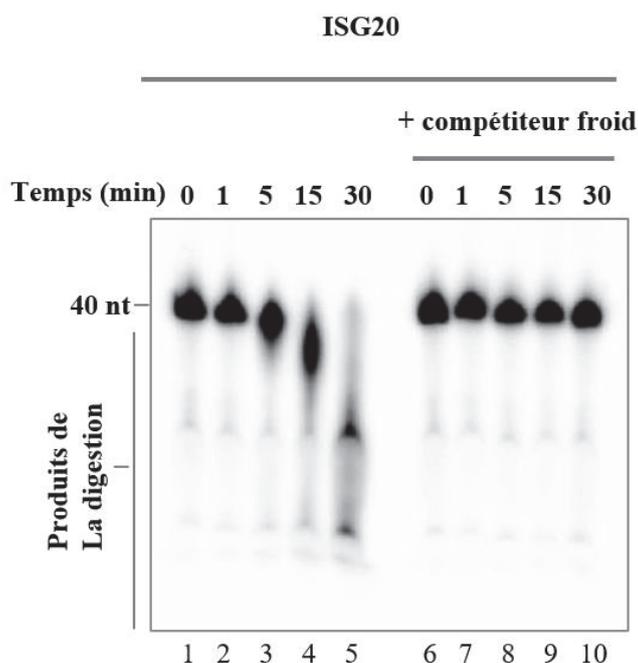
## IX.2. Etude de la processivité d'ISG20

Une question fondamentale est de savoir si l'activité exonucléase d'ISG20 repose sur un mécanisme processif ou distributif : cette propriété enzymatique n'a jamais été correctement caractérisée jusqu'à présent. Nous avons effectué un test de processivité à l'état stationnaire. La processivité d'une exonucléase est la capacité de l'enzyme à catalyser plusieurs cycles de digestion sans se détacher de son substrat. Une enzyme distributive catalyse un ou deux cycles de digestion avant de se décrocher de son substrat pour s'accrocher à nouveau. Techniquement cela peut être observé en adaptant les conditions de réaction. Pour cela, j'ai incubé l'ARN linéaire radiomarqué AC-PolyA avec ou sans un piège d'ARN froid (**figure 76**). Le début de la réaction est initié par l'ajout d'ISG20. Des échantillons de la réaction de digestion sont prélevés à différents temps et la réaction est arrêtée comme détaillé dans le chapitre Matériel et Méthodes (voir **chapitre XII.5.3.1**). Les échantillons sont analysés par une électrophorèse en condition dénaturante sur un gel de séquence.



**Figure 76.** Mécanismes d'action processif et distributif vis-à-vis d'un substrat ARN linéaire radiomarqué (en gris) en présence d'ARN linéaire compétiteur froid (en bleu).

En comparant le profil de digestion en absence et en présence de compétiteur froid, nous pouvons voir que l'ARN n'est pas digéré au cours des 30 minutes de cinétique en présence du compétiteur (**figure 77**, puits 6 à 10) alors que des bandes de digestion à différentes tailles sont présentes en absence du compétiteur (puits 3 à 5). En effet, à partir de 5 minutes après l'ajout d'ISG20, des produits de digestion apparaissent et au bout de 30 minutes de réaction, l'ARN est quasiment totalement digéré. Ainsi, la présence du compétiteur inhibe l'activité d'ISG20, ce qui signifie qu'ISG20 est une enzyme distributive.



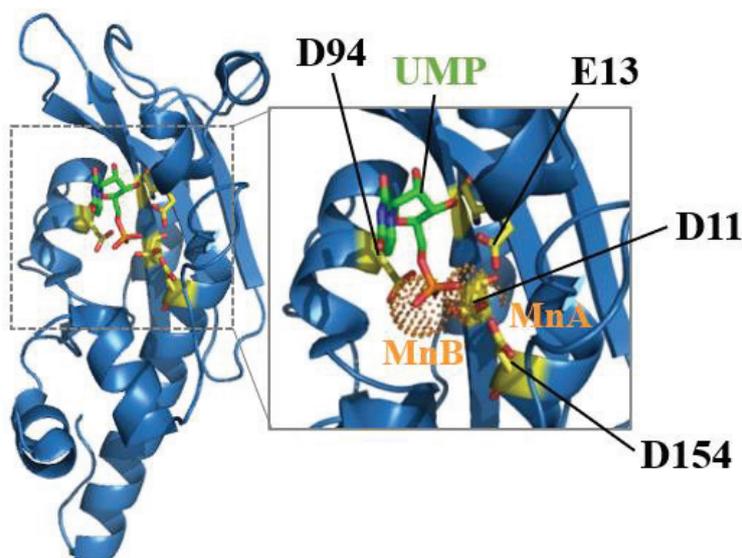
**Figure 77.** ISG20 a un mode d'action distributif. L'activité exonucléase d'ISG20 sur un ARNsb marqué radioactivement au  $^{32}\text{P}$  en absence et en présence d'un ARNsb froid a été étudiée pendant 30 minutes et contrôlée par une électrophorèse sur gel de polyacrylamide 15% en condition dénaturante.

### IX.3. Etude du mutant catalytique D94

Dans le cadre de notre collaboration avec l'équipe du Dr. Cimarelli, j'ai travaillé sur le mutant catalytique D94 d'ISG20 (**figure 78**). L'activité enzymatique d'ISG20 humaine nécessite deux cations métalliques et quatre résidus qui constituent le site catalytique : D11, E13, D94 et D154. La structure

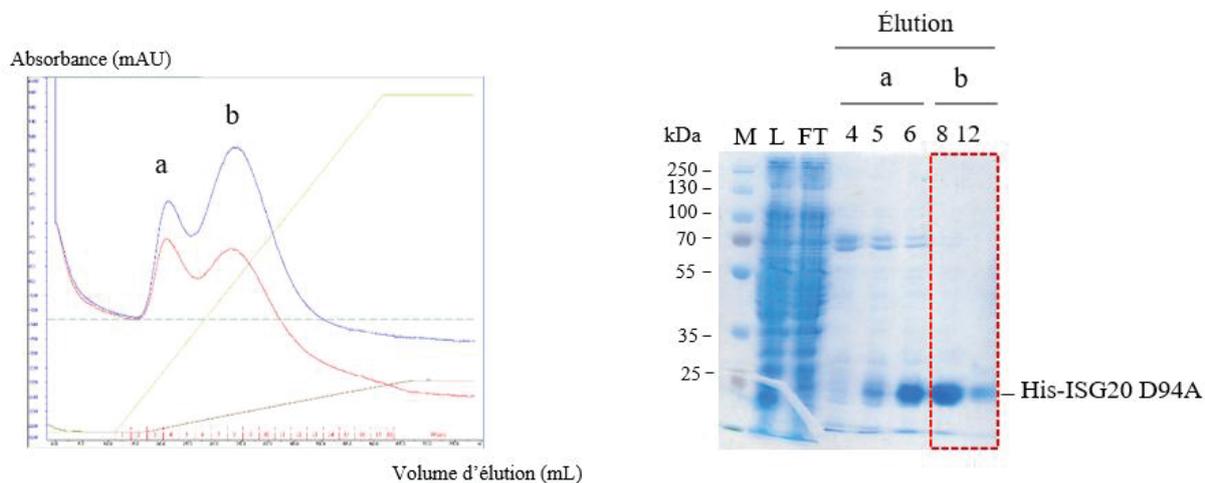
cristalline d'ISG20 résolue avec de l'UMP a permis de mettre en évidence que l'ion MnA est impliqué dans l'attaque nucléophile du groupement phosphate de l'acide nucléique. En effet, l'ion MnA interagit avec les oxygènes des fonctions carboxylates des résidus D11, E13 et D154, et les oxygènes du groupement phosphate de l'UMP (Horio *et al.*, 2004). Le rôle de l'ion MnB n'est pas encore totalement compris. D'après la structure, l'ion MnB interagit avec les oxygènes des fonctions carboxylates des résidus D11, du groupement phosphate de l'UMP et de quatre molécules d'eau. Parmi ces molécules d'eau, deux sont impliquées dans des interactions hydrogènes avec les oxygènes de la fonction carboxylique du résidu D94.

Les résultats obtenus dans l'équipe du Dr. Cimarelli ont montré que le mutant D94 perd son activité antivirale contre le virus de la stomatite vésiculaire (VSV) (Wu *et al.*, 2019). Pour mieux comprendre ce phénotype, j'ai étudié l'interaction du mutant D94 et de la protéine *wt* vis-à-vis d'un substrat ARN.



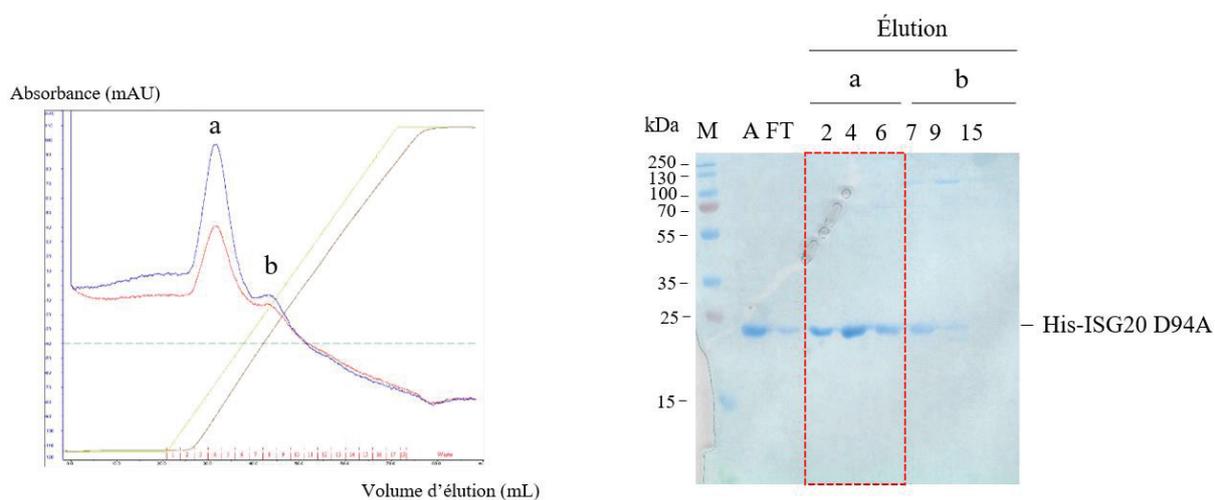
**Figure 78.** Structure cristalline d'ISG20 humaine (Horio *et al.*, 2004 ; code PDB : 1WLJ). ISG20 est représentée en rubans (bleu). Les résidus catalytiques sont représentés en bâton et les ions manganèse sous forme de sphères (jaune).

La protéine his-ISG20 D94A est purifiée par deux étapes de chromatographie (voir **chapitre XII.1.6**, Matériel et Méthodes), la première est une chromatographie IMAC. Les deux pics d'éluion (pics a et b) observés sur le chromatogramme (**figure 79**, à gauche) sont analysés par SDS-PAGE. La protéine recombinante his-ISG20 D94A a une masse moléculaire de 23 kDa. Sur le gel de polyacrylamide où ont été chargés les fractions d'éluion (**figure 79**, à droite), on peut voir que le premier pic correspond à l'éluion des protéines d'intérêt avec des contaminants aux alentours de 70 kDa, tandis que le second pic correspond à l'éluion des protéines d'intérêt sans ces contaminants. Il est donc conservé et purifié par la suite par une chromatographie d'affinité sur colonne Heparin.



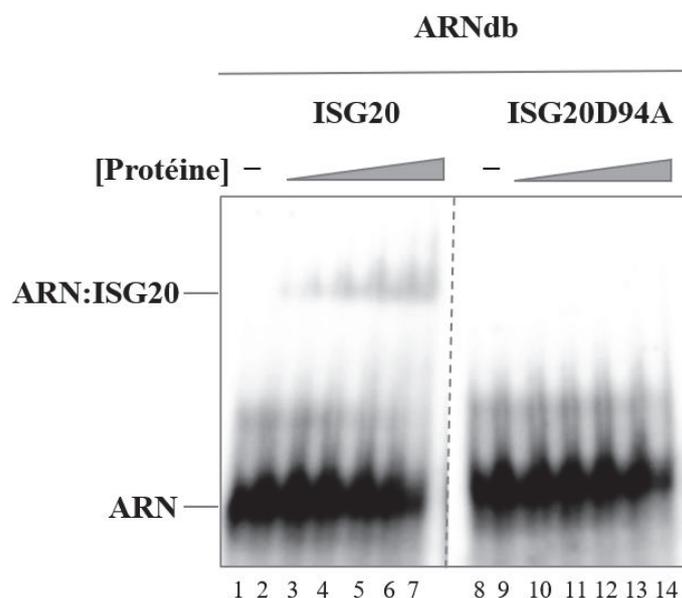
**Figure 79.** *A gauche*, chromatogramme de la purification de his-ISG20 D94A sur une colonne HisTrap crude 5 mL. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'éluion en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et celle en rouge à la mesure de l'absorbance à 260 nm. Le tracé vert correspond au pourcentage de tampon d'éluion et le tracé marron à la conductivité. *A droite*, SDS-PAGE 12% sur les fractions de la purification de his-IS20 D94A sur colonne HisTrap (M : marqueur de masses moléculaires ; L : lysat ; FT : fraction non retenue). Les fractions encadrées en rouge sont celles qui ont été conservées pour la suite de la purification. Les numéros des fractions d'éluion contrôlées sont précisés sur le gel.

Deux pics sont observés sur le chromatogramme de la seconde étape de purification (**figure 80**, à gauche). Ils sont caractérisés par SDS-PAGE (**figure 80**, à droite). Le puits 3, correspondant à la fraction non retenue, présente une légère bande à 23 kDa correspondant à his-ISG20 D94A, ainsi toute la protéine présente au départ (puits 2) n'est pas retenue sur la colonne. His-ISG20 D94A est éluée sans contaminants dans le pic d'éluion a, tandis qu'une faible quantité de his-ISG20 D94A est éluée en présence de contaminants aux alentours de 130 kDa dans le second pic. Le premier pic est conservé pour les expériences fonctionnelles.



**Figure 80.** *A gauche*, chromatogramme de la purification de his-ISG20 D94A sur une colonne Heparin HiTrap 5 mL. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'éluion en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et celle en rouge à la mesure de l'absorbance à 260 nm. Le tracé vert correspond au pourcentage de tampon Héparine B. *A droite*, SDS-PAGE 15% sur les fractions de la purification de his-ISG20 D94A sur colonne Heparin (M : marqueur de masses moléculaires ; A : échantillon avant l'injection sur la colonne de purification ; FT : fraction non retenue). Les fractions encadrées en rouge sont celles qui ont été conservées pour la suite des expériences. Les numéros des fractions d'éluion contrôlées sont précisés sur le gel.

J'ai réalisé une expérience de retard sur gel pour étudier l'interaction d'ISG20<sub>wt</sub> et mutante en présence d'un ARNdb. Pour cela, j'ai incubé les deux protéines, séparément, à des concentrations croissantes avec l'ARN structuré TAR-9SL pendant 30 minutes à 37°C comme décrit dans le chapitre Matériel et Méthodes (voir **chapitre XII.5.1.2**). Ensuite, les échantillons sont analysés par une électrophorèse sur gel de polyacrylamide en condition native. Les puits 1 et 8 sont des contrôles qui permettent de déterminer la distance de migration de l'ARN libre (**figure 81**). Dans les cas des ratios molaires 1:1 à 1:16 pour la protéine ISG20<sub>wt</sub>, une bande de très faible migration est observée (puits 2 à 7) et correspond au complexe ISG20:TAR-9SL. L'intensité de cette bande augmente de manière proportionnelle à l'augmentation du ratio ARN:protéine. Ainsi, plus il y a d'ISG20<sub>wt</sub>, plus y a de complexes ISG20:TAR-9SL formé. Cependant, dans le cas de la protéine mutante D94A, nous n'observons pas cette bande de faible migration correspondant au complexe ARN:ISG20. La mutation D94 empêche la liaison de l'ARN structuré à ISG20, donc ce résidu catalytique est impliqué dans la liaison du substrat. Ce résultat suggère que le second ion Mn<sup>2+</sup> du site actif pourrait intervenir dans la liaison de l'ARN.

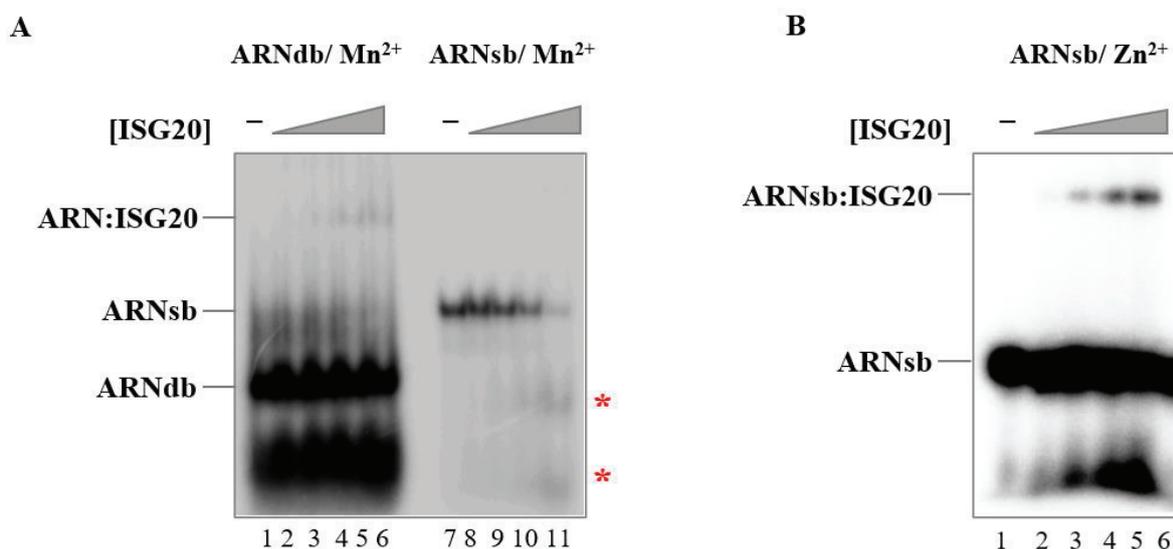


**Figure 81.** La mutation D94A dans le site catalytique empêche la liaison d'ISG20 à l'ARN. Expérience de retard sur gel montrant l'interaction de *his*-ISG20<sub>wt</sub> et *his*-ISG20 D94A avec un ARNdb radiomarqué.

#### IX.4. Etude de la liaison d'ISG20 à des substrats ARNdb et ARNsb

Afin de caractériser la liaison d'ISG20 à des substrats ARNsb et ARNdb, j'ai réalisé un test de retard sur gel en utilisant soit l'ARN linéaire AG(50)-mer soit l'ARN structuré TAR-9SL avec ISG20 à différents ratios molaires ARN:protéine comme décrit dans le chapitre Matériel et Méthodes (voir **chapitre XII.5.1.3**). Sur le gel natif (**figure 82.A**) nous observons une bande correspondant au complexe ISG20:TAR-9SL de très faible intensité dans le cas des ratios molaires 1:1 à 1:8 (puits 2 à 5). L'intensité de cette bande augmente de manière proportionnelle à l'augmentation du ratio ARN:protéine, donc plus il y a d'ISG20, plus y a de complexes ISG20:TAR-9SL formé. Dans le cas de l'ARN linéaire AG(50)-mer, des bandes supplémentaires avec une distance de migration supérieure à l'ARN libre apparaissent lorsque le ratio molaire ARN:protéine augmente. Ces bandes correspondent à la digestion de l'ARN linéaire. Ainsi, lorsque la quantité d'ISG20 augmente, il y a apparition de formes digérées de l'ARN, et leur quantité augmente proportionnellement à la quantité d'ISG20. L'ARN linéaire est donc digéré en présence d'ISG20, ce qui n'est pas le cas de l'ARN structuré qui se lie uniquement à ISG20 sans être digéré.

La digestion de l'ARNsb par ISG20 a déjà été montré *in vitro* (Nguyen *et al.*, 2001), les auteurs ont également mis en évidence que la présence d'ions  $Zn^{2+}$  à la place de  $Mn^{2+}$  permettait de limiter la digestion médiée par ISG20. J'ai donc répété la même expérience que précédemment avec un ARN linéaire AC-PolyA en présence d'ions  $Zn^{2+}$  au lieu de  $Mn^{2+}$  dans le tampon de digestion pour étudier la liaison d'ISG20 à l'ARNsb (voir **chapitre XII.5.1.4**, Matériel et Méthodes). Sur la **figure 82.B**, nous observons une liaison de l'ARNsb à ISG20 (puits 2 à 6). Plus la concentration d'ISG20 augmente, plus la quantité de complexe ARN:ISG20 augmente et plus nous observons de fragments d'ARN digérés.



**Figure 82.** ISG20wt lie l'ARN structuré et linéaire mais elle ne dégrade que l'ARN linéaire. **(A)** Expérience de retard sur gel montrant l'interaction de his-ISG20wt avec un ARNsb (AG(50)-mer) et un ARNdb (TAR-9SL) en présence de  $Mn^{2+}$ . Les étoiles rouges correspondent aux produits de digestion. **(B)** Expérience de retard sur gel montrant l'interaction de his-ISG20wt avec un ARNsb (AC-PolyA) en présence de  $Zn^{2+}$ . L'étoile rouge correspond aux produits de digestion.

## X. Mécanismes de protection des ARNm cellulaires et/ou d'échappement viral

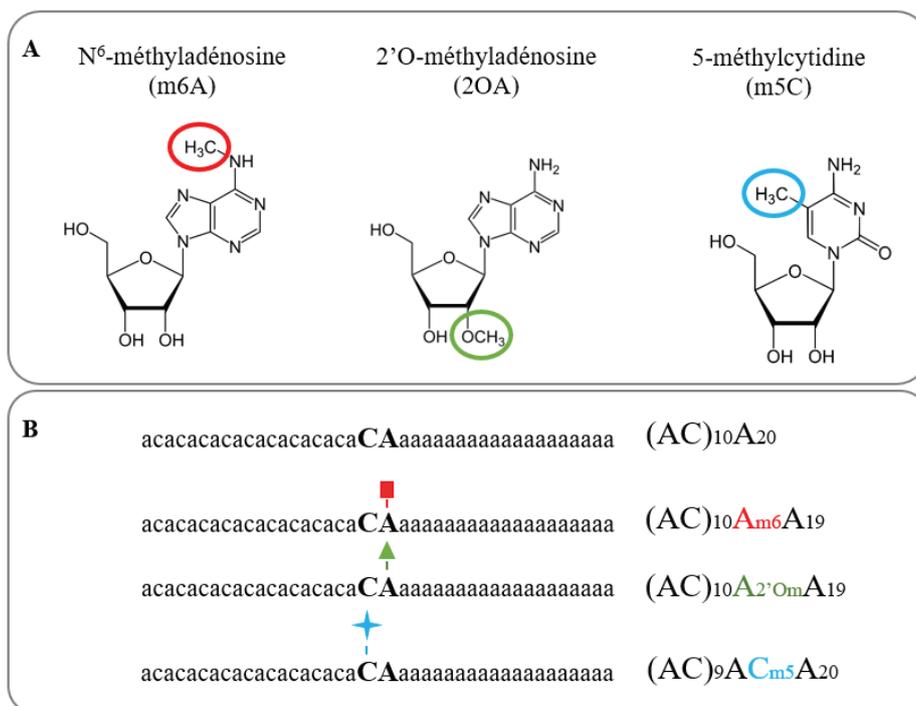
Le mécanisme d'action d'ISG20 lors de la réponse antivirale n'est pas encore connu à ce jour (Deymier *et al.*, 2022). Il a été montré par des expériences de biologie cellulaire qu'ISG20 est capable de faire la distinction entre le soi et le non-soi (Wu *et al.*, 2019). J'ai étudié différents déterminants de l'ARN pour identifier des mécanismes de protection contre l'activité d'ISG20. Mes travaux portent sur : (1) les modifications épitranscriptomiques de l'ARN, (2) la structure de l'ARN, et (3) l'extrémité PolyA en 3' de l'ARN qui permet la liaison à des protéines.

### X.1. La protection par les modifications épitranscriptomiques

Les modifications épitranscriptomiques de l'ARN sont localisées à différents endroits dans l'ARN et sont associées à différentes fonctions, comme la régulation de la traduction et de la dégradation des ARNm (Nachtergaele et He, 2017). J'ai étudié l'effet de trois modifications épitranscriptomiques d'un ARNsb sur l'activité d'ISG20.

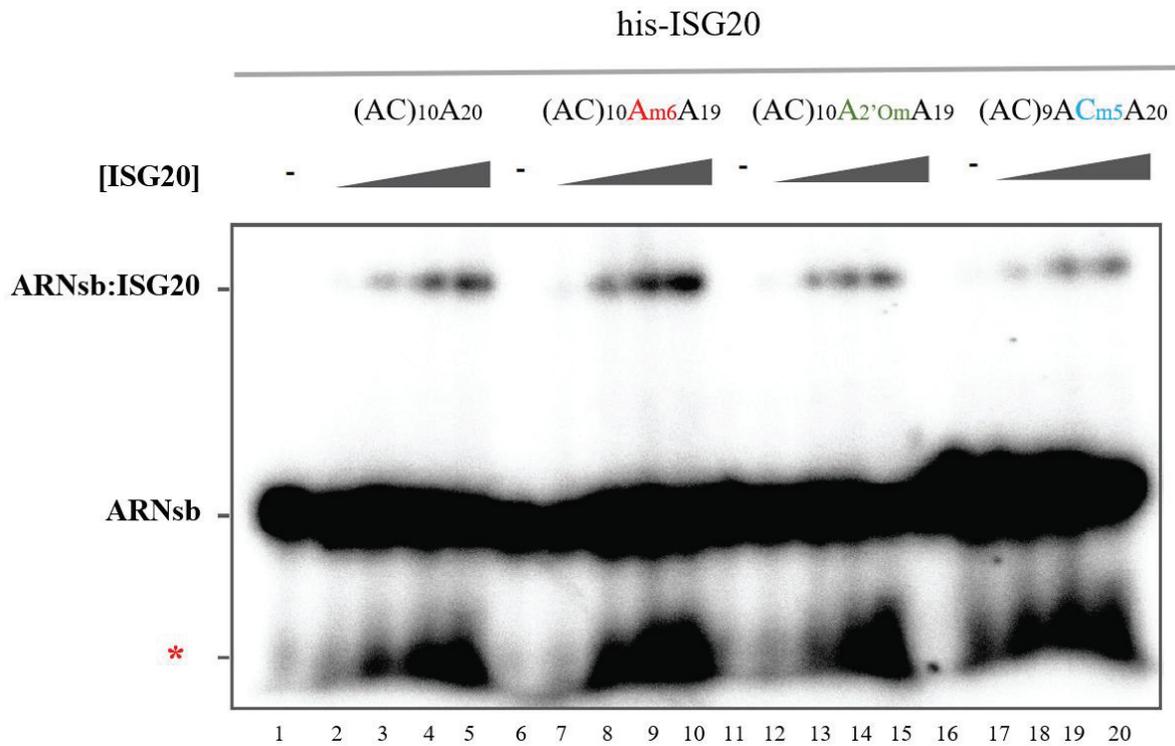
### X.1.1. Détermination de l'interaction d'ISG20 avec les ARN méthylés

Avant de réaliser les tests de digestion avec les substrats ARN méthylés, j'ai vérifié que la méthylation n'affecte pas la liaison d'ISG20 à l'ARN. Pour cela, j'ai réalisé des tests en incubant différents ARN linéaires AC-PolyA non méthylés ou modifiés par les méthylation suivantes : m6A, 2OA ou m5C (**figure 83**). L'incubation est réalisée avec ISG20 à différents ratios molaires ARN:protéine dans un tampon de digestion contenant du ZnSO<sub>4</sub> au lieu de MnCl<sub>2</sub>, comme décrit dans le chapitre Matériel et Méthodes (voir **chapitre XII.5.1.5**).



**Figure 83.** Les différents ARN linéaires méthylés utilisés dans cette étude. **A.** Représentation des structures chimiques des nucléosides méthylés. **B.** Représentation de la séquence nucléotidique des différents ARN méthylés et de la localisation de la méthylation sur la séquence.

Nous observons une digestion croissante de l'ARN linéaire en présence d'une concentration croissante d'ISG20 (**figure 84**). Nous observons également une liaison de l'ARN à ISG20 pour des ratios molaires ARN:protéine de 1:8, 1:16 et 1:32 (puits 3 à 5, 8 à 10, 13 à 15 et 18 à 20). Les quantités de complexe ARN:ISG20 et de fragments d'ARN digérés augmentent de façon proportionnelle avec la quantité d'ISG20. La liaison est observée pour tous les ARN (sans et avec méthylation), ce qui signifie que la méthylation n'empêche pas la liaison de l'ARN à ISG20.



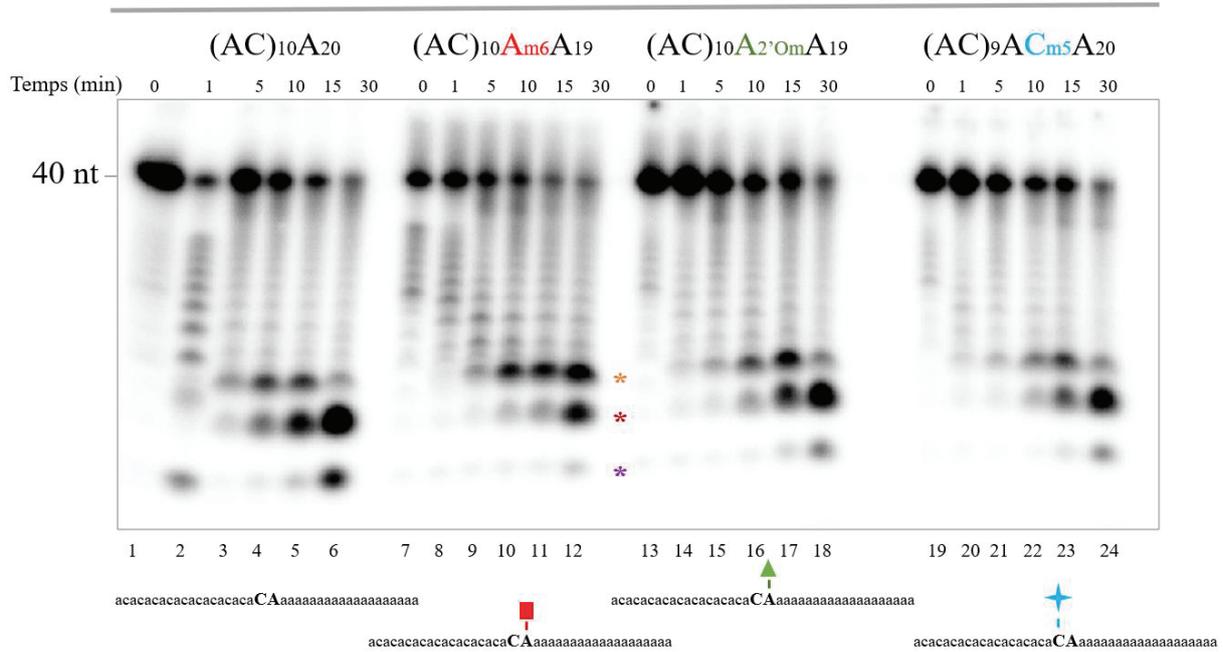
**Figure 84.** Expérience de retard sur gel montrant l'interaction de his-ISG20wt avec les ARNsb AC-PolyA non-méthylé et méthylés radiomarqués en présence de Zn<sup>2+</sup>. L'étoile rouge correspond aux produits de digestion.

#### X.1.2. Test de digestion avec les ARN méthylés

J'ai étudié l'effet des méthylation de l'ARN sur la digestion médiée par ISG20 en réalisant des tests de digestion avec des ARN présentant ou non (AC-polyA) différentes méthylation (m6A, 2OA ou m5C). La réaction de digestion est réalisée à 37°C et des prélèvements sont réalisés à différents temps pendant 30 minutes (voir **chapitre XII.5.3.2**). Ils sont analysés par une électrophorèse en condition dénaturante.

En comparant la cinétique de digestion des différents ARN, nous observons des produits de digestion de tailles similaires mais à des intensités différentes entre l'ARN non-méthylé et les ARN méthylés (**figure 85** ; étoiles violette, rouge et orange). En effet dans le cas de l'ARN non-méthylé, au bout de 30 minutes de réaction, le produit de digestion de plus petite taille (étoile violette) est en plus grande quantité (puits 6) en comparaison aux ARN méthylés (puits 12, 18 et 24). Le produit de digestion de plus petite taille est en plus faible quantité dans le cas de l'ARN avec la méthylation m6A. De plus dans le cas des ARN méthylés, après 15 minutes de réaction, le produit de digestion majoritaire (étoile orange) est de plus grande taille que dans le cas de l'ARN non-méthylé (étoile rouge). Ces résultats suggèrent que les méthylation de l'ARN de type m6A, 2OA et m5C diminuent l'activité d'ISG20 avec une diminution plus importante dans le cas de la méthylation m6A.

## his-ISG20



**Figure 85.** La méthylation de l'ARN diminue la digestion médiée par ISG20wt. L'activité exonucléase d'ISG20 sur des ARN non-méthylé ou méthylés a été étudiée pendant 30 minutes et contrôlée par une électrophorèse sur gel de polyacrylamide en condition dénaturante. Les étoiles violette, rouge et orange correspondent à des produits de digestion.

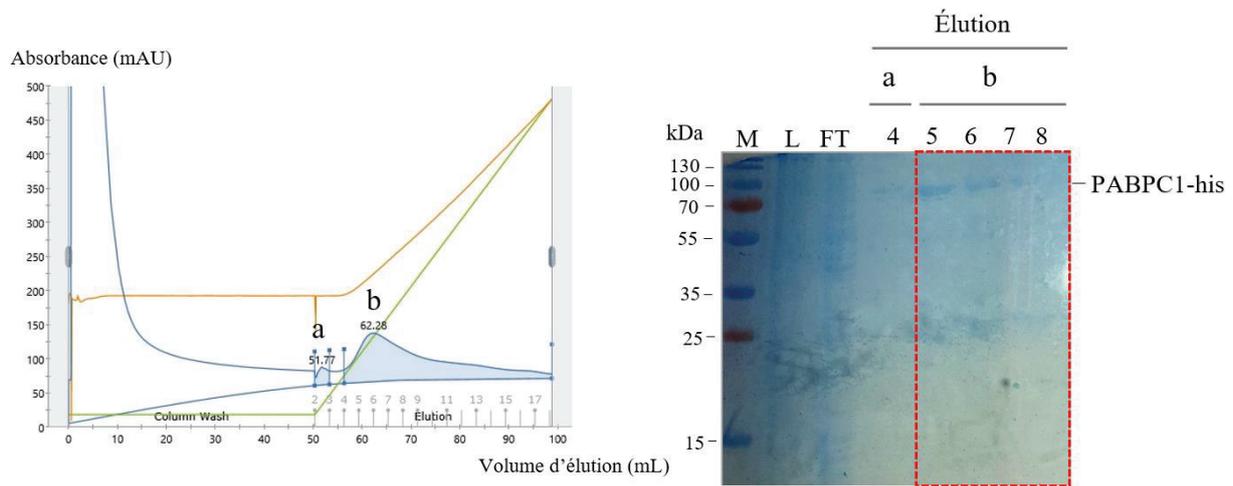
Cependant pendant que je réalisais ces travaux sur les méthylations de l'ARN, une étude a été publiée montrant l'effet de la 2'O-méthylation sur l'activité d'ISG20, avec une étude des résidus d'ISG20 impliqués dans la reconnaissance de la méthylation (El Kazzi *et al.*, 2022). Suite à cela, je me suis concentrée sur deux autres modes de protection de l'ARN qui sont la structure et l'addition d'une extrémité PolyA en 3'.

### X.2. La protection par une extrémité 3' polyadénylée

Tous les ARN cellulaires à l'exception de quelques-uns présentent une extrémité PolyA en 3' (Shatkin et Manley, 2000). Cette extrémité participe à différentes étapes du métabolisme des ARN, comme la régulation de la traduction ou la protection des ARN de la dégradation médiée par des exonucléases (Dreyfus et Régner, 2002). Cette protection se fait par le recrutement de différents facteurs protéiques dont les PABP (Mangus *et al.*, 2003). Pour cela, j'ai étudié l'effet de la liaison d'une protéine sur une extrémité 3'PolyA sur l'activité d'ISG20.

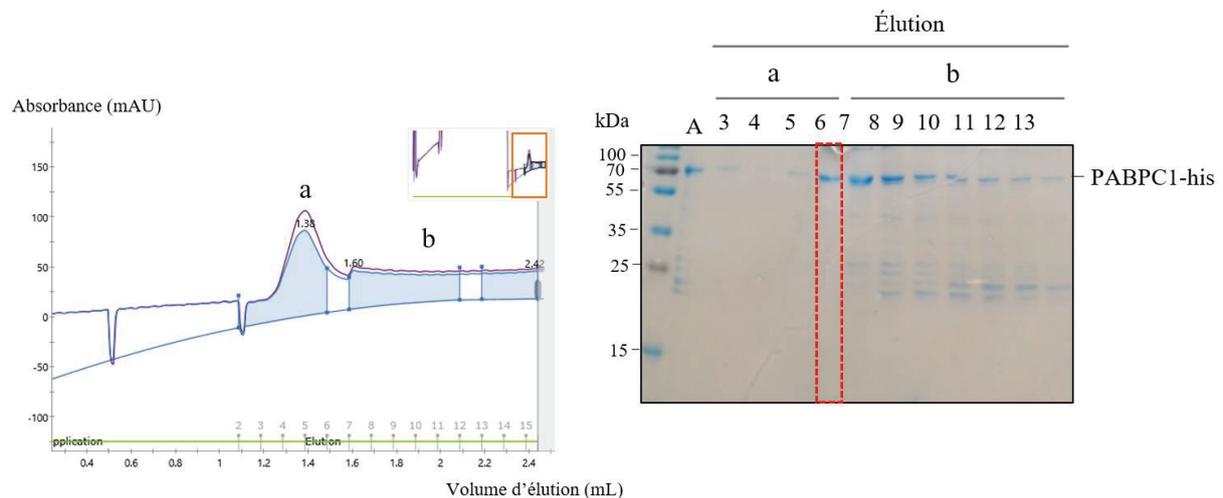
#### X.2.1. Purification de la protéine recombinante PABPC1-his

La protéine PABPC1-his est purifiée en deux étapes de chromatographie, la première est une chromatographie IMAC (voir **chapitre XII.1.7**, Matériel et Méthodes). Deux pics d'éluion sont observés (pics a et b) sur le chromatogramme (**figure 87**, à gauche). Une analyse par SDS-PAGE montre que le second pic correspond à l'éluion des protéines d'intérêt à une masse moléculaire de 70 kDa.



**Figure 87.** *A gauche*, chromatogramme de la purification de PABPC1-his sur une colonne HisTrap HP 5 mL. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'élution en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et celle en violet à la mesure de l'absorbance à 260 nm. Le tracé vert correspond au pourcentage de tampon d'élution et le tracé orange correspond à la conductivité. *A droite*, SDS-PAGE 15% sur les fractions de la purification de PABPC1-his sur colonne HisTrap (M : marqueur de masses moléculaires ; L : lysat ; FT : fraction non retenue). Les fractions encadrées en rouge sont celles qui ont été conservées pour la suite de la purification. Les numéros des fractions d'élution contrôlées sont précisés sur le gel.

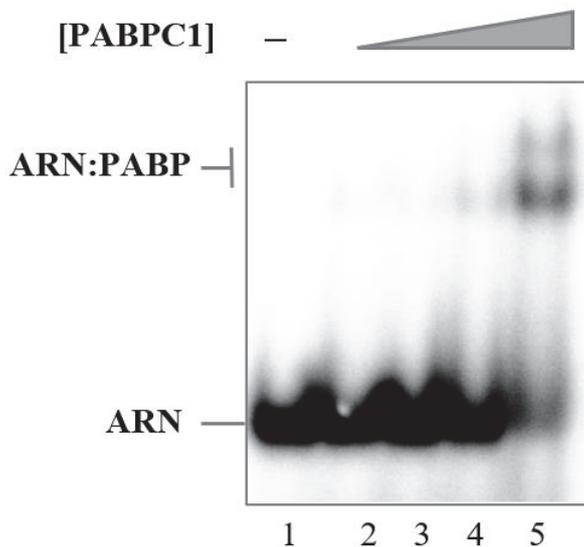
La seconde étape de purification de PABPC1-his se fait sur une colonne Superdex 75 Increase 5/150 GL par chromatographie SEC. Les deux pics observés sur le chromatogramme (**figure 88**, à gauche) sont caractérisés par SDS-PAGE (**figure 88**, à droite). PABPC1-his est éluée dans la seconde partie du premier pic (pic a) sans contaminant et dans le second pic (pic b) avec des contaminants. La fraction 6 d'élution (encadré rouge) qui correspond à PABPC1-his pure est éluée à 1,55 mL, donc PABPC1-his est sous forme monomérique en solution (**Annexes, figure 95**).



**Figure 88.** *A gauche*, chromatogramme de la purification de PABPC1-his sur une colonne Superdex 75 5/150 GL. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'élution en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et celle en violet à la mesure de l'absorbance à 260 nm. Le tracé vert correspond au pourcentage de tampon SEC. *A droite*, SDS-PAGE 15% sur les fractions de la purification SEC de PABPC1-his (M : marqueur de masses moléculaires ; A : échantillon avant l'injection sur la colonne Superdex 75 10/300 GL). Les fractions encadrées en rouge sont celles qui ont été conservées pour la suite des expériences. Les numéros des fractions d'élution contrôlées sont précisés sur le gel.

### X.2.2. Détermination de l'interaction de PABPC1 avec l'ARN AC-PolyA

Une expérience de retard sur gel est réalisée au préalable, afin de s'assurer de l'interaction entre la PABPC1 et l'ARN AC-PolyA et d'optimiser les conditions d'interaction. L'ARN AC-PolyA est incubé avec la PABPC1 pendant 15 minutes à 37°C selon les conditions détaillées dans le chapitre Matériel et Méthodes (voir **chapitre XII.5.1.6**). PABPC1-his lie l'ARN AC-PolyA pour des ratios molaires ARN:protéine compris entre 1:20 et 1:50 dans les conditions testées.



**Figure 89.** PABPC1 lie l'ARN AC-PolyA. Expérience de retard sur gel montrant l'interaction de la PABPC1 avec un ARNs radiomarqué au  $^{32}P$ .

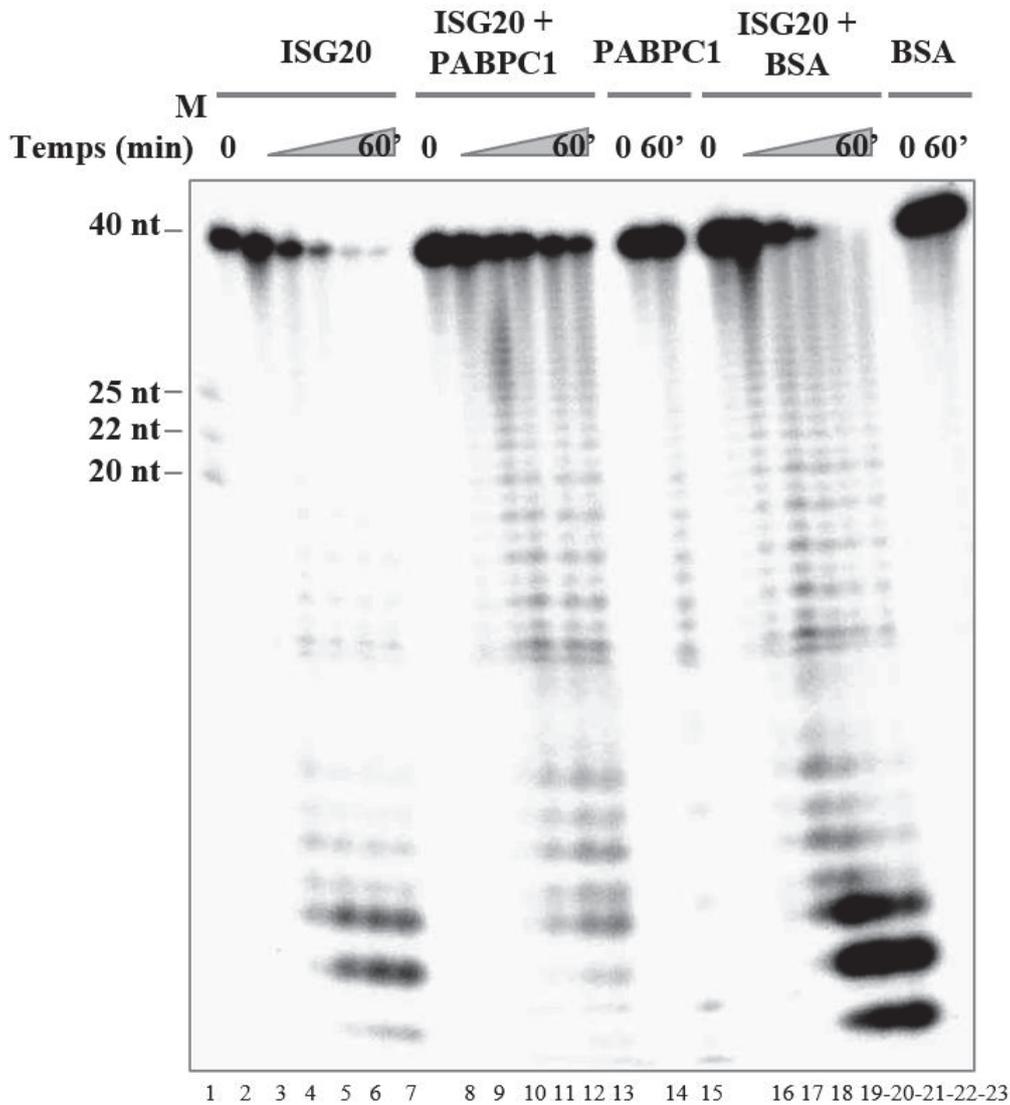
### X.2.3. Test de protection par la PABPC1

Des tests de digestion sont réalisés par la suite, en pré-incubant l'ARN AC-PolyA avec la PABPC1 ou la BSA pendant 20 minutes à 37°C. Puis ISG20 est ajouté pour initier la digestion. La réaction de digestion est réalisée à 37°C et des prélèvements sont réalisés à différents temps pendant 60 minutes. Une analyse par électrophorèse est réalisée en condition dénaturante sur un gel de séquence, comme décrit dans le **chapitre XII.5.3.3** du Matériel et Méthodes.

En absence de PABPC1 et de BSA, nous pouvons voir que l'ARN est presque totalement digéré à partir de 30 minutes (**figure 90**, puits 6 et 7). En comparaison, dans le cas d'une préincubation de l'ARN avec la PABPC1, l'ARN est moins digéré qu'en absence de PABPC1. La préincubation de l'ARN avec la PABPC1 le protège de la digestion par ISG20. Cette protection n'est pas observée dans le cas d'une préincubation avec la BSA. Au contraire, la digestion médiée par ISG20 est améliorée en présence de BSA, une hypothèse serait que la BSA procure un environnement plus visqueux qui est favorable à l'activité d'ISG20 comme démontré par une meilleure conservation de l'activité de la protéine au cours du temps lors d'un stockage dans 50% de glycérol.

Les deux contrôles, où l'ARN est incubé avec la PABPC1 ou la BSA sans ISG20 (puits 14 et 15, 23 et 24), permettent de s'assurer que la digestion observée est médiée par ISG20.

Ces résultats suggèrent qu'une extrémité 3'-polyadénylée permettant la liaison d'une PABP protège l'ARN de la digestion médiée par ISG20. Ce moyen de protection permet de protéger les ARN cellulaires, notamment les ARNm, lors d'une réponse IFN et permet également aux ARN viraux polyadénylée d'échapper à l'activité antivirale d'ISG20.



**Figure 90.** La liaison de la PABPC1 à l'ARN, le protège de la dégradation par ISG20. L'activité exonucléase d'ISG20 sur des ARNs radiomarqués seuls ou pré-incubés avec la PABPC1 ou la BSA a été étudiée pendant 1h et contrôlée par une électrophorèse sur gel de polyacrylamide en condition dénaturante. M : marqueur de taille.

### X.3. La protection par la structure

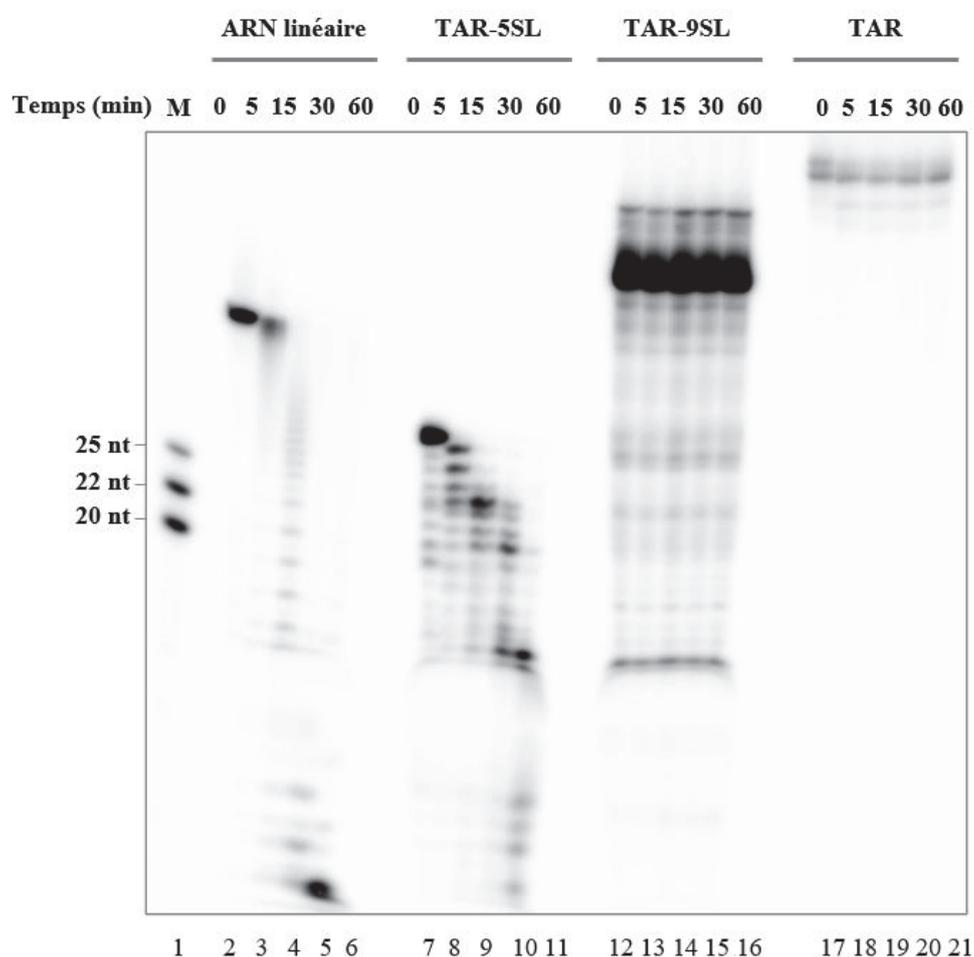
Les deux mécanismes étudiés précédemment, les modifications épitranscriptomiques et l'extrémité PolyA, sont des caractéristiques importantes dans le métabolisme des ARN cellulaires et en particulier des ARNm. Ces deux mécanismes peuvent être détournés par des virus pour se protéger de l'activité antivirale d'ISG20 (El Kazzi *et al.*, 2022). De plus, d'après les résultats que j'ai obtenu sur le mode d'action d'ISG20, nous avons pu voir qu'ISG20 lie mais ne digère pas l'ARN TAR-9SL lors d'une expérience de retard sur gel. Nous avons donc proposé l'hypothèse que la structure de l'ARNg viral pourrait le protéger de la digestion médiée par ISG20. Pour cela, j'ai étudié la résistance des ARN structurés à l'activité d'ISG20 en réalisant un test de digestion.

J'ai incubé des ARN présentant des degrés de structuration différents avec ISG20 (voir **chapitre XII.5.3.4**, Matériel et Méthodes). La réaction de digestion est réalisée à 37°C et des prélèvements sont effectués à différents temps pendant 60 minutes. Ils sont ensuite analysés par une électrophorèse en condition dénaturante sur un gel de séquence.

En comparant le profil de digestion entre les différents ARN, nous pouvons voir que l'ARN linéaire est totalement digéré au bout de 60 minutes de réaction (**figure 86**, puits 6). En comparaison avec l'ARN linéaire et l'ARN TAR-5SL, les ARN TAR-9SL et TAR ne sont pas digérés au cours de la réaction (puits 12 à 21). Ils sont ainsi résistants à l'action d'ISG20, ce qui n'est pas le cas de l'ARN TAR-5SL (puits 7 à 11). Ce résultat suggère qu'une tige de 9 pb est suffisante pour résister à l'activité d'ISG20.

De plus, en comparant la digestion de l'ARN linéaire à celle de l'ARN TAR-5SL : nous voyons que l'ARN TAR-5SL est digéré moins rapidement que l'ARN linéaire. Après 5 minutes de réaction, une échelle de 5 nucléotides est présente sur le gel. Elle correspond à la tige basale de l'ARN TAR-5SL (**figure 86**, puits 8 ; **figure 73.A**). Après 30 minutes de réaction, nous pouvons voir la digestion de l'hexaboucle de TAR 5-SL (bande intense de plus grande migration) et la digestion de la tige apicale de TAR-5SL (bande intense de plus faible migration). L'ARN TAR-5SL est totalement digéré après 1h de réaction.

En comparant la distance de migration entre les ARN linéaire et structurés, nous observons une migration aberrante des ARN structurés, par exemple l'ARN TAR-9SL de 34 nucléotides de long migre moins que l'ARN linéaire de 40 nucléotides. Cela peut s'expliquer par le fait que les structures ne sont peut-être pas complètement dénaturées dans ces conditions.



**Figure 86.** ISG20wt ne digère pas un ARN structuré avec une tige supérieure à 9 pb. L'activité exonucléase d'ISG20 sur des ARN linéaire ou structurés a été étudiée pendant 1h et contrôlée par une électrophorèse sur gel de polyacrylamide en condition dénaturante. M : marqueur de taille.

## Chapitre 5 : Discussion et perspectives (seconde partie)

L'objectif était de mieux comprendre le mode d'action d'ISG20 et notamment les déterminants de l'ARN qui expliqueraient la résistance de certains ARN à la dégradation par ISG20. J'ai travaillé sur des éléments cis et trans localisés à l'extrémité 3' des ARN. Mes résultats montrent que l'extrémité 3' PolyA qui permet la liaison de la PABP et la structure de l'ARN, un certain degré de structuration comme l'ARN TAR-9SL, protègent l'ARN de la dégradation médiée par ISG20. L'ARN TAR-5SL, présentant une tige-boucle de 5 pb, est digéré par ISG20, ce résultat a également été observé dans une autre étude (El Kazzi *et al.*, 2022). Dans cette étude, ils ont mis en évidence la protection de l'ARN par des modifications 2OA (El Kazzi *et al.*, 2022). Les auteurs démontrent que la digestion médiée par ISG20 s'arrête deux nucléotides en amont du site de la méthylation et au niveau de la méthylation suite à un clash stérique entre les résidus R53 et D90 d'ISG20 avec le nucléotide méthylé. Dans mes expériences avec des substrats ARN 2'O-méthylés, je n'observe pas cet arrêt en amont du site de méthylation et au site de méthylation. De plus, j'observe un effet protecteur de la modification m6A, ce qui n'est pas le cas dans les résultats publiés par El Kazzi *et al.* (2022). Ces différences peuvent s'expliquer par une préparation différente de l'ARN méthylé, synthétique ou préparé par transcription *in vitro*, ce qui peut influencer la précision du site de méthylation.

ISG20 inhibe la réplication d'un grand nombre de virus (Deymier *et al.*, 2022), mais le mécanisme moléculaire antiviral est peu compris. Il a été proposé que son activité antivirale soit liée à la dégradation directe de l'ARN viral, mais il a aussi été montré qu'ISG20 inhibe la traduction de l'ARN viral sans le dégrader directement (Weiss *et al.*, 2018 ; Wu *et al.*, 2019). Nos résultats pourraient expliquer ces différents modes d'action décrits pour ISG20 : les ARN viraux non-polyadénylés ou qui ne lient pas de protéine ou encore qui ne présentent pas de structure à l'extrémité 3' sont sensibles à l'activité d'ISG20 et seront directement dégradés. Au contraire les ARN polyadénylés ou liés à une protéine ou fortement structurés à l'extrémité 3' seront résistants à l'activité d'ISG20 et alors l'activité antivirale d'ISG20 sera médiée par une inhibition de la traduction en dégradant par exemple un ARN cellulaire. Ce mécanisme est vérifié pour le virus de la dengue qui présente une extrémité 3' très structurée (Murugesan et Manoharan, 2020) et qui est inhibé par ISG20 dans la seconde phase de l'infection virale, c'est-à-dire dans les étapes de traduction et de réplication (Jiang *et al.*, 2010). D'autres exemples supportent ce mécanisme comme le VSV (Espert *et al.*, 2003 ; Wu *et al.*, 2019), le virus chikungunya (Weiss *et al.*, 2018) et le VIH-1 (Espert *et al.*, 2005) qui ont un ARN « protégé » polyadénylé à l'extrémité 3' (Hwang *et al.*, 1998 ; Khan *et al.*, 2002 ; Karn et Stoltzfus, 2012) et dont la traduction est inhibée par ISG20. Curieusement, l'ARN du virus de l'hépatite B qui présente une extrémité PolyA en 3' est décrit comme étant dégradé par ISG20 (Imam *et al.*, 2020). Toutefois, un autre mécanisme est décrit par les auteurs : ils démontrent que l'ARN présente une modification m6A qui est reconnue par ISG20 en présence d'un partenaire cellulaire et induit la dégradation de l'ARN par ISG20 (Imam *et al.*, 2020). La combinaison de plusieurs déterminants semble avoir des effets différents sur l'activité d'ISG20 selon le virus, en effet El Kazzi *et al.* (2022) ont montré par des études *in vitro* et *in vivo* que la modification 2OA sur une extrémité 3' PolyA protège davantage qu'une simple modification 2OA en 3' dans le cas de l'ARNg du VIH-1. Il serait intéressant d'analyser les virus qui sont décrits comme résistants à ISG20 et d'identifier les déterminants en 3' de l'ARN pour étudier si ces modes de protection sont modulables en fonction du type de virus. Il serait également intéressant d'étudier l'effet de la liaison de nucléocapsides virales en 3' de l'ARNg viral sur l'activité antivirale d'ISG20.

De plus de manière intéressante, j'ai mis en évidence *in vitro* qu'ISG20 lie l'ARN structuré sans le dégrader. Ce résultat suggère un autre mode d'action d'ISG20 et il serait intéressant d'étudier si l'interaction d'ISG20 avec ces ARN « protégés » joue un rôle dans le devenir de ces ARN, par exemple en entraînant une gêne stérique ou encore un changement de localisation subcellulaire. Le mutant catalytique D94 d'ISG20 entraîne une perte complète de l'activité exonucléase d'ISG20 *in vitro* et *in vivo* (Wu *et al.*, 2019). De manière intéressante, j'ai découvert que cette mutation entraîne une perte de liaison d'ISG20 à l'ARN. Par la suite, ce mutant pourrait être exploité pour étudier l'effet de la perte de liaison d'ISG20 dans les cellules, ce qui permettrait de mieux comprendre l'activité antivirale d'ISG20. Quelques résidus d'ISG20 ont été décrits comme étant impliqués dans la reconnaissance du substrat ARN. Les résidus R53 et M14 interagissent directement avec le groupement 2'-OH du ribose dans la structure cristalline d'ISG20 en complexe avec de l'UMP suggérant un rôle de ces résidus dans la reconnaissance de l'ARN (Horio *et al.*, 2004). De plus, les travaux de El Kazzi *et al.* (2022) ont identifié les résidus R53 et D90 comme étant impliqués dans la reconnaissance de la modification 2OA.

J'ai mis en évidence qu'ISG20 est une enzyme distributive. Ce mode d'action peut être relié à une autre activité d'ISG20 décrite dans la littérature qui est la régulation des microARN (Sim *et al.*, 2022). Il a été montré *in vitro* et *in vivo* qu'ISG20, comme d'autres exonucléases 3'-5' et notamment SDN1 (« *small RNA degrading nuclease 1* ») (Chen *et al.*, 2018), permet la régulation génique en raccourcissant des microARN de 22 nucléotides de long à une longueur de 14 nucléotides. ISG20 pourrait donc réaliser une digestion partielle de l'ARN viral et conduire à une régulation de la traduction, ou réguler l'expression d'autres facteurs impliqués dans la réponse antivirale.

La structure des deux premiers domaines de SDN1 en complexe avec un substrat ARN et celle du domaine C-terminal, résolues par cristallographie aux rayons X, montrent que la protéine contient un domaine semblable aux exonucléases 3'-5' et un domaine C-terminal qui contient un motif de reconnaissance de l'ARN (RRM, « *RNA recognition motif* ») (Chen *et al.*, 2018). Des études *in vitro* montrent que ces deux domaines agissent coopérativement pour la reconnaissance d'un substrat ARN et pour permettre de le lier correctement : avec le domaine DEDDh qui digère l'extrémité 3' d'un ARNsb, tandis que le domaine RRM lie la région 5' de l'ARN. Ce domaine RRM est critique pour la processivité de SDN1 vis-à-vis de substrats ARNsb et lorsqu'il est absent, SDN1 est une protéine distributive. Il serait intéressant d'étudier l'interaction d'ISG20 avec d'autres protéines cellulaires contenant des domaines RRM et de vérifier si cette interaction peut rendre ISG20 processive et moduler son activité exonucléase dans un contexte d'infection virale.

## Chapitre 6 : Conclusion générale

L'ensemble de mes travaux de thèse menés sur l'IN et ISG20 souligne l'importance de l'ARN viral tout au long du cycle de vie du VIH-1. En effet, cet ARN n'est pas qu'un « simple » génome, mais il est un véritable régulateur de la réplication et de la propagation virale. Il évolue au contact de facteurs cellulaires et viraux, permettant d'un côté de détourner les machineries cellulaires au bénéfice de la réplication virale et de l'autre de résister aux mécanismes de défense cellulaires.

Mes résultats sur l'IN suggèrent un rôle de cette protéine dans la transcription provirale en interagissant avec l'ARN TAR et le complexe P-TEFb impliqué dans l'élongation de la transcription. Cela ouvre une perspective pour la stratégie « *block and lock* » (blocage et verrouillage) étudiée en ce moment pour le VIH-1. Elle repose sur le développement d'inhibiteurs de la transcription du provirus, qui permettent de conserver le VIH-1 sous forme latente (Vansant *et al.*, 2020 ; Da Rocha *et al.*, 2022). Ceux-ci ciblent la protéine rétrovirale Tat et le complexe P-TEFb cellulaire (Rice, 2017 ; Asamitsu *et al.*, 2018 ; Vansant *et al.*, 2020).

De plus, les virus peuvent évoluer en développant des stratégies pour détourner les enzymes cellulaires responsables des modifications de l'ARN. Dans le cas du VIH-1, des méthyltransférases cellulaires sont utilisées pour modifier son génome et le protéger de la dégradation médiée par ISG20 (El Kazzi *et al.*, 2022). Mon étude met en évidence deux autres mécanismes de protection de l'ARN contre l'activité de ce facteur de restriction. Le premier est la polyadénylation et le recrutement de la PABP, et le second est la forte structuration de l'extrémité 3'.

## Chapitre 7 : Matériel et Méthodes

Le matériel et méthodes qui concerne les expériences détaillées dans l'article de recherche sur l'IN n'est pas détaillé dans ce chapitre, il se trouve dans la partie matériel et méthodes de l'article en question.

### XI. Techniques de Biologie Moléculaire

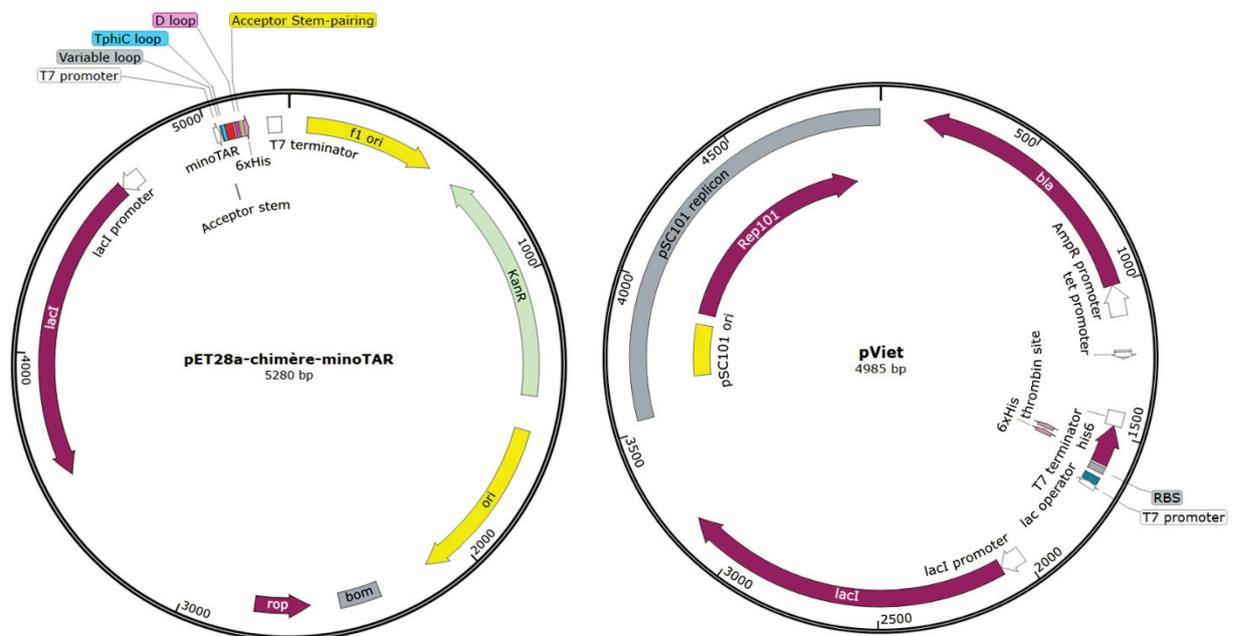
#### XI.1. Plasmides

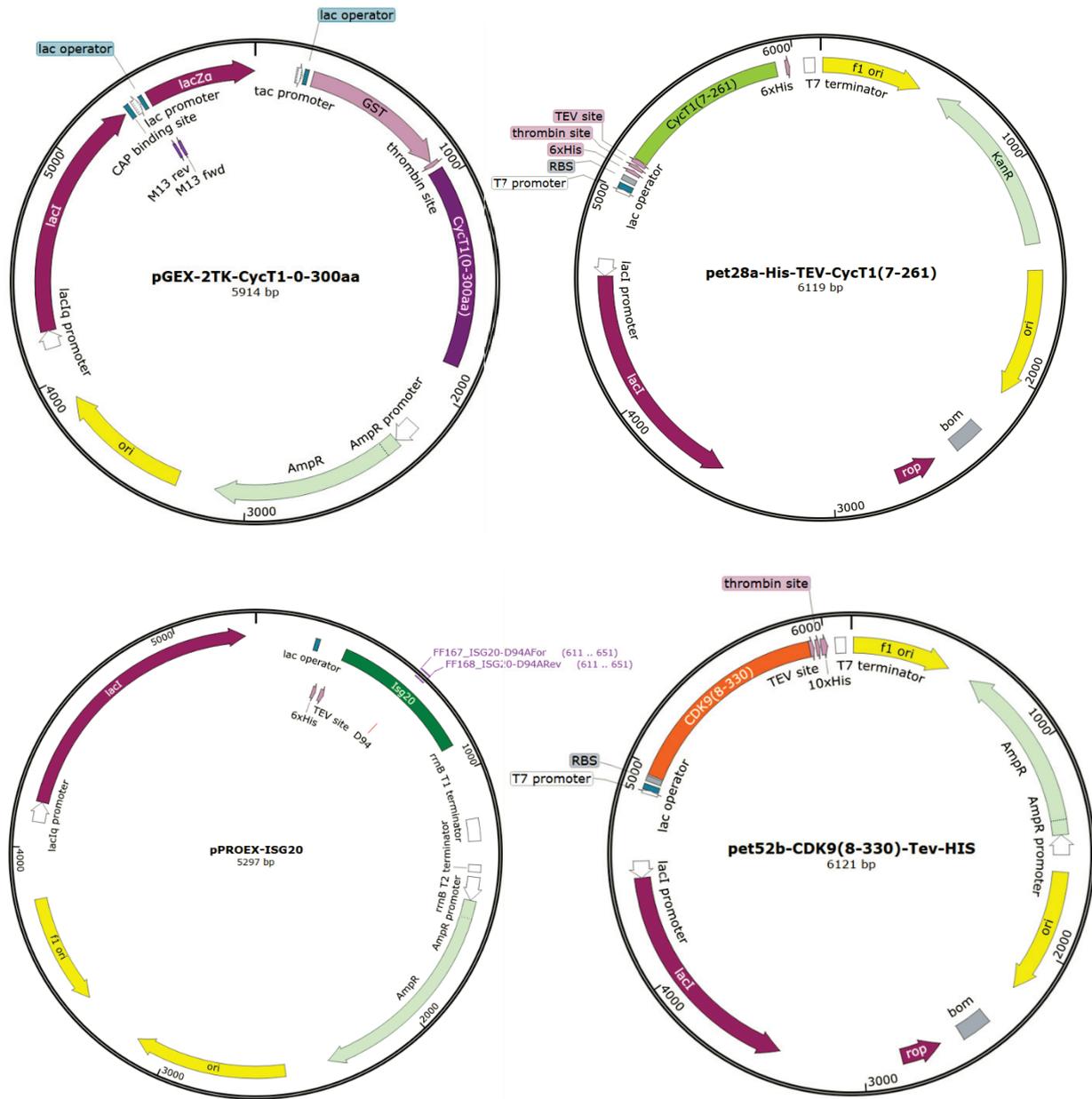
Pour ce travail de thèse, j'ai produit plusieurs protéines recombinantes. Les plasmides utilisés pour la production de ces protéines ont été donné par des collaborateurs (**figure 87**) :

**Projet IN :** Les plasmides pViet-IN-CTD- $\Delta$ CT (aa 220-270) et pViet-IN-CTD (aa 220-288) ont été donné par Marc Ruff (IGBMC, Strasbourg) et ont permis de produire respectivement les protéines his-IN-CTD- $\Delta$ CT et IN-CTD. Les plasmides pAC8-CDK9-cycline T1 (Kolesnikova *et al.*, 2022), pGEX-2TK-cycline-T1 (aa 1-300) ont été donné par Olivier Bensaude (IBENS, Paris) et permettent de produire respectivement les protéines GST-cycline T1, his-cycline T1 et CDK9-his à partir des constructions pET28a-cycline-T1 (aa 7-261) et pET52b-CDK9 (aa 8-330) conçues dans l'équipe.

**Projet ISG20 :** Le plasmide pPROEX-ISG20 (aa 1-181) a été donné par Andrea Cimarelli (CIRI, Lyon) et le plasmide pET-PABPC1 (aa 1-636) (Mure *et al.*, 2018) a été donné par Fabrice Mure (CIRI, Lyon). Ces deux plasmides ont permis de produire respectivement les protéines his-ISG20*wt* et PABPC1-his. De plus, à partir du plasmide pPROEX-ISG20, une mutagenèse dirigée par réaction en chaîne de la polymérase (PCR) a été réalisée dans l'équipe afin de produire par la suite la protéine recombinante his-ISG20 D94A.

Pour le projet IN, le plasmide pET28a-chimère-minoTAR a été conçu au laboratoire et permet d'exprimer l'ARN chimérique minoTAR.





**Figure 56.** Les cartes plasmidiques des constructions utilisées au cours de mes travaux de thèse.

Des étiquettes d'affinité sont fusionnées au gène d'intérêt pour faciliter la purification de la protéine d'intérêt (tableau 2).

Construction	Vecteur d'expression	Étiquette de purification (position)	Site de clivage	Antibiotique
Cycline T1 (aa 1-300)	pGEX-2TK	GST (N-terminale)	Site thrombine LVPR GS	Ampicilline
Cycline T1 (aa 7-261)	pET28a	6-histidine (N-terminale)	Site thrombine LVPR GS Site TEV ENLYFQ G	Kanamycine
CDK9 (aa 8-330)	pET52b	10-histidine (C-terminale)	Site TEV ENLYFQ G	Ampicilline

			Site thrombine LVPR GS	
ISG20 <sub>wt</sub> (aa 1-181)	pPROEX	6-histidine (N-terminale)	Site TEV ENLYFQ G	Ampicilline
ISG20 D94A (aa 1-181)	pPROEX	6-histidine (N-terminale)	Site TEV ENLYFQ G	Ampicilline
minoTAR (79 nucléotides)	pET28a			Kanamycine

**Tableau 2.** Caractéristiques des constructions plasmidiques utilisées au cours de mes travaux de thèse. TEV, « tobacco etch virus » ou protéase du virus de la gravure du tabac.

## XI.2. Transformation des bactéries

### XI.2.1. Protocole classique

La souche *E. coli* BL21(DE3) Rosetta/pLysS (Novagen, Sigma Aldrich) est transformée avec le plasmide d'intérêt la veille au soir par la méthode du choc thermique : 50 µL de cellules *E. coli* avec 100 ng d'ADN sont mis 20 minutes dans la glace. Puis un choc thermique est réalisé 2 minutes à 42°C suivie d'une incubation 5 minutes dans la glace. Par la suite, 600 µL de milieu SOC (tryptone 2%, extrait de levure 0,5%, NaCl 10 mM, KCl 2,5 mM, MgCl<sub>2</sub> 10 mM, MgSO<sub>4</sub> 10 mM et glucose 20 mM) sont ajoutés. Après 1h à 37°C en rotation à 450 rpm, une centrifugation pendant 2 minutes à 775 g est réalisée. Le surnageant est éliminé de moitié, la moitié conservée est utilisée pour resuspendre le culot qui est ensuite étalé sur une boîte de Pétri LB – antibiotique sur la nuit à 37°C. Dans le cas de l'ampicilline, la concentration finale est de 50 µg/mL et dans le cas de la kanamycine, elle est de 25 µg/mL.

La composition du milieu LB est la suivante : tryptone 1%, extrait de levure 0,5% et NaCl 170 mM.

### XI.2.2. Protocole pour les protéines dont l'expression est toxique dans *E. coli*

L'expression de his-ISG20<sub>wt</sub> est toxique dans *E. coli*, pour cela j'ai optimisé le protocole précédent par la méthode du choc thermique en réalisant une incubation pendant 2h à 30°C au lieu d'1h à 37°C. De plus, la boîte de Pétri LB – antibiotique est incubée sur deux nuits à 30°C au lieu d'une incubation sur la nuit à 37°C.

## XI.3. Extractions et purifications des acides nucléiques

### XI.3.1. Extraction et purification de l'ADN plasmidique sur colonne

Plusieurs colonies d'une boîte de Pétri de transformation sont isolées et re-suspendues dans 8 mL de LB – antibiotique. Les différentes préparations sont incubées à 37°C toute la nuit. Le lendemain, l'ADN plasmidique est extrait selon le kit et le protocole « *Nucleospin isolation of low-copy plasmids* » (Machery-Nagel). L'ADN plasmidique pur est élué dans un tampon contenant du Tris-HCl pH 8,5 à 5 mM. Les échantillons sont conservés à -20°C.

### XI.3.2. Extraction de l'ARN par traitement au PCI puis précipitation à l'éthanol

L'ARN est extrait par une solution phénol:chloroforme:alcool isoamylique (PCI), en rapport volume/volume de 25:24:1, saturée avec du Tris pH 8 à 10 mM et de l'EDTA à 1 mM (Sigma-Aldrich). Le traitement par le PCI est réalisé deux fois : ajout d'1 volume de solution PCI, puis la solution est vortexée 1 minute, ensuite une centrifugation pendant 2 minutes à 20 817 g à 4°C est réalisée afin d'obtenir la séparation entre les phases organique et aqueuse. Finalement la phase aqueuse contenant l'ARN est récupérée. L'ARN est ensuite précipité en ajoutant 0,1 volume d'acétate de sodium à 3 M et 0,02 g/mL de glycogène, puis 3 volumes d'éthanol froid absolu sont ajoutés. Après agitation, l'échantillon est conservé 1h à -80°C. Ensuite, une centrifugation de 45 minutes à 20 817 g à 4°C est réalisée. Le surnageant est éliminé et l'ARN précipité est lavé avec de l'éthanol à 70%. L'éthanol est ensuite éliminé par une centrifugation de 5 minutes à 20 817 g à 4°C. Le culot d'ARN précipité est séché et l'ARN est re-suspendu dans 30 µL d'eau. L'ARN est conservé à -20°C.

### XI.4. La quantification des acides nucléiques

La concentration est mesurée avec un spectrophotomètre à microvolume NanoDrop™ 2000/2000c (Thermo Fisher Scientific) en réalisant des mesures d'absorbance à 260 nm (**tableau 3**). Par la suite, la qualité de l'ARN est contrôlée par une électrophorèse sur un gel d'acrylamide à 15% en condition dénaturante.

	<b>A 260 nm, 1 unité de DO correspond à :</b>	<b>Coefficient d'absorption molaire <math>\epsilon</math> (L/mol/cm)</b>
ADNdb	50 µg/mL	6 200
ADNsb	33 µg/mL	8 350
ARN	40 µg/mL	7 700

**Tableau 3.** Données permettant la quantification des acides nucléiques avec un spectrophotomètre.

### XI.5. Séparation analytique et préparative des ARN

L'électrophorèse sur gel de polyacrylamide pour séparer des échantillons d'ARN se fait dans un gel continu.

La composition du gel dénaturant est : 10% ou 15% d'acrylamide:bisacrylamide (19:1), Tris Borate EDTA (TBE) 1X, 6 M urée, 0,1% persulfate d'ammonium (APS) et 7 mM TEMED. Les échantillons préalablement au chargement sur gel sont chauffés 3 minutes à 98°C après ajout du tampon de charge à 1X final (formamide 95%, EDTA 18 mM, SDS 0,025%, xylène cyanol 0,01% et bleu de bromophénol 0,01%). La migration se fait entre 300 et 600V dans un tampon TBE 1X.

La composition du gel natif est : 6% d'acrylamide/bisacrylamide (19:1), Tris Acétate EDTA (TAE) 1X, 0,1% APS et 7 mM TEMED. Dans les échantillons, préalablement au chargement, du tampon de charge natif 1X final (30% glycérol, 0,01% xylène cyanol et 0,01% bleu de bromophénol) est ajouté. La migration se fait à 130 V dans le tampon TAE 0,5X.

La visualisation des bandes sur le gel de polyacrylamide se fait avec le bromure d'éthidium (BET) en utilisant un transilluminateur UV.

## XI.6. Conception des constructions plasmidiques

J'ai réalisé différentes constructions plasmidiques à partir de plasmides fournis (listés dans le **chapitre XI.1**) en réalisant des expériences de clonage et de mutagenèse *in vitro*. Les échantillons ont ensuite été séquencés selon la méthode de Sanger par la société Eurofins Genomics (service GATC, Ebersberg, Allemagne).

### XI.6.1. Constructions pET28a-cycline-T1 (aa 7-261) et pET52b-CDK9 (aa 8-330)

La première PCR permet d'amplifier la forme tronquée aa 7-261 de la cycline T1 à partir du plasmide pAC8-CDK9-cycline T1 ; de même pour la forme tronquée aa 8-330 de la CDK9. Pour la cycline T1, les oligonucléotides A1 et A2 sont utilisés, tandis que pour la CDK9 ce sont les oligonucléotides A3 et A4 (**tableau 4**). En pratique, pour la PCR : 10 ng de plasmide pAC8-CDK9-cycline T1 sont incubés avec 1  $\mu$ M d'oligonucléotides, 1X de tampon de la Q5 polymérase (New England BioLabs), 2 mM d'un mix de dNTP (Thermo Fisher Scientific) et 2 unités de Q5 polymérase (NEB). Ensuite avec un thermocycler GeneAmp® PCR System 9700 (Applied Biosystems) le programme suivant est réalisé : 5 minutes à 98°C ; 30 cycles de : 30 secondes à 98°C, 1 minute à 56°C et 2 minutes à 72°C ; puis 20 minutes à 72°C et l'échantillon est conservé à 4°C.

Le plasmide pET28a et la séquence ADN codant pour la cycline T1 (aa 7-261) amplifiée sont digérés par des enzymes de restriction en 5' et en 3' : 1  $\mu$ g de plasmide est mélangé avec 1  $\mu$ g d'insert dans 1X de tampon Cutsmart (NEB) en présence de 30 unités de NdeI (NEB) et 30 unités de EcoRI (NEB). L'échantillon est incubé pendant 1h à 37°C.

Le plasmide pET52b et la séquence ADN codant pour la CDK9 (aa 8-330) amplifiée sont digérés par des enzymes de restriction en 5' et en 3' : 1  $\mu$ g de plasmide est mélangé avec 1  $\mu$ g d'insert dans 1X de tampon Cutsmart (NEB) en présence de 15 unités de NcoI (NEB) et 15 unités de SacI (NEB). L'échantillon est incubé pendant 1h à 37°C.

Ensuite, les plasmides digérés sont traités pendant 1h à 37°C avec 20 unités de CIAP (« *Calf Intestinal Alkaline Phosphatase* ») (Thermo Fischer Scientific). Les plasmides sont ensuite traités par le PCI : ajout de 0,1 volume de NaAc 3 M et 0,02 g/mL de glycogène ; puis ajout de 1 volume de PCI. Les échantillons sont vortexés pendant 2 minutes à 4°C à 20 817 g. L'ADN présent dans le surnageant est précipité avec 3 volumes d'éthanol absolu froid pendant 1h à -20°C. Puis une centrifugation de 5 minutes à 20 817 g à 4°C permet d'éliminer le surnageant contenant l'éthanol. Le culot d'ADN précipité est séché et re-suspendu dans 30  $\mu$ L d'eau.

Une réaction de ligation est réalisée en mélangeant 50 ng de plasmide digéré seul ou avec 100 ng d'insert dans un tampon de ligation 1X (Thermo Fischer Scientific) et 5 unités de ligase T4 (Thermo Fischer Scientific). Les échantillons sont incubés sur la nuit à 16°C. Puis, 10  $\mu$ L d'échantillon sont utilisés pour transformer 100  $\mu$ L de bactéries compétentes TOP10 avec un gène de résistance à la kanamycine pour la cycline T1 et à l'ampicilline pour la CDK9. Selon le nombre de colonies présentes sur la boîte de transformation correspondant à l'échantillon sans insert et à celle correspondant à l'échantillon avec insert, un certain nombre de colonies sont sélectionnées sur la boîte de transformation avec insert. Une préculture est réalisée en inoculant 5 mL LB – antibiotique avec chacune des colonies sélectionnées sur la nuit à 37°C sous agitation. Ensuite, l'ADN est extrait selon le kit de Machery-Nagel, et il est séquencé.

Nom	Séquence (5' → 3')
A1	gccatattggaaaacctgtattttcagggcaacaacaacaacgggtggat
A2	ctcgaattcttagcatgccctccaattccaatgcg
A3	taccatggtggagtgccctttttgtg
A4	gcgagctcgcctgaaaatacaggttttcgggtggagagcatgcccttg

**Tableau 4.** Les oligonucléotides ADN utilisés pour produire les constructions plasmidiques pET28a-cycline-T1 (aa 7-261) et pET52b-CDK9 (aa 8-330).

#### XI.6.2. Mutagenèse dirigée par PCR : mutant ISG20 D94A

Une mutation non-conservative de l'aspartate 94 en alanine dans la séquence codante de l'ISG20 humaine (aa 1-181) avec une étiquette 6-histidine en N-terminale clonée dans le vecteur pPROEX a été introduite par PCR en utilisant les amorces A5 et A6. Les amorces sont énumérées dans le **tableau 5**.

Le protocole de mutagenèse dirigée est le suivant : 200 ng de plasmide, 0,5 µM de chaque oligonucléotide (A5 et A6), 10 mM d'un mix de dNTP, 1X de tampon de réaction pour la Q5 polymérase sont mélangés avec 2 unités de Q5 polymérase. Ensuite le thermocycler permet de réaliser le programme suivant : 5 minutes à 98°C ; 25 cycles de : 30 secondes à 98°C, 1 minute à 56°C et 7 minutes à 72°C ; puis 20 minutes à 72°C et l'échantillon est conservé à 4°C.

Le plasmide matrice contenant la séquence protéique d'ISG20<sub>wt</sub> est dégradé par l'enzyme DpnI à 2 unités pendant 4h à 37°C. 10 µL de réaction sont transformés par choc thermique dans 100 µL de cellules TOP10. Des colonies sont sélectionnées et une préculture est réalisée, puis l'ADN est extrait et séquencé.

Nom	Séquence (5' → 3')	Plasmides d'expression
A5	gggtcatgacctgaagcacgcgtccaggcactgaaagagg	pPROEX-ISG20
A6	cctctttcagtgctggaacgcgtgcttcaggctcatgaccc	pPROEX-ISG20

**Tableau 5.** Les oligonucléotides ADN utilisés pour produire la construction plasmidique pPROEX-ISG20 D94A.

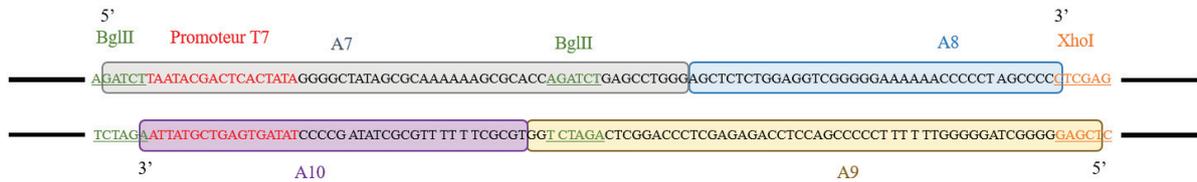
#### XI.6.3. Construction pET28a-chimère-minoTAR

L'objectif est d'obtenir une construction exprimant l'ARN TAR au sein d'une séquence d'un ARNt.

La première étape est l'hybridation des oligonucléotides : les quatre oligonucléotides (A7, A8, A9 et A10, **tableau 6**) à une concentration finale de 20 µM sont mélangés dans un tampon ligase 1X (NEB) et chauffés pendant 1 minute à 95°C puis un refroidissement lent jusqu'à une température de 16°C est réalisé (**figure 88**).

La deuxième étape est la digestion du plasmide par des enzymes de restriction : 200 ng de plasmide pET28a sont incubés avec 15 unités de BglII (NEB) et 30 unités de XhoI (NEB) dans un tampon RED 1X (NEB) sur la nuit à 37°C. Ensuite, l'échantillon est traité avec 20 unités de CIAP et le plasmide digéré est traité par le PCI.

Une réaction de ligation est réalisée en mélangeant 50 ng de plasmide digéré seul ou avec 200 ng ou 500 ng d'insert (hybride minoTAR avec les sites de restriction BglII/XhoI aux extrémités 5'/3') dans un tampon de ligation 1X et 5 unités de ligase T4. Les échantillons sont incubés sur la nuit à 16°C. Puis, 10 µL d'échantillon sont utilisés pour transformer 100 µL de bactéries compétentes TOP10 avec un gène de résistance à la kanamycine. A partir de ces colonies, une préculture est réalisée puis l'ADN est extrait et séquencé.



**Figure 88.** Représentations des quatre amorces chevauchantes (A7, A8, A9 et A10) utilisées pour permettre la conception de la construction exprimant un ARN chimérique minoTAR au sein du vecteur d'expression pET28a (traits noirs) en utilisant les enzymes de restriction BglII et XhoI.

Nom	Séquence (5' → 3')
A7	ggggctatagcgcaaaaaagcgcaccagatctgagcctgggagctctctggaggtcgggggaaaaaaccccctagcccc
A8	agctctctggaggtcgggggaaaaaaccccctagcccc
A9	tcgagggggctaggggggtttttcccccgacctccagagagctcccagggtcagatctgg
A10	tgcgctttttgcgctatagcccctatagtgagtcgtattaa

**Tableau 6.** Les oligonucléotides ADN utilisés pour produire la construction plasmidique pET28a-chimère-minoTAR.

## XII. Etudes fonctionnelles

### XII.1. Surproduction et purification des protéines de fusion

La composition des tampons de purification est détaillée dans la partie Annexes (**tableau 16**).

#### XII.1.1. his-IN-CTD (aa 220-288) et his-IN-CTD- $\Delta$ CT (aa 220-270)

Les protocoles d'expression et de purification des protéines recombinantes his-IN-CTD (aa 220 – 288) et his-IN-CTD- $\Delta$ CT (aa 220-270) utilisées pour les expériences détaillées dans l'article de recherche (Rocchi, Louvat *et al.*, 2022) sont données dans le matériel et méthodes de l'article.

#### XII.1.2. GST-cycline T1 (aa 1-300)

Une transformation bactérienne est réalisée avec la construction du plasmide pPROEX codant pour la protéine cycline T1 (aa 1-300) étiquetée GST en N-terminale. Les cellules sont cultivées dans 50 mL d'un milieu LB-ampicilline 50  $\mu$ g/mL pendant une nuit à 37°C sous agitation. Les cellules sont transvasées dans un milieu LB-ampicilline de 2 L, puis incubées pendant 4h à 37°C sous agitation et l'expression des protéines est induite avec 1 mM d'IPTG pendant 17h à 18°C sous agitation continue. Ensuite, les cellules sont pelletées en centrifugeant 15 minutes à 5 000 g à 4°C.

Les culots cellulaires sont remis en suspension dans un tampon de lyse complété avec un demi comprimé de Protease Inhibitor Cocktail (ROCHE, cOmplete™). La préparation est soniquée pendant 4 minutes sur glace à 50% de puissance de l'appareil Branson Ultrasonics™ Sonifier S-250A (Thermo Fischer Scientific). Le lysat obtenu est centrifugé 45 minutes à 11 000 g à 4°C.

Une purification en « batch » de GST-cycline T1 est réalisée avec une résine Gluthatione Sepharose 4B (Thermo Fisher Scientific). 1 volume de résine est équilibré deux fois avec 5 volumes de tampon de lyse. Le tampon de lyse est éliminé en centrifugeant 1 minute à 4°C à 1 800 g. La résine est incubée

avec le lysat bactérien pendant 2h à 4°C en rotation. Ensuite une étape de centrifugation permet d'éliminer la fraction d'échantillon non retenue contenue dans le surnageant. 5 volumes de tampon de lyse sont ajoutés puis éliminés par centrifugation. Cette étape de lavage est répétée 3 fois. Puis, l'élution est réalisée en trois temps successifs avec 1 volume de tampon d'élution préparé extemporanément. La première fraction d'élution est récupérée tout de suite après l'ajout du tampon d'élution ; la deuxième est récupérée après 15 minutes d'incubation ; et la troisième après 5 minutes d'incubation. Les fractions d'élution sont récupérées par une étape de centrifugation à chaque fois.

Les fractions d'intérêt sont rassemblées et dialysées 1h à 4°C contre le tampon de stockage. L'échantillon est dosé grâce au spectrophotomètre NanoDrop™ 2000/2000c en réalisant des mesures d'absorbance à 280 nm ; puis il est congelé à l'azote liquide pour une conservation à -80°C.

La protéine recombinante GST-cycline T1 a une séquence protéique comprenant 533 acides aminés (**tableau 7**), une masse moléculaire de 61 kDa et un pHi de 6,4.

MSPILGYWKIKGLVQPTRLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDG DVKLTQSMAIIRYIADKHNMLGGCPKERAISMLEGAVLDIRYGVSRIAYSKDFETLKVDL SKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKK RIEAIPOIDKYLKSSKYIAWPLQGWQATFGGGDHPKSDLVPRGSRRASVGSMEGERKNNN KRWYFTREQLENSPSRRFVDPDKELSYRQQAANLLQDMGQRLNVSQLTINTAIVYMHRF YMIQSFTQFPGNSVAPAALFLAAKVEEQPKKLEHVIKVAHTCLHPQESLPDTRSEAYLQQV QDLVILESIIQLTGFELTIDHPHTHVVKCTQLVRASKDLAQTSYFMATNSLHLTTFSLQYTP PVVACVCIHLACKWSNWEIPVSTDGKHWEYVDATVTLELLDELTHEFLQILEKTPNRLK RIWNWRACEAAKTKADDRGTDEKTSEQTILNMISQSSSDTTIAGLM
---

**Tableau 7.** La séquence de la protéine recombinante GST-cycline T1.

### XII.1.3. his-cycline T1 (aa 7-261)

Une transformation bactérienne est réalisée avec la construction du plasmide pET28a codant pour la protéine cycline T1 (aa 7-261) étiquetée 6-histidine en N-terminale. Les cellules sont cultivées comme détaillé précédemment pour GST-cycline T1, avec de la kanamycine à la place de l'ampicilline. Les culots cellulaires sont remis en suspension dans un tampon de lyse complété avec 1 mM d'ATP et un demi comprimé de Protease Inhibitor Cocktail. La préparation est soniquée et centrifugée dans les conditions détaillées précédemment.

Le surnageant est ensuite chargé sur une colonne HisTrap crude de 5 mL (Cytiva) équilibrée avec 2 volume colonne (VC) de tampon de lyse grâce à une pompe péristaltique. Par la suite, la colonne est lavée grâce à la pompe péristaltique avec 3 VC de tampon de lavage. La purification est réalisée avec un système ÄKTA pure (Cytiva) en deux temps : premièrement, une étape de lavage avec 5 VC de tampon de lyse, puis dans un second temps, une étape d'élution de la protéine. L'étape d'élution est réalisée avec un gradient linéaire sur 10 VC allant de 0% à 100% du tampon d'élution, suivi d'une étape à 100% de tampon d'élution sur 0,5 VC. Les fractions d'élution sont contrôlées par SDS-PAGE. Les fractions d'intérêt sont rassemblées et concentrées avec des tubes Amicon® Ultra-15 (Merck Millipore) de seuil de rétention de 3 kDa en centrifugeant à 4°C à 4 500 g pour obtenir un volume final de 2 mL. Au préalable, la membrane du concentrateur est équilibrée avec du tampon de dialyse en centrifugeant à 4°C à 4 500 g pendant 5 minutes.

L'échantillon protéique est dialysé sur la nuit à 4°C, sous agitation contre un tampon de stockage, puis il est dosé et stocké à -80°C.

La protéine recombinante his-cycline T1 a une séquence protéique comprenant 279 acides aminés (**tableau 8**), une masse moléculaire de 32 kDa et un pHi de 6,9.

```
HHHHHHSSGLVPRGSHMENLYFQGNNKRWYFTREQLNSPSRRFGVDPDKELSYRQQA  
ANLLQDMGQRLNVSQLTINTAIVYMRFYMIQSFTQFPGNSVAPAALFLAAKVEEQPKKLE  
HVIKVAHTCLHPQESLPDTRSEAYLQQVQDLVILESILQTLGFELTIDHPHPTHVVKCTQLVR  
ASKDLAQTSYFMATNSLHLTTFSLQYTPPVVACVCIHLACKWSNWEIPVSTDGKHWWEYV  
DATVTLELLDELTHEFLQILEKTPNRLKRIWNWRAC
```

**Tableau 8.** La séquence de la protéine recombinante his-cycline T1.

#### XII.1.4. CDK9-his (aa 8-330)

La protéine recombinante CDK9-his a une séquence protéique comprenant 352 acides aminés (**tableau 9**), une masse moléculaire de 41 kDa et un pHi de 8,4.

```
MVECPFCDEVSKYEKLAKIGQGTFGEVFKARHRKTGQKVALKKVLMENEKEGFPITALREI  
KILQLLKHENVVNLIEICRTKASPYNRCKGSIYLVDFCEHDLAAGLLSNVLVKFTLSEIKRV  
MQMLLNGLYYIHRNKILHRDMKAANVLITRDGVLKLADFGGLARAFSLAKNSQPNRYTNR  
VVTLWYRPPPELLLGERDYGPPIDLWGAGCIMAEMWTRSPIMQGNTEQHQLALISQLCGSIT  
PEVWPNVDNYELYEKLELVKGGQRKVKDRLKAYVRDPYALDLIDKLLVLDPAQRIDSDD  
ALNHDFFWSDPMPSDLKGMSTENLYFQGELALVPRGSSAHHHHHHHHHHH
```

**Tableau 9.** La séquence de la protéine recombinante CDK9-his.

#### Essai de purification de CDK9-his seule

Une transformation bactérienne est réalisée avec la construction du plasmide pET52b codant pour la protéine CDK9 (aa 8-330) étiquetée 10-histidine en C-terminale. Les cellules sont cultivées comme détaillé précédemment avec de l'ampicilline. La lyse bactérienne et la purification sur une colonne HisTrap crude de 5 mL sont réalisées dans les conditions décrites précédemment pour la protéine his-cycline T1.

#### Co-lyse en conditions native et dénaturante de culots bactériens ayant exprimé his-cycline T1 et CDK9-his

Les culots cellulaires de his-cycline T1 et CDK9-his sont lysés ensemble et traités dans les conditions décrites pour la lyse et la purification de CDK9-his seule, la seule différence est que dans le cas de la co-lyse il n'y a pas d'étape de lavage à grande concentration de NaCl pour éviter de décrocher le complexe cycline T1:CDK9.

#### Transformation et culture bactérienne selon le protocole adapté pour les protéines toxiques

Une transformation bactérienne est réalisée avec la construction du plasmide pET52b codant pour la protéine CDK9 (aa 8-330) étiquetée 10-histidine en C-terminale selon le protocole décrit précédemment pour les protéines dont l'expression est toxique chez *E. coli*. Cependant au lieu d'étaler l'échantillon bactérien sur un milieu solide (boîte de Pétri LB-ampicilline), l'échantillon bactérien est inoculé dans 50 mL de milieu LB-ampicilline 50 µg/mL pendant une nuit à 30°C sous agitation. Puis les cellules sont transvasées dans un milieu LB-ampicilline de 2 L pendant 8h à 30°C sous agitation et l'expression des

protéines est induite avec 0,1 mM d'IPTG pendant 2h à 37°C sous agitation continue. Les cellules sont pelletées en centrifugeant 15 minutes à 5 000 g à 4°C, puis stockées à -20°C.

#### XII.1.5. his-ISG20<sub>wt</sub> (aa 1-181)

Une transformation bactérienne est réalisée avec la construction du plasmide pPROEX codant pour la protéine ISG20<sub>wt</sub> (aa 1-181) étiquetée 6-histidine en N-terminale selon le protocole décrit précédemment pour les protéines dont l'expression est toxique chez *E. coli*. Les cellules sont cultivées dans 200 mL d'un milieu LB-ampicilline 50 µg/mL pendant 1 jour à 25°C sous agitation. Puis les cellules sont transvasées dans un milieu LB-ampicilline de 2 L et incubées pendant 16h à 25°C sous agitation. L'expression des protéines est induite avec 0,1 mM d'IPTG pendant 3h30 à 25°C sous agitation continue. Les cellules sont pelletées en centrifugeant 15 min à 5 000 g à 4°C.

Les culots cellulaires sont remis en suspension dans un tampon de lyse complété avec 1 mM d'ATP, 1 mM de MgCl<sub>2</sub> et un demi comprimé de Protease Inhibitor Cocktail. La préparation est soniquée et centrifugée. Le surnageant est sur une colonne HisTrap HP de 5 mL (Cytiva) équilibrée avec 2 VC de tampon de lyse grâce à une pompe péristaltique. Par la suite, la colonne est lavée grâce à la pompe péristaltique avec 3 VC de tampon de lavage. La purification est réalisée avec un système ÄKTA pure en deux temps comme détaillé pour his-cycline T1. Les fractions d'intérêt sont rassemblées et dialysées sur la nuit à 4°C, sous agitation contre un tampon SEC.

Le lendemain, l'échantillon est concentré avec des tubes Amicon® Ultra-4 (Merck Millipore) de seuil de rétention de 10 kDa en centrifugeant à 4°C à 4 500 g pour obtenir un volume final de 850 µL. L'échantillon concentré est injecté grâce au système ÄKTA pure sur une colonne Superdex 75 10/300 GL (Cytiva) équilibrée avec 2 VC de tampon SEC. L'élution est réalisée sur un gradient isocratique de 1,5 VC. Les fractions d'intérêt sont rassemblées et dialysées 1h à 4°C contre un tampon de stockage. L'échantillon est conservé à -20°C.

La protéine recombinante his-ISG20<sub>wt</sub> a une séquence protéique comprenant 205 acides aminés (**tableau 10**), une masse moléculaire de 23 kDa et un pHi de 7,4.

HHHHHHHDYDIPTTENLYFQGAMDPMAGSREVVAMDCMVGLGPHRESGLARCSLVNVH GAVLYDKFIRPEGEITDYRTRVSGVTPQHVMGATPFAVARLEILQLLKGKLVVGHDLKHDF QALKEDMSGYTIYDTSTDRLWREAKLDHCRRVSLRVLSERLLHKSQNSLLGHSSVEDAR ATMELYQISQRIRARRGLPRLAVSD
---

**Tableau 10.** La séquence de la protéine recombinante his-ISG20<sub>wt</sub>.

#### XII.1.6. his-ISG20 D94A (aa 1-181)

Une transformation bactérienne est réalisée avec la construction du plasmide pPROEX codant pour la protéine mutante ISG20 D94A (aa 1-181) étiquetée 6-histidine en N-terminale. Les cellules sont cultivées et lysées comme détaillé précédemment pour GST-cycline T1. Le surnageant est purifié sur une colonne HisTrap crude de 5 mL dans les conditions détaillées pour his-cycline T1. Les fractions d'intérêt sont rassemblées et dialysées sur la nuit à 4°C, sous agitation contre un tampon Héparine A.

Le lendemain, l'échantillon est chargé sur une colonne Heparin HiTrap de 5 mL (Cytiva) équilibrée avec 2 VC de tampon Héparine A grâce à une pompe péristaltique. La purification est réalisée avec un système ÄKTA purifier en deux temps : premièrement, une étape de lavage avec 5 VC de tampon Héparine A, puis dans un second temps, une étape d'élution de la protéine en utilisant un gradient linéaire sur 10 VC allant de 0% à 100% de tampon Héparine B. Les fractions d'intérêt sont rassemblées et dialysées sur la nuit à 4°C, sous agitation contre le tampon Héparine A.

Le lendemain, l'échantillon est concentré avec des tubes Amicon® Ultra-4 de seuil de rétention de 10 kDa en centrifugeant à 4°C à 5 000 g. Puis il est dialysé 1h à 4°C contre un tampon de stockage. L'échantillon est dosé puis conservé à -80°C.

La protéine recombinante his-ISG20 D94A a une séquence protéique comprenant 205 acides aminés (**tableau 11**), une masse moléculaire de 23 kDa et un pHi de 8,0.

HHHHHHHDYDIPTTENLYFQGAMDPMAGSREVVAMDCMVGLGPHRESGLARCSLVNVH GAVLYDKFIRPEGEITDYRTRVSGVTPQHMGATPFAVARLEILQLLKGKLVVGHDLKHA QALKEDMSGYTIYDTSTDRLWLWREAKLDHCRRVSLRVLSEKLLHKSISQNSLLGHSSVEDAR ATMELYQISQRIRARRGLPRLAVSD
---

**Tableau 11.** La séquence de la protéine recombinante his-ISG20 D94A.

#### XII.1.7. PABPC1-his (aa 1-636)

Une transformation bactérienne est réalisée avec la construction du plasmide pET codant pour la protéine PABPC1 (aa 1-636) étiquetée 6-histidine en C-terminale. Les cellules sont cultivées et lysées comme détaillé précédemment pour GST-cycline T1.

Le surnageant est ensuite chargé sur une colonne HisTrap HP de 5 mL équilibrée avec 2 VC de tampon de lyse grâce à une pompe péristaltique. La purification est réalisée avec un système ÄKTA pure en deux temps : premièrement, 10 VC de tampon de lyse, puis dans un second temps, une étape d'éluion de la protéine en utilisant un gradient linéaire sur 10 VC allant de 0% à 100% du tampon d'éluion. Les fractions d'intérêt sont rassemblées et dialysées sur la nuit à 4°C, sous agitation contre un tampon SEC. L'échantillon est concentré avec des tubes Amicon® Ultra-4 de seuil de rétention de 50 kDa en centrifugeant à 4°C à 4 500 g pour obtenir un volume final de 100 µL. Puis il est injecté grâce au système ÄKTA pure sur une colonne Superdex 75 5/150 GL (Cytiva) équilibrée avec 2 VC de tampon SEC. L'éluion est réalisée sur un gradient isocratique de 1,5 VC. Les fractions d'intérêt sont rassemblées et dialysées sur la nuit à 4°C contre un tampon de stockage. L'échantillon est concentré comme décrit précédemment puis dosé et conservé à -80°C.

#### XII.2. Electrophorèse en gel de polyacrylamide en condition dénaturante pour les protéines

Pour la préparation de gels SDS-PAGE, un gel de concentration est coulé au-dessus d'un gel de migration. Le gel de migration est composé de 12% ou 15% d'acrylamide:bisacrylamide (37,5:1), 0,375 M Tris-HCl pH 8,8, 0,4% SDS, 0,1% APS et 7 mM TEMED. Tandis que le gel de concentration contient 4% d'acrylamide:bisacrylamide (37,5:1), 0,375 M Tris-HCl pH 6,8, 0,4% SDS, 0,1% APS et 7 mM TEMED.

Avant le chargement des échantillons, du tampon de Laemmli à 1x final est ajouté (2% SDS, 10% glycérol, 5% β-mercaptoéthanol, 2% bleu de bromophénol et 62,5 mM Tris-HCl pH 6,8). La migration se fait à 90 V dans le gel de concentration puis à 200 V dans le gel de migration. Les gels sont baignés dans un tampon de migration contenant : 25 mM Tris pH 8,3, 192 mM glycine et 0,1% SDS. Pour identifier la masse moléculaire des bandes, un marqueur de masse moléculaire est utilisé : PageRuler™ Plus Prestained Protein Ladder (Thermo Fisher Scientific), allant de 10 à 250 kDa.

La solution commerciale InstantBlue® Protein Stain (Expedeon) est utilisée pour colorer les protéines sur un gel de polyacrylamide en réalisant une incubation de 15 minutes sous agitation.





une centrifugation à 10 000 g à 20°C pendant 30 minutes est réalisée. Du NaCl est ajouté dans le surnageant pour obtenir une concentration finale en NaCl de 100 mM. Puis l'échantillon est chargé sur une colonne HiTrapQ-XL de 5 mL (Cytiva) équilibrée avec 2 VC de tampon de lyse avec une seringue. La purification est réalisée avec un système ÄKTA purifier en deux temps : premièrement, une étape de lavage avec 5 VC de tampon de lyse, puis dans un second temps, une étape d'élution de l'ARN en utilisant un gradient linéaire sur 15 VC allant de 0% à 100% du tampon d'élution (10 mM Tris-HCl pH 7,5, 10 mM MgCl<sub>2</sub> et 1 M NaCl). Les fractions d'élution sont contrôlées par une électrophorèse sur gel de polyacrylamide en condition dénaturante. Les fractions d'intérêt sont rassemblées et injectées grâce au système ÄKTA pure sur une colonne Superdex 75 5/150 GL équilibrée avec 2 VC de tampon d'élution. L'élution est réalisée sur un gradient isocratique de 1,5 VC. Les fractions d'élution sont contrôlées, puis les fractions d'intérêt sont rassemblées et purifiées sur gel de polyacrylamide en condition dénaturante. Une élution par diffusion et une électro-élution sont réalisées ensuite.

Le protocole ci-dessus est inspiré de celui publié par Ponchon *et al.* (2009).

### XII.3.3. Purification de l'ARN sur gel de polyacrylamide en condition dénaturante

La purification de l'ARN sur gel de polyacrylamide en condition dénaturante permet de purifier un ARN d'intérêt en récupérant la bande d'intérêt directement sur le gel. Par la suite, l'ARN est élué de la bande par diffusion ou alors en appliquant un courant électrique, il s'agit alors d'une électro-élution.

En pratique pour la purification sur gel, l'ARN est dilué de moitié dans un tampon de charge à 1X final (95% formamide, 18 mM EDTA, 0,025% SDS, 0,01% xylène cyanol et 0,01% bleu de bromophénol), puis chargé sur un gel dénaturant 10% ou 15% d'acrylamide:bisacrylamide (19:1), TBE 1X, 7M urée, 0,1% APS et 7 mM TEMED. Les échantillons sont chauffés 5 minutes à 98°C puis chargé sur gel. La migration se fait à 300 V dans le tampon TBE 1X. La révélation des ARN se fait par le BET avec un transilluminateur UV ou alors par imagerie au phosphore à l'aide du scanner Typhoon<sup>TM</sup> FLA 9500 (GE HealthCare).

#### XII.3.3.1. Elution par diffusion

La bande correspondant à l'ARN d'intérêt est coupée avec une lame de rasoir et placée dans un tube contenant 500 µL de tampon d'élution de composition finale suivante : 1 mM EDTA, 0,2% SDS, 0,5 M d'acétate de sodium et 0,01 volume d'une solution de PCI.

L'échantillon est conservé à -20°C pendant 30 minutes puis chauffé à 95°C pendant 5 minutes. Il est ensuite placé sur une roue en rotation toute la nuit à température ambiante. Le lendemain, l'échantillon est chauffé pendant 30 minutes à 37°C puis un traitement par le PCI est réalisé suivi d'une précipitation à l'éthanol comme décrit précédemment. L'ARN est chauffé dans un tampon de repliement (20 mM HEPES pH 7,5, 200 mM NaCl, 2 mM MgCl<sub>2</sub> et 2 mM DTT) pendant 5 minutes à 95 °C, puis un refroidissement lent et contrôlé de 40 minutes a été réalisé. L'ARN est utilisé par la suite pour la formation de complexes ribonucléoprotéiques avec his-IN-CTD (essais de cristallogénèse du 30/03/2022).

#### XII.3.3.2. Electro-élution

Comme précédemment la bande d'intérêt est isolée, puis elle est placée dans une membrane de dialyse au sein d'un appareil d'électrophorèse avec un tampon TBE 1X. Ensuite, une migration de 2h à 150 V est réalisée. Le tampon d'électrophorèse et le tampon au sein de la membrane de dialyse sont contrôlés par une électrophorèse sur gel de polyacrylamide en condition dénaturante. L'ARN est traité par le PCI, précipité à l'éthanol, puis chauffé dans un tampon de repliement.

L'ARN est ensuite incubé avec his-IN-CTD pour la formation des complexes ribonucléoprotéiques (essais de cristallogénèse du 09/09/2022).

#### XII.4. Marquage radioactif de l'ARN catalysé par la T4 polynucléotide kinase

Les ARN concernés par ce marquage radioactif sont les ARN TAR-5SL, TAR-9SL, TAR, AC-PolyA et AG(50)-mer.

50 pmol d'ARN sont radiomarqués à l'extrémité 5' en utilisant 10 unités de T4 polynucléotide kinase (T4 PNK) (NEB) mélangées à 3  $\mu\text{L}$  de  $\gamma^{32}\text{P}$ -ATP (3000 Ci/mmol, 10 mCi/mL, PerkinElmer) dans le tampon de la T4 PNK à 1X (NEB) pendant 2h à 37 °C. La T4 PNK est ensuite inactivée par une incubation pendant 20 minutes à 65°C. Les ARN radioactifs sont purifiés sur des colonnes G-25 illustra Microspin (Cytiva) selon le protocole de purification du fournisseur ; et ensuite sur gel de polyacrylamide dénaturant à 10% avec une élution par diffusion comme décrit précédemment. La qualité des ARN est contrôlée sur un gel de polyacrylamide 15% dénaturant. Les ARN sont dosés et conservés à -20°C.

#### XII.5. Complexes ribonucléoprotéiques

##### XII.5.1. Etude de l'interaction protéine – ARN par retard sur gel

###### XII.5.1.1. Etude de l'interaction entre his-IN-CTD et l'ARN minoTAR

Les échantillons sont préparés en mélangeant 0,8  $\mu\text{M}$  d'ARN minoTAR avec des concentrations croissantes de his-IN-CTD (0 ; 0,8 ; 1,6 ; 3,2 ; 4,8 ; 6,4 et 12,8  $\mu\text{M}$ ), correspondant à des ratios molaires ARN:protéine croissants (1:0, 1:1, 1:2, 1:4, 1:6, 1:8 et 1:16). La liaison se fait dans un tampon contenant 25 mM HEPES pH 7,5, 2,5 mM  $\text{MgCl}_2$ , 0,1  $\mu\text{M}$   $\text{ZnSO}_4$ , 2,5 mM DTT, 0,1  $\mu\text{g}/\mu\text{L}$  BSA et 8% PEG 8000 à différentes concentrations de NaCl : 100, 500 et 1 000 mM. Les échantillons sont incubés à 37°C pendant 30 minutes avant d'être analysés par une électrophorèse sur gel de polyacrylamide à 6% en condition native. La migration est réalisée dans un tampon TAE 0,5X sur un gel de dimension suivante : 18 cm x 20 cm.

La coloration et la révélation du gel se font avec du BET et un transilluminateur UV.

###### XII.5.1.2. Etude de l'interaction entre his-ISG20 $_{wt}$ , his-ISG20 D94A et un ARNdb

Les échantillons sont préparés en mélangeant 100 nM d'ARN TAR-9SL marqué radioactivement au  $^{32}\text{P}$  avec des concentrations croissantes de protéines (0 ; 100 ; 200 ; 400 ; 800 et 1 600 nM), correspondant à des ratios molaires ARN:protéine croissants (1:0, 1:1, 1:2, 1:4, 1:8 et 1:16). Deux protéines différentes sont testées : his-ISG20 $_{wt}$  et his-ISG20 D94A. Le mélange est réalisé dans un tampon contenant 50 mM Tris-HCl pH 7,5, 150 mM NaCl, 1 mM DTT, 10 mM  $\text{Mn}^{2+}$ , 0,1  $\mu\text{g}/\mu\text{L}$  BSA et 10% glycérol. Les échantillons sont incubés à 37°C pendant 30 minutes, puis ils sont analysés par une électrophorèse en condition native comme détaillé précédemment.

Les dimensions du gel sont les suivantes : 20 cm x 34 cm. Le gel a été exposé sur la nuit à -20°C. Les résultats ont été visualisés avec le scanner Typhoon<sup>TM</sup> FLA 9500.

### XII.5.1.3. Etude de l'interaction entre his-ISG20<sub>wt</sub>, un ARN<sub>db</sub> et un ARN<sub>sb</sub>

Les échantillons sont préparés en mélangeant 50 nM d'ARN marqué radioactivement au <sup>32</sup>P avec des concentrations croissantes de protéines his-ISG20<sub>wt</sub> (0 ; 50 ; 100 ; 200 et 400 nM), correspondant à des ratios molaires ARN:protéine croissants (1:0, 1:1, 1:2, 1:4 et 1:8). Deux ARN structurés différemment sont étudiés : l'ARN TAR-9SL et l'ARN AG(50)-mer. Le mélange est réalisé dans un tampon contenant 50 mM Tris-HCl pH 7,5, 150 mM NaCl, 5 mM β-mercaptoéthanol, 0,1 µg/µL BSA, 5% glycérol, 2,5 mM EDTA et 10 mM Mn<sup>2+</sup>. Les échantillons sont incubés à 30°C pendant 30 minutes, puis ils sont analysés par électrophorèse et le résultat est visualisé dans les conditions décrites dans le **chapitre XII.5.1.2.**

### XII.5.1.4. Etude de l'interaction entre his-ISG20<sub>wt</sub> et un ARN<sub>sb</sub> en présence de Zn<sup>2+</sup>

Les échantillons sont préparés en mélangeant 30 nM d'ARN AC-PolyA marqué radioactivement au <sup>32</sup>P avec des concentrations croissantes de his-ISG20<sub>wt</sub> (0 ; 120 ; 240 ; 480 et 960 nM), correspondant à des ratios molaires ARN:protéine croissants (1:0, 1:4, 1:8, 1:16 et 1:32). Le mélange se fait dans un tampon contenant 10 mM Tris-HCl pH 7,5, 50 mM NaCl, 2 mM DTT, 0,1 µg/µL BSA, 10 mM ZnSO<sub>4</sub> et 5% glycérol, pendant 20 minutes à 30°C. L'analyse des échantillons par électrophorèse et leur visualisation sont réalisées dans les conditions décrites dans le **chapitre XII.5.1.2.**

### XII.5.1.5. Etude de l'interaction entre his-ISG20<sub>wt</sub> et des ARN<sub>sb</sub> avec différentes modifications post-transcriptionnelles

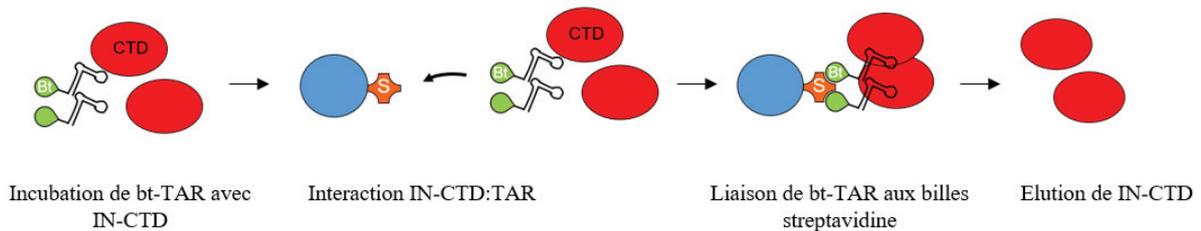
Cette étude est réalisée dans les mêmes conditions que précédemment (voir **chapitre XII.5.1.4**) avec les ARN AC-PolyA sans méthylation ou présentant une des méthylations suivantes : m6A, 2OA et m5C.

### XII.5.1.6. Etude de l'interaction entre PABPC1-his et AC-PolyA

Les échantillons sont préparés en mélangeant 10 nM d'ARN AC-PolyA marqué radioactivement au <sup>32</sup>P avec des concentrations croissantes de PABPC1-his (0 ; 50 ; 100 ; 200 et 500 nM), correspondant à des ratios molaires ARN:protéine croissants (1:0, 1:5, 1:10, 1:20 et 1:50). Le mélange est réalisé dans un tampon contenant 25 mM Tris-HCl pH 7,5, 5 mM MgCl<sub>2</sub>, 100 mM NaCl, 2,5 mM DTT, 0,1 µg/µL BSA et 10% glycérol. Les échantillons ont été incubés à 37°C pendant 15 minutes. L'analyse des échantillons (électrophorèse puis visualisation) se fait dans les conditions décrites dans le **chapitre XII.5.1.2.**

## XII.5.2. Test d'interaction par ARN « pull-down » avec IN-CTD

Une expérience de co-précipitation permet d'étudier l'interaction d'une biomolécule avec une autre. Dans notre cas, il s'agit de l'interaction entre un ARN et une ou des protéines. Pour cela, un ARN biotinylé est incubé avec les protéines étudiées (**figure 90**). Puis ce mélange est mis en contact avec des billes présentant de la streptavidine à leur surface. Des étapes de lavages permettent de retirer les interactions aspécifiques. Finalement, les protéines liées à l'ARN sont analysées par SDS-PAGE.



**Figure 90.** Schéma des étapes d'une expérience de co-précipitation avec des billes streptavidine.

Les protéines (his-IN-CTD- $\Delta$ ACT, his-IN-CTD et GST-cycline T1) à une concentration finale de 4  $\mu$ M sont mélangées dans le tampon de liaison suivant : 100 mM NaCl, 40 mM HEPES pH 7,5, 5 mM MgCl<sub>2</sub>, 10% glycérol et 2 mM  $\beta$ -mercaptoéthanol. 5  $\mu$ L d'échantillon sont prélevés afin de pouvoir contrôler par SDS-PAGE la présence des protéines d'intérêt dans les bonnes quantités. Ensuite, les échantillons sont complétés ou non avec 0,5  $\mu$ M d'ARN TAR-bt dans un volume final de 60  $\mu$ L. Les échantillons sont incubés 30 minutes à 37°C. Puis, 10  $\mu$ L de billes magnétiques de streptavidine (Dynabeads MyOne, Thermo Fisher Scientific) et 200  $\mu$ L de tampon de lavage (40 mM HEPES pH 7,5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 10% glycérol, 2 mM  $\beta$ -mercaptoéthanol et 10 mM imidazole) sont ajoutés. Les échantillons sont incubés pendant 1h30 en rotation à 4°C. Les billes magnétiques sont lavées deux fois avec 500  $\mu$ L de tampon de lavage grâce à un portoir magnétique. L'éluion des protéines est réalisée en incubant les billes 10 minutes avec 10  $\mu$ L de LD 4X. Les fractions d'éluion sont récupérées et contrôlées par SDS-PAGE.

Pour empêcher les interactions aspécifiques des tubes Protein LoBind® (Eppendorf) sont utilisés et de l'imidazole est ajouté dans le tampon de lavage.

### XII.5.3. Test de digestion enzymatique médié par ISG20

#### XII.5.3.1. Test de la processivité ou distributivité d'ISG20

Le test de digestion est réalisé en mélangeant 500 nM d'ARN AC-PolyA marqué radioactivement au <sup>32</sup>P avec 5 nM de his-ISG20 (ratio molaire ARN:protéine de 100:1) en présence ou non de 50  $\mu$ M d'ARN AC-PolyA non-marqué (ratio molaire ARN chaud:ARN froid de 1:10). Le mélange se fait dans le tampon suivant : 50 mM Tris-HCl pH 7,4, 2,5 mM MnCl<sub>2</sub>, 1 mM  $\beta$ -mercaptoéthanol, 0,4 mM DTT, 0,1% Triton X-100. La réaction de digestion est réalisée à 37°C et débute avec l'ajout de his-ISG20. Des prélèvements de 3  $\mu$ L sont réalisés à différents temps pendant 30 minutes (0, 1, 5, 15 et 30 minutes). L'échantillon prélevé est immédiatement mélangé dans 1 volume de tampon de « quenching » (5 mM EDTA, 0,5% SDS, 0,1X LD II, 0,5 M urée et 34% glycérol) pour arrêter la réaction. Les échantillons sont ensuite chauffés 5 minutes à 98°C et conservés à -20°C.

Ensuite, les échantillons sont contrôlés par électrophorèse sur gel de polyacrylamide à 15% en condition dénaturante. La migration est réalisée dans du tampon TBE 1X à 600V sur un gel de dimension 34 cm par 20 cm. Après migration, le gel est exposé sur la nuit à -80°C contre un écran adéquat dans une cassette dédiée. Les résultats sont visualisés avec le scanner Typhoon<sup>TM</sup> FLA 9500.

#### XII.5.3.2. Test de digestion avec des substrats ARN linéaires et méthylés

Le test de digestion est réalisé en mélangeant 600 nM d'ARN marqué radioactivement au <sup>32</sup>P dans un tampon optimisé (50 mM Tris-HCl pH 7,5, 20 mM MnCl<sub>2</sub>, 1 mM DTT, 150 mM NaCl et 10% glycérol) avec 2,4  $\mu$ M de his-ISG20 (ratio molaire ARN:protéine de 1:4). Quatre ARN différents sont utilisés : l'ARN AC-PolyA sans méthylation, l'ARN m6A, l'ARN 2OA et l'ARN m5C. L'ajout de his-ISG20 permet d'initier la réaction de digestion. La réaction de digestion est réalisée à 37°C et des prélèvements

de 2  $\mu\text{L}$  sont réalisés à différents temps pendant 30 minutes (0, 1, 5, 10, 15 et 30 minutes). L'échantillon prélevé est immédiatement mélangé dans 5 volumes de tampon de « *quenching* » pour arrêter la réaction. Les échantillons sont ensuite chauffés 5 minutes à 98°C et conservés à -20°C. Par la suite, ils sont contrôlés par électrophorèse sur gel de polyacrylamide à 20% en condition dénaturante. La migration est réalisée dans du tampon TBE 1X à 400V sur des gels de 20 cm par 34 cm. Après migration, le gel est exposé sur la nuit à -80°C et les résultats sont visualisés avec le scanner Typhoon™ FLA 9500.

### XII.5.3.3. Test de digestion avec la PABP

Le test de digestion est réalisé en mélangeant 30 nM d'ARN AC-PolyA marqué radioactivement au  $^{32}\text{P}$  avec 900 nM de PABPC1-his ou 900 nM de BSA (ratio molaire ARN:protéine de 1:30) dans un tampon optimisé (10 mM Tris-HCl pH 7,4, 10 mM  $\text{MnCl}_2$ , 1 mM DTT, 100 mM NaCl et 10% glycérol). Une incubation de 20 minutes à 37°C est réalisée. Puis 120 nM de his-ISG20 (ratio molaire ARN:protéine de 1:4) sont ajoutés pour initier la réaction de digestion. La réaction de digestion est réalisée à 37°C et des prélèvements de 4  $\mu\text{L}$  sont réalisés à différents temps pendant 60 minutes (0, 1, 5, 15, 30 et 60 minutes). L'échantillon prélevé est immédiatement mélangé dans 1 volume de tampon de « *quenching* » pour arrêter la réaction. Les échantillons sont ensuite chauffés 5 minutes à 98°C et conservés à -20°C. Les échantillons sont analysés (électrophorèse puis révélation) dans les conditions détaillées dans le **chapitre XII.5.3.1.**

Des réactions contrôles sont réalisées en incubant l'ARN uniquement avec la PABPC1-his ou la BSA sans his-ISG20. Pour ces réactions, des prélèvements sont effectués dans les conditions décrites ci-dessus à 0 et 60 minutes.

Le marqueur de taille est composé de trois oligonucléotides ADN de différentes tailles : 19, 22 et 25 nucléotides. Ces oligonucléotides sont marqués radioactivement au  $^{32}\text{P}$  ; puis du tampon de « *quenching* » est ajouté.

### XII.5.3.4. Test de digestion avec des substrats ARN linéaire et structurés

Le test de digestion est réalisé en mélangeant 500 nM d'ARN marqué radioactivement au  $^{32}\text{P}$  dans un tampon optimisé (50 mM Tris-HCl pH 7,4, 2,5 mM  $\text{MnCl}_2$ , 1 mM  $\beta$ -mercaptoéthanol, 0,4 mM DTT, 0,1% Triton X-100) avec 5 nM de his-ISG20 (ratio molaire ARN:protéine de 100:1). Quatre ARN différents sont utilisés : l'ARN AC-PolyA, l'ARN TAR-5SL, l'ARN TAR-9SL et l'ARN TAR. Comme précédemment, l'ajout de his-ISG20 permet d'initier la réaction de digestion. La réaction de digestion est réalisée à 37°C et des prélèvements de 3  $\mu\text{L}$  sont réalisés à différents temps pendant 60 minutes (0, 5, 15, 30 et 60 minutes). L'échantillon prélevé est immédiatement mélangé dans 1 volume de tampon de « *quenching* » pour arrêter la réaction. Les échantillons sont traités et analysés dans les conditions détaillées dans le **chapitre XII.5.3.1.**

## XIII. Etudes structurales

Les études structurales concernent uniquement le complexe formé entre IN-CTD et l'ARN minoTAR.

### XIII.1. Analyse de la monodispersité des échantillons par diffusion dynamique de la lumière

La DLS permet de mesurer le rayon hydrodynamique ( $R_H$ ) d'une macromolécule et son coefficient de polydispersité (PDI) (Falke et Betzel, 2019). Cette technique repose sur le mouvement Brownien d'une particule en solution, donnant accès au coefficient de diffusion par la formule de Stokes-Einstein :

$R_H = \frac{k \times T}{6 \times \pi \times \eta \times D_T}$  avec : T, la température absolue ;  $D_T$ , le coefficient de diffusion de translation ;  $\eta$ , la viscosité dynamique du milieu de dilution ; et k, la constante de Boltzmann.

Le rayon hydrodynamique  $R_H$  est le rayon de la macromolécule en tenant compte de sa couche d'hydratation. Le détecteur permet de mesurer les fluctuations de l'intensité à l'échelle de la microseconde et de relier mathématiquement par la fonction de corrélation ces fluctuations au coefficient de diffusion  $D_T$ , cela donne accès au rayon hydrodynamique  $R_H$ . De plus, le coefficient de polydispersité (PDI) d'un échantillon, c'est-à-dire l'homogénéité de la macromolécule en solution, est représentatif de la largeur de distribution de la taille des macromolécules.

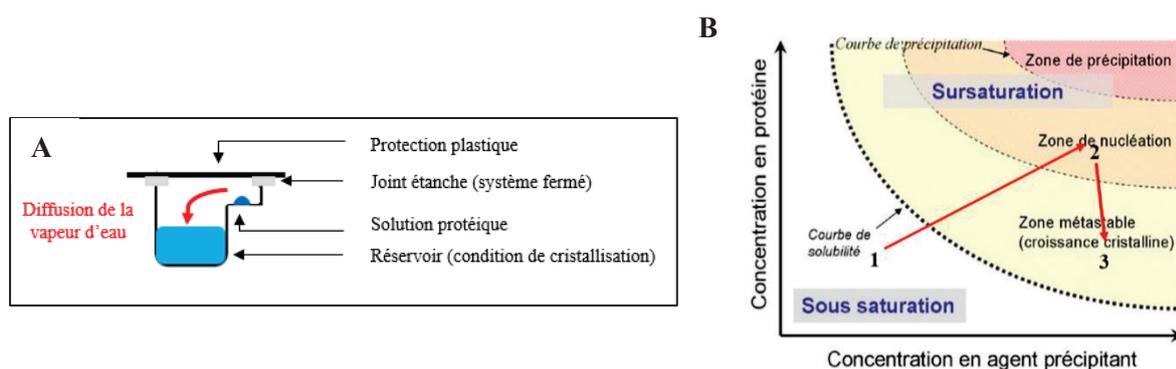
Les échantillons sont préparés en mélangeant 2,5  $\mu\text{M}$  d'ARN minoTAR avec 20  $\mu\text{M}$  de his-IN-CTD (ratio molaire ARN:protéine de 1:8) dans un tampon contenant 50 mM HEPES pH 7,5, 5 mM  $\text{MgCl}_2$ , 0,1  $\mu\text{M}$   $\text{ZnSO}_4$ , 5 mM DTT, 0,1  $\mu\text{g}/\mu\text{L}$  BSA et 16% PEG 8000 à deux concentrations de NaCl : 100 et 250 mM. Les échantillons sont incubés à 37 °C pendant 30 minutes dans un volume final de 120  $\mu\text{L}$ . Puis ils sont analysés dans une microcuvette en plastique BRAND™ (Thermo Fisher Scientific) avec un appareil Zetasizer Nano S90 (Malvern Panalytical). 15 mesures sont réalisées en triplicata à 25°C.

## XIII.2. Cristallographie aux rayons X

### XIII.2.1. La méthode de la diffusion de vapeur d'eau

Cette méthode repose sur la diffusion de vapeur d'eau d'une goutte contenant un mélange de solution protéique et de solution précipitante, vers le réservoir contenant la solution précipitante en excès dans un système fermé (Dessau et Modis, 2011) (**figure 91.A**). Le mouvement de la vapeur d'eau de la goutte vers le réservoir entraîne une concentration lente de la protéine et de l'agent cristallisant jusqu'à ce que l'équilibre soit atteint, c'est-à-dire que la concentration en agent précipitant dans la goutte et le réservoir soit égale. La concentration de la protéine conduit à un phénomène de nucléation, puis il y a une croissance cristalline qui peut aboutir à des cristaux de taille exploitable pour la diffraction aux rayons X (**figure 91.B**).

La condition permettant de cristalliser l'échantillon est inconnue, ainsi un criblage des conditions de cristallisation est réalisé avec des plaques 96 puits.



**Figure 91.** En A, schéma illustrant la technique de la goutte assise. En B, diagramme de solubilité illustrant les phases de cristallisation de protéines.

### XIII.2.2. Cryo-protection et transfert des cristaux

Les cristaux à tester par diffraction aux rayons X sont pêchés grâce à une cryo-boucle et rapidement trempés dans une goutte de solution de cristallisation contenant un cryo-protectant (15% éthylène

glycol). Puis ils sont congelés à l'azote liquide et transportés à un synchrotron où ils sont soumis à un faisceau de rayons X. Les cristaux sont maintenus à basse température dans un flux d'azote gazeux pendant l'enregistrement des données de diffraction (Drenth, 2007).

### XIII.2.3. Obtention et enregistrement des données de diffraction

#### XIII.2.3.1. Principe de diffraction

Lorsqu'un cristal est irradié par les rayons X, il y a réémission de l'onde par chacun des atomes du cristal en fonction de leur position et de leur nature dans un processus de diffusion élastique. Du fait de la périodicité du cristal, il y a des interférences constructives ou destructives entre les ondes réémises. Ce phénomène de diffraction conduit à une réémission des ondes dans des directions privilégiées. L'ensemble des taches de diffraction constitue le réseau réciproque et pour chaque tache des coordonnées (h, k, l) sont calculées.  $F(hkl)$  est un nombre complexe appelé facteur de structure qui permet de décrire l'onde diffractée par un cristal.

$$F(hkl) = |F(hkl)| \times e^{i\phi(hkl)} = \sum_{j=1}^{j=N} f_j \times e^{i2\pi(hx_j+ky_j+lz_j)}$$

$\phi$  : la phase du facteur de structure pour la réflexion (h, k, l) ; N : le nombre d'atomes dans la maille ;  $f_j$  : le facteur de diffusion de l'atome j et x, y, z : sa position.

Les symétries de rotation cristallographiques du réseau direct (cristal) se retrouvent dans le réseau réciproque. En théorie à partir du facteur de structure, il est possible de remonter à la densité électronique  $\rho(x, y, z)$  en un point quelconque de coordonnées (x, y, z) dans le réseau direct. Cependant, l'information concernant les phases est perdue donc il est impossible de calculer la fonction de densité électronique.

$$\rho(x, y, z) = \frac{1}{V} \times \sum_{h,k,l=-\infty}^{+\infty} F(hkl) \times e^{-i2\pi(hx+ky+lz)}$$

V : le volume du cristal.

Pour obtenir une information de phase et déterminer des structures tridimensionnelles par cristallographie aux rayons X, différentes méthodes peuvent être utilisées : le remplacement moléculaire (à l'aide d'une macromolécule de structure homologue), les dérivés lourds isomorphes (en faisant diffuser des atomes lourds dans le cristal) et la dispersion anormale (en exploitant la variation du facteur de diffusion d'un type d'atome en fonction de la longueur d'onde  $\lambda$ ). Ces méthodes peuvent mener à la construction puis à l'affinement d'un modèle atomique.

#### XIII.2.3.2. Essais de cristallogénèse du 30/03/2022

Deux complexes sont formés dans les conditions décrites précédemment pour les expériences de DLS à une concentration en NaCl de 250 mM et à deux ratios ARN:protéine différents :

- Un complexe à un ratio de 1:4 (0,3 mM d'ARN minoTAR + 1,2 mM de his-IN-CTD) ;
- Et un complexe à un ratio de 1:16 (1,5  $\mu$ M d'ARN minoTAR + 24  $\mu$ M de his-IN-CTD).

Pour chacun de ces complexes, quatre kits commerciaux sont utilisés comme indiqué ci-dessous (**tableau 14**).

Nom du kit commercial	Type de plaques	$\mu\text{L}$ de solution de cristallisation	Goutte
Mini Screen (Hampton Research)	24 puits	200	1 goutte suspendue (1 $\mu\text{L}$ de complexe + 2 $\mu\text{L}$ de solution de cristallisation)
MIDAS plus (Molecular Dimensions)	96 puits	70	1 goutte assise (180 nL de complexe + 90 nL de solution de cristallisation)
PEG I (Qiagen)	96 puits	70	1 goutte assise (180 nL de complexe + 90 nL de solution de cristallisation)
Natrix HT (Hampton Research)	96 puits	70	1 goutte assise (180 nL de complexe + 90 nL de solution de cristallisation)

**Tableau 14.** Détails sur les kits, les plaques et les conditions de réalisation des essais de cristallogénèse du 30/03/2022.

Le prélèvement du tampon de cristallisation dans le réservoir et son dépôt dans la goutte pour les plaques 96 puits (MRC, Molecular Dimensions) sont réalisés avec le Mosquito® LV genomics (SPT Labtech). Les plaques 24 puits sont réalisées à la main. Par la suite, les plaques sont recouvertes d'un film plastique et mises à 19°C. Une photo des gouttes est prise de façon journalière grâce au Rock-Imager®-182 (FORMULATRIX).

A partir de ces essais, un cristal a été pêché dans la goutte de la condition B6 du kit Mini Screen (ratio ARN:protéine de 1:4) 40 jours après la réalisation des gouttes. Puis, il a été analysé sur la ligne de lumière ID23-1 à l'ESRF (« *European Synchrotron Radiation Facility* », Grenoble).

### XIII.2.3.3. Essais de cristallogénèse du 09/09/2022

Deux complexes à deux ratios ARN:protéine différents sont formés dans les conditions décrites précédemment pour les expériences de DLS à une concentration en NaCl de 250 mM :

- Un complexe à un ratio de 1:2 (0,4 mM d'ARN minoTAR + 0,8 mM de his-IN-CTD) ;
- Et un complexe à un ratio de 1:4 (0,1 mM d'ARN minoTAR + 0,4 mM de his-IN-CTD).

Les gouttes sont réalisées pour chaque complexe avec trois kits commerciaux différents et stockées à deux températures différentes comme indiqué ci-dessous (**tableau 15**).

Nom du kit commercial	Type de plaques	$\mu\text{L}$ de solution de cristallisation	Gouttes	Température de stockage de la plaque
MIDAS plus	96 puits	70	2 gouttes assises (170 nL de complexe + 85 nL de solution de cristallisation)	19°C
Natrix HT	96 puits	70	2 gouttes assises (170 nL de complexe + 85 nL de solution de cristallisation)	19°C
PEG I	96 puits	70	2 gouttes assises (170 nL de complexe + 85 nL de solution de cristallisation)	19°C
MIDAS plus	96 puits	70	2 gouttes assises (170 nL de complexe + 85 nL de solution de cristallisation)	37°C

Natrix HT	96 puits	70	2 gouttes assises (170 nL de complexe + 85 nL de solution de cristallisation)	37°C
PEG I	96 puits	70	2 gouttes assises (170 nL de complexe + 85 nL de solution de cristallisation)	37°C

**Tableau 15.** Détails sur les kits, les plaques et les conditions de réalisation des essais de cristallogénèse du 09/09/2022.

La goutte 1 correspond au complexe avec un ratio molaire ARN:protéine de 1:2, et la goutte 2 correspond à celui avec un ratio de 1:4.

A partir de ces essais sept cristaux ont été pêchés dans les conditions suivantes à 37°C : E1 du kit PEGI (goutte 2), E2 du kit PEGI (goutte 2), E3 du kit PEGI (gouttes 1 et 2), H3 du kit Natrix (goutte 2), H12 du kit Natrix (goutte 2) et B11 (goutte 2) du kit MIDAS. Puis, ils ont été analysés sur la ligne de lumière PX2 au synchrotron SOLEIL (Source Optimisée de Lumière d'Énergie Intermédiaire du LURE, Saint-Aubin).

#### XIII.2.3.4. Enregistrement des données

Lors de l'enregistrement des données sur les lignes de lumières au synchrotron SOLEIL ou à l'ESRF, une première caractérisation du cristal est réalisée. Le cristal est soumis à un rayonnement X à 40% de transmission et quatre images sont collectées avec un pas d'oscillation du cristal de 1°. Avec ces quatre images, il est possible de discriminer un cristal de sel d'un cristal de protéine et ainsi de poursuivre ou non la collecte de données selon le résultat obtenu.

## Références bibliographiques

### [A]

Abbas A., Lichtman A., Pillai Shiv. Cellular and Molecular Immunology. 9th Edition. Elsevier, 2016. 608 p.

Aboul-ela F. Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Research*. 1996 Oct 15;24(20):3974–81.

AIDES. VIH et SIDA : Kézako ? [En ligne]. 2023. [consulté le 13 mars 2023]. <https://www.aides.org/vih-sida-c-quoi>

Andrake MD, Skalka AM. Retroviral Integrase: Then and Now. *Annu Rev Virol*. 2015 Nov 9;2(1):241–64.

Anshabo AT, Milne R, Wang S, Albrecht H. CDK9: A Comprehensive Review of Its Biology, and Its Role as a Potential Target for Anti-Cancer Agents. *Front Oncol*. 2021 May 10;11:678559.

Ao Z, Danappa Jayappa K, Wang B, Zheng Y, Kung S, Rassart E, et al. Importin  $\alpha 3$  Interacts with HIV-1 Integrase and Contributes to HIV-1 Nuclear Import and Replication. *J Virol*. 2010 Sep;84(17):8650–63.

Arnolds KL, Spencer JV. CXCR4: A virus's best friend? *Infection, Genetics and Evolution*. 2014 Jul;25:146–56.

Arts EJ, Hazuda DJ. HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harbor Perspectives in Medicine*. 2012 Apr 1;2(4):a007161–a007161.

Asamitsu K, Fujinaga K, Okamoto T. HIV Tat/P-TEFb Interaction: A Potential Target for Novel Anti-HIV Therapies. *Molecules*. 2018 Apr 17;23(4):933.

### [B]

Balakrishnan M, Yant SR, Tsai L, O'Sullivan C, Bam RA, Tsai A, et al. Non-Catalytic Site HIV-1 Integrase Inhibitors Disrupt Core Maturation and Induce a Reverse Transcription Block in Target Cells. Harrich D, editor. *PLoS ONE*. 2013 Sep 9;8(9):e74163.

Ballandras A, Moreau K, Robert X, Confort MP, Merceron R, Haser R, et al. A Crystal Structure of the Catalytic Core Domain of an Avian Sarcoma and Leukemia Virus Integrase Suggests an Alternate Dimeric Assembly. Qiu J, editor. *PLoS ONE*. 2011 Aug 9;6(8):e23032.

Ballandras-Colas A, Brown M, Cook NJ, Dewdney TG, Demeler B, Cherepanov P, et al. Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature*. 2016 Feb;530(7590):358–61.

Ballandras-Colas A, Maskell DP, Serrao E, Locke J, Swuec P, Jónsson SR, Kotecha A, Cook NJ, Pye VE, Taylor IA, Andrésdóttir V, Engelman AN, Costa A, Cherepanov P. A supramolecular assembly mediates lentiviral DNA integration. *Science*. 2017 Jan 6;355(6320):93-95.

Baumli S, Lolli G, Lowe ED, Troiani S, Rusconi L, Bullock AN, et al. The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. *EMBO J*. 2008 Jul 9;27(13):1907–18.

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*. 1983 May 20;220(4599):868–71.

Belashov IA, Crawford DW, Cavender CE, Dai P, Beardslee PC, Mathews DH, et al. Structure of HIV TAR in complex with a Lab-Evolved RRM provides insight into duplex RNA recognition and synthesis of a constrained peptide that impairs transcription. *Nucleic Acids Research*. 2018 Jul 27;46(13):6401–15.

- Beloglazova N, Flick R, Tchigvintsev A, Brown G, Popovic A, Nocek B, et al. Nuclease Activity of the Human SAMHD1 Protein Implicated in the Aicardi-Goutières Syndrome and HIV-1 Restriction. *Journal of Biological Chemistry*. 2013 Mar;288(12):8101–10.
- Benleulmi MS, Matysiak J, Robert X, Miskey C, Mauro E, Lapaillerie D, et al. Modulation of the functional association between the HIV-1 intasome and the nucleosome by histone amino-terminal tails. *Retrovirology*. 2017 Dec;14(1):54.
- Beyrer C, McCormack S, Grulich A. Pre-Exposure Prophylaxis for HIV Infection as a Public Health Tool. *J Law Med Ethics*. 2022;50(S1):24-28.
- Bhatt V, Shi K, Salamango DJ, Moeller NH, Pandey KK, Bera S, Bohl HO, Kurniawan F, Orellana K, Zhang W, Grandgenett DP, Harris RS, Sundborger-Lunna AC, Aihara H. Structural basis of host protein hijacking in human T-cell leukemia virus integration. *Nat Commun*. 2020 Jun 19;11(1):3121.
- Bieniasz PD, Grdina TA, Bogerd HP, Cullen BR. Recruitment of cyclin T1/P-TEFb to an HIV type 1 long terminal repeat promoter proximal RNA target is both necessary and sufficient for full activation of transcription. *Proc Natl Acad Sci USA*. 1999 Jul 6;96(14):7791–6.
- Bocian A, Sławek S, Jaromin M, Hus KK, Buczkowicz J, Łysiak D, et al. Comparison of Methods for Measuring Protein Concentration in Venom Samples. *Animals (Basel)*. 2020 Mar 8;10(3):448.
- 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 USA*. 2019 May 21;116(21):10372–81.
- Bukrinskaya A. HIV-1 matrix protein: A mysterious regulator of the viral life cycle. *Virus Research*. 2007 Mar;124(1–2):1–11.
- Bushman FD, Craigie R. Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. *Proc Natl Acad Sci USA*. 1991 Feb 15;88(4):1339–43.
- [C]
- C. Quaresma AJ, Bugai A, Barboric M. Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. *Nucleic Acids Res*. 2016 Sep 19;44(16):7527–39.
- Cachay E. R. Infection par le virus de l'immunodéficience humaine (VIH). [En ligne]. 2021. [consulté le 13 mars 2023]. <https://www.msmanuals.com/fr/accueil/infections/infection-par-le-virus-de-l-immunod%C3%A9ficience-humaine-vih/infection-par-le-virus-de-l-immunod%C3%A9ficience-humaine-vih>
- Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat Struct Mol Biol*. 1997 Jul;4(7):567–77.
- Campbell-Yesufu OT, Gandhi RT. Update on human immunodeficiency virus (HIV)-2 infection. *Clin Infect Dis*. 2011 Mar 15;52(6):780-7.
- Centers for Disease Control and Prevention. Pre-Exposure Prophylaxis (PrEP). [En ligne]. 2022. [consulté le 13 mars 2023]. <https://www.cdc.gov/hiv/risk/prep/index.html>
- Cereseto A, Manganaro L, Gutierrez MI, Terreni M, Fittipaldi A, Lusic M, et al. Acetylation of HIV-1 integrase by p300 regulates viral integration. *EMBO J*. 2005 Sep 7;24(17):3070–81.
- Chavali SS, Bonn-Breach R, Wedekind JE. Face-time with TAR: Portraits of an HIV-1 RNA with diverse modes of effector recognition relevant for drug discovery. *Journal of Biological Chemistry*. 2019 Jun;294(24):9326–41.
- Chen H, Engelman A. The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc Natl Acad Sci USA*. 1998 Dec 22;95(26):15270–4.
- Chen H, Wei SQ, Engelman A. Multiple Integrase Functions Are Required to Form the Native Structure of the Human Immunodeficiency Virus Type I Intasome. *Journal of Biological Chemistry*. 1999 Jun;274(24):17358–64.

Chen J, Liu L, You C, Gu J, Ruan W, Zhang L, et al. Structural and biochemical insights into small RNA 3' end trimming by Arabidopsis SDN1. *Nat Commun.* 2018 Sep 4;9(1):3585.

Chen JCH, Krucinski J, Miercke LJW, Finer-Moore JS, Tang AH, Leavitt AD, et al. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding. *Proc Natl Acad Sci USA.* 2000 Jul 18;97(15):8233–8.

Chen Z, Yan Y, Munshi S, Li Y, Zugay-Murphy J, Xu B, Witmer M, Felock P, Wolfe A, Sardana V, Emini EA, Hazuda D, Kuo LC. X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50-293)--an initial glance of the viral DNA binding platform. *J Mol Biol.* 2000 Feb 18;296(2):521-33.

Cherepanov P, Devroe E, Silver PA, Engelman A. Identification of an Evolutionarily Conserved Domain in Human Lens Epithelium-derived Growth Factor/Transcriptional Co-activator p75 (LEDGF/p75) That Binds HIV-1 Integrase. *Journal of Biological Chemistry.* 2004 Nov;279(47):48883–92.

Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y, et al. HIV-1 Integrase Forms Stable Tetramers and Associates with LEDGF/p75 Protein in Human Cells. *Journal of Biological Chemistry.* 2003 Jan;278(1):372–81.

Cherepanov P, Sun ZYJ, Rahman S, Maertens G, Wagner G, Engelman A. Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nat Struct Mol Biol.* 2005 Jun;12(6):526–32.

Choi J, Ryoo J, Oh C, Hwang S, Ahn K. SAMHD1 specifically restricts retroviruses through its RNase activity. *Retrovirology.* 2015 Dec;12(1):46.

Chou S, Upton H, Bao K, Schulze-Gahmen U, Samelson AJ, He N, et al. HIV-1 Tat recruits transcription elongation factors dispersed along a flexible AFF4 scaffold. *Proc Natl Acad Sci USA [Internet].* 2013 Jan 8 [cited 2023 Mar 13];110(2). Available from: <https://pnas.org/doi/full/10.1073/pnas.1216971110>

Churcher MJ, Lamont C, Hamy F, Dingwall C, Green SM, Lowe AD, Butler JG, Gait MJ, Karn J. High affinity binding of TAR RNA by the human immunodeficiency virus type-1 tat protein requires base-pairs in the RNA stem and amino acid residues flanking the basic region. *J Mol Biol.* 1993 Mar 5;230(1):90-110.

Clark E, Nava B, Caputi M. Tat is a multifunctional viral protein that modulates cellular gene expression and functions. *Oncotarget.* 2017 Apr 18;8(16):27569–81.

Coffin JM. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science.* 1995 Jan 27;267(5197):483-9.

Compton AA, Bruel T, Porrot F, Mallet A, Sachse M, Euvrard M, et al. IFITM Proteins Incorporated into HIV-1 Virions Impair Viral Fusion and Spread. *Cell Host & Microbe.* 2014 Dec;16(6):736–47.

Crowe S, Mills J, McGrath MS. Quantitative immunocytofluorographic analysis of CD4 surface antigen expression and HIV infection of human peripheral blood monocyte/macrophages. *AIDS Res Hum Retroviruses.* 1987 Summer;3(2):135-145.

## [D]

D'Orso I, Frankel AD. RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation. *Nat Struct Mol Biol.* 2010 Jul;17(7):815–21.

Da Rocha M, Poirier M, Bonhamn B, Bell B. Bloquer la transcription du VIH pour verrouiller le virus. *Virologie (Montrouge).* 2022 Feb 24;26(1):23-40.

Dana-Farber Cancer Institute, Achuthan V, Perreira JM, Ahn JJ, Brass AL, Engelman AN. Capsid-CPSF6 interaction: Master regulator of nuclear HIV-1 positioning and integration. *JoLS.* 2019 Jun 1;1(1):39–45.

Dar MJ, Monel B, Krishnan L, Shun MC, Di Nunzio F, Helland DE, et al. Biochemical and virological analysis of the 18-residue C-terminal tail of HIV-1 integrase. *Retrovirology.* 2009 Dec;6(1):94.

- Das S, Cano J, Kalpana GV. Multimerization and DNA binding properties of INI1/hSNF5 and its functional significance. *J Biol Chem*. 2009 Jul 24;284(30):19903-14.
- Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol*. 2002 Jun 21;319(5):1097-113.
- Delaugerre C. Barrière génétique à la résistance des antirétroviraux. Le point sur le raltégravir, premier inhibiteur d'intégrase [Genetic barrier to antiretroviral drug-resistance. Focus on raltegravir, the first integrase inhibitor]. *Med Mal Infect*. 2010 Sep;40 Suppl 1:S1-10. French.
- Dessau MA, Modis Y. Protein crystallization for X-ray crystallography. *J Vis Exp*. 2011 Jan 16;(47).
- Deymier S, Louvat C, Fiorini F, Cimarelli A. ISG20: an enigmatic antiviral RNase targeting multiple viruses. *FEBS Open Bio*. 2022 Jun;12(6):1096–111.
- Dingwall C, Ernberg I, Gait MJ, Green SM, Heaphy S, Karn J, et al. Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. *Proc Natl Acad Sci USA*. 1989 Sep;86(18):6925–9.
- Dingwall C, Ernberg I, Gait MJ, Green SM, Heaphy S, Karn J, et al. HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure. *The EMBO Journal*. 1990 Dec;9(12):4145–53.
- Dobard CW, Briones MS, Chow SA. Molecular Mechanisms by Which Human Immunodeficiency Virus Type 1 Integrase Stimulates the Early Steps of Reverse Transcription. *J Virol*. 2007 Sep 15;81(18):10037–46.
- Doyle T, Goujon C, Malim MH. HIV-1 and interferons: who's interfering with whom? *Nat Rev Microbiol*. 2015 Jul;13(7):403–13.
- Drenth J. Principles of Protein X-Ray Crystallography. New York: Springer Science+Business Media, 2007. 332 p.
- Dreyfus M, Régnier P. The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell*. 2002 Nov 27;111(5):611-3.
- Dubé M, Paquay C, Roy BB, Bego MG, Mercier J, Cohen ÉA. HIV-1 Vpu Antagonizes BST-2 by Interfering Mainly with the Trafficking of Newly Synthesized BST-2 to the Cell Surface. *Traffic*. 2011 Dec;12(12):1714–29.
- Dutrieux J, Maarifi G, Portilho DM, Arhel NJ, Chelbi-Alix MK, Nisole S. PML/TRIM19-Dependent Inhibition of Retroviral Reverse-Transcription by Daxx. Aiken C, editor. *PLoS Pathog*. 2015 Nov 13;11(11):e1005280.
- Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Daviest DR. Crystal Structure of the Catalytic Domain of HIV-1 Integras: Similarity to Other Polynucleotidyl Transferases. 1994;266.
- [E]**
- Eijkelenboom APAM, Puras Lutzke RA, Boelens R, Plasterk RHA, Kaptein R, Hård K. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nat Struct Mol Biol*. 1995 Sep;2(9):807–10.
- El Kazzi P, Rabah N, Chamontin C, Poulain L, Ferron F, Debart F, et al. Internal RNA 2'O-methylation in the HIV-1 genome counteracts ISG20 nuclease-mediated antiviral effect. *Nucleic Acids Research*. 2022 Nov 10;gkac996.
- Elliott JL, Eschbach JE, Koneru PC, Li W, Puray-Chavez M, Townsend D, et al. Integrase-RNA interactions underscore the critical role of integrase in HIV-1 virion morphogenesis. *eLife*. 2020 Sep 22;9:e54311.
- Elliott JL, Kutluay SB. Going beyond Integration: The Emerging Role of HIV-1 Integrase in Virion Morphogenesis. *Viruses*. 2020 Sep 9;12(9):1005.

Engelman A, Cherepanov P. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat Rev Microbiol.* 2012 Mar 16;10(4):279-90.

Engelman A, Kessl JJ, Kvaratskhelia M. Allosteric inhibition of HIV-1 integrase activity. *Current Opinion in Chemical Biology.* 2013 Jun;17(3):339-45.

Engelman AN, Cherepanov P. Retroviral intasomes arising. *Current Opinion in Structural Biology.* 2017 Dec;47:23-9.

Engelman AN. Multifaceted HIV integrase functionalities and therapeutic strategies for their inhibition. *Journal of Biological Chemistry.* 2019 Oct;294(41):15137-57.

Engelman AN, Kvaratskhelia M. Multimodal Functionalities of HIV-1 Integrase. *Viruses.* 2022 Apr 28;14(5):926.

Espert L, Degols G, Gongora C, Blondel D, Williams BR, Silverman RH, Mechti N. ISG20, a new interferon-induced RNase specific for single-stranded RNA, defines an alternative antiviral pathway against RNA genomic viruses. *J Biol Chem.* 2003 May 2;278(18):16151-8.

Espert L, Degols G, Lin YL, Vincent T, Benkirane M, Mechti N. Interferon-induced exonuclease ISG20 exhibits an antiviral activity against human immunodeficiency virus type 1. *Journal of General Virology.* 2005 Aug 1;86(8):2221-9.

## [F]

Falke S, Betzel C. Dynamic Light Scattering (DLS). *Radiation in Bioanalysis.* 2019 Jul 29;8:173-93.

Fanales-Belasio E, Raimondo M, Suligoi B, Buttò S. HIV virology and pathogenetic mechanisms of infection: a brief overview. *Ann Ist Super Sanità* [Internet]. 2010 Mar [cited 2023 Mar 13];46(1). Available from: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0021-25712010000100002&lng=en&nrm=iso&tlng=en](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0021-25712010000100002&lng=en&nrm=iso&tlng=en)

Farnet CM, Bushman FD. HIV-1 cDNA Integration: Requirement of HMG I(Y) Protein for Function of Preintegration Complexes In Vitro. *Cell.* 1997 Feb;88(4):483-92.

Feng L, Larue RC, Slaughter A, Kessl JJ, Kvaratskhelia M. HIV-1 integrase multimerization as a therapeutic target. *Curr Top Microbiol Immunol.* 2015;389:93-119.

Fontana J, Jurado KA, Cheng N, Ly NL, Fuchs JR, Gorelick RJ, et al. Distribution and Redistribution of HIV-1 Nucleocapsid Protein in Immature, Mature, and Integrase-Inhibited Virions: a Role for Integrase in Maturation. Sundquist WI, editor. *J Virol.* 2015 Oct;89(19):9765-80.

Fortner A, Bucur O. mRNA-based vaccine technology for HIV.

Fujinaga K, Taube R, Wimmer J, Cujec TP, Peterlin BM. Interactions between human cyclin T, Tat, and the transactivation response element (TAR) are disrupted by a cysteine to tyrosine substitution found in mouse cyclin T. *Proc Natl Acad Sci USA.* 1999 Feb 16;96(4):1285-90.

## [G]

Gallay P, Hope T, Chin D, Trono D. HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proc Natl Acad Sci USA.* 1997 Sep 2;94(18):9825-30.

Ganser-Pornillos BK, Chandrasekaran V, Pornillos O, Sodroski JG, Sundquist WI, Yeager M. Hexagonal assembly of a restricting TRIM5 $\alpha$  protein. *Proc Natl Acad Sci USA.* 2011 Jan 11;108(2):534-9.

Ganser-Pornillos BK, Pornillos O. Restriction of HIV-1 and other retroviruses by TRIM5. *Nat Rev Microbiol.* 2019 Sep;17(9):546-56.

Garber ME, Wei P, KewalRamani VN, Mayall TP, Herrmann CH, Rice AP, et al. The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. *Genes Dev.* 1998 Nov 15;12(22):3512-27.

- Ge H. Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *The EMBO Journal*. 1998 Nov 16;17(22):6723–9.
- Goldstone DC, Ennis-Adeniran V, Hedden JJ, Groom HCT, Rice GI, Christodoulou E, et al. HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature*. 2011 Dec;480(7377):379–82.
- Goujon C, Darlix JL, Cimarelli A. La protéine virale Vpx des lentivirus de la lignée VIH-2/VISSM [The vpx protein of HIV-2]. *Virologie (Montrouge)*. 2009 Oct 1;13(5):259-269. French.
- Goujon C, Moncorgé O, Bauby H, Doyle T, Ward CC, Schaller T, et al. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature*. 2013 Oct;502(7472):559–62.
- Grandgenett DP, Aihara H. Oligomerization of Retrovirus Integrases. In: Harris JR, Bhella D, editors. *Virus Protein and Nucleoprotein Complexes* [Internet]. Singapore: Springer Singapore; 2018 [cited 2023 Mar 13]. p. 211–43. (Subcellular Biochemistry; vol. 88). Available from: [http://link.springer.com/10.1007/978-981-10-8456-0\\_10](http://link.springer.com/10.1007/978-981-10-8456-0_10)
- Grawenhoff J, Engelman AN. Retroviral integrase protein and intasome nucleoprotein complex structures. *WJBC*. 2017;8(1):32
- Gu J, Babayeva ND, Suwa Y, Baranovskiy AG, Price DH, Tahirov TH. Crystal structure of HIV-1 Tat complexed with human P-TEFb and AFF4. *Cell Cycle*. 2014 Jun;13(11):1788–97.
- Gupta K, Brady T, Dyer BM, Malani N, Hwang Y, Male F, Nolte RT, Wang L, Velthuisen E, Jeffrey J, Van Duyne GD, Bushman FD. Allosteric inhibition of human immunodeficiency virus integrase: late block during viral replication and abnormal multimerization involving specific protein domains. *J Biol Chem*. 2014 Jul 25;289(30):20477-88.
- Gupta K, Turkki V, Sherrill-Mix S, Hwang Y, Eilers G, Taylor L, McDanal C, Wang P, Temelkoff D, Nolte RT, Velthuisen E, Jeffrey J, Van Duyne GD, Bushman FD. Structural Basis for Inhibitor-Induced Aggregation of HIV Integrase. *PLoS Biol*. 2016 Dec 9;14(12):e1002584.
- [H]**
- Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature*. 2010 Mar;464(7286):232–6.
- Hare S, Maertens GN, Cherepanov P. 3'-Processing and strand transfer catalysed by retroviral integrase in crystallo: Integration reactions in crystallo. *The EMBO Journal*. 2012 Jun 29;31(13):3020–8.
- Harris RS, Hultquist JF, Evans DT. The Restriction Factors of Human Immunodeficiency Virus. *Journal of Biological Chemistry*. 2012 Nov;287(49):40875–83.
- Hauser H, Lopez LA, Yang SJ, Oldenburg JE, Exline CM, Guatelli JC, et al. RHesIeVar-ch1 Vpu and HIV-2 Env counteract BST-2/tetherin by sequestration in a perinuclear compartment. 2010;
- Hehl EA, Joshi P, Kalpana GV, Prasad VR. Interaction between Human Immunodeficiency Virus Type 1 Reverse Transcriptase and Integrase Proteins. *J Virol*. 2004 May 15;78(10):5056–67.
- HIV.gov. A Timeline of HIV and AIDS. [En ligne]. 2022. [consulté le 24 mars 2023]. <https://www.hiv.gov/hiv-basics/overview/history/hiv-and-aids-timeline/>
- HIV.gov. Long-Acting HIV Prevention Tools. [En ligne]. 2022. [consulté le 13 mars 2023]. <https://www.hiv.gov/hiv-basics/hiv-prevention/potential-future-options/long-acting-prep>
- Horio T, Murai M, Inoue T, Hamasaki T, Tanaka T, Ohgi T. Crystal structure of human ISG20, an interferon-induced antiviral ribonuclease. *FEBS Letters*. 2004 Nov 5;577(1–2):111–6.
- Hrecka K, Hao C, Gierszewska M, Swanson SK, Kesik-Brodacka M, Srivastava S, et al. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature*. 2011 Jun;474(7353):658–61.

Hurley JH, Hanson PI. Membrane budding and scission by the ESCRT machinery: it's all in the neck. *Nat Rev Mol Cell Biol.* 2010 Aug;11(8):556–66.

Huthoff H, Malim MH. Cytidine deamination and resistance to retroviral infection: towards a structural understanding of the APOBEC proteins. *Virology.* 2005 Apr;334(2):147–53.

Huynh K, Kahwaji CI. HIV Testing. 2022 Apr 21. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan.

Hwang LN, Englund N, Pattnaik AK. Polyadenylation of vesicular stomatitis virus mRNA dictates efficient transcription termination at the intercistronic gene junctions. *J Virol.* 1998 Mar;72(3):1805-13.

## [I]

Imam H, Kim GW, Mir SA, Khan M, Siddiqui A. Interferon-stimulated gene 20 (ISG20) selectively degrades N6-methyladenosine modified Hepatitis B Virus transcripts. Hu J, editor. *PLoS Pathog.* 2020 Feb 14;16(2):e1008338.

Inserm. Sida et VIH, A quand la guérison ? [En ligne]. 2018. [consulté le 13 mars 2023]. <https://www.inserm.fr/dossier/sida-et-vih/>

Institut Pasteur. Recherche d'aujourd'hui, médecine de demain. Paris : Editions de La Martinière ; 2017. 207 p.

Institut Pasteur. SIDA/VIH. [En ligne]. 2021. [consulté le 13 mars 2023]. <https://www.pasteur.fr/fr/centre-medical/fiches-maladies/sida-vih>

International Committee on Taxonomy of Viruses. Taxonomy Browser. [En ligne]. 2023. [consulté le 13 mars 2023]. <https://ictv.global/taxonomy>

International Committee on Taxonomy of Viruses. ICTV 9th Report (2011) – Retroviridae. [En ligne]. 2023. [consulté le 28 mars 2023]. [https://ictv.global/report\\_9th/RT/Retroviridae](https://ictv.global/report_9th/RT/Retroviridae)

## [J]

Janeway Jr, C.A., Travers, P., Walport, M., et al. Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science, 2001. 752 p.

Jaspart A, Calmels C, Cosnefroy O, Bellecave P, Pinson P, Claverol S, et al. GCN2 phosphorylates HIV-1 integrase and decreases HIV-1 replication by limiting viral integration. *Sci Rep.* 2017 May 23;7(1):2283.

Jeang KT, Shank PR, Rabson AB, Kumar A. Synthesis of functional human immunodeficiency virus tat protein in baculovirus as determined by a cell-cell fusion assay. *J Virol.* 1988 Oct;62(10):3874–8.

Jeang KT, Xiao H, Rich EA. Multifaceted Activities of the HIV-1 Transactivator of Transcription, Tat. *Journal of Biological Chemistry.* 1999 Oct;274(41):28837–40.

Ji X, Wu Y, Yan J, Mehrens J, Yang H, DeLucia M, et al. Mechanism of allosteric activation of SAMHD1 by dGTP. *Nat Struct Mol Biol.* 2013 Nov;20(11):1304–9.

Jiang D, Weidner JM, Qing M, Pan XB, Guo H, Xu C, Zhang X, Birk A, Chang J, Shi PY, Block TM, Guo JT. Identification of five interferon-induced cellular proteins that inhibit west nile virus and dengue virus infections. *J Virol.* 2010 Aug;84(16):8332-41.

Jimenez-Guardeño JM, Apolonia L, Betancor G, Malim MH. Immunoproteasome activation enables human TRIM5 $\alpha$  restriction of HIV-1. *Nat Microbiol.* 2019 Mar 18;4(6):933–40.

Jóźwik IK, Passos DO, Lyumkis D. Structural Biology of HIV Integrase Strand Transfer Inhibitors. *Trends Pharmacol Sci.* 2020 Sep;41(9):611-626.

Jurado KA, Wang H, Slaughter A, Feng L, Kessl JJ, Koh Y, et al. Allosteric integrase inhibitor potency is determined through the inhibition of HIV-1 particle maturation. *Proc Natl Acad Sci USA.* 2013 May 21;110(21):8690–5.

## [K]

Kajaste-Rudnitski A, Marelli SS, Pultrone C, Pertel T, Uchil PD, Mechti N, et al. TRIM22 Inhibits HIV-1 Transcription Independently of Its E3 Ubiquitin Ligase Activity, Tat, and NF- $\kappa$ B-Responsive Long Terminal Repeat Elements. *J Virol*. 2011 May 15;85(10):5183–96.

Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP. Binding and Stimulation of HIV-1 Integrase by a Human Homolog of Yeast Transcription Factor SNF5. *Science*. 1994 Dec 23;266(5193):2002–6.

Kammler S, Otte M, Hauber I, Kjems J, Hauber J, Schaal H. The strength of the HIV-1 3' splice sites affects Rev function. *Retrovirology*. 2006 Dec 4;3:89.

Kane M, Yadav SS, Bitzegeio J, Kutluay SB, Zang T, Wilson SJ, et al. MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature*. 2013 Oct;502(7472):563–6.

Kaneko T. The SH3 domain- a family of versatile peptide- and protein-recognition module. *Front Biosci*. 2008;Volume(13):4938.

Karn J, Stoltzfus CM. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med*. 2012 Feb;2(2):a006916.

Kessl JJ, Jena N, Koh Y, Taskent-Sezgin H, Slaughter A, Feng L, de Silva S, Wu L, Le Grice SF, Engelman A, Fuchs JR, Kvaratskhelia M. Multimode, cooperative mechanism of action of allosteric HIV-1 integrase inhibitors. *J Biol Chem*. 2012 May 11;287(20):16801-11.

Kessl JJ, Kutluay SB, Townsend D, Rebensburg S, Slaughter A, Larue RC, et al. HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. *Cell*. 2016 Aug;166(5):1257-1268.e12.

Khan AH, Morita K, Parquet MDC, Hasebe F, Mathenge EGM, Igarashi A. Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site. *J Gen Virol*. 2002 Dec;83(Pt 12):3075-3084.

Khan AS, Bodem J, Buseyne F, Gessain A, Johnson W, Kuhn JH, Kuzmak J, Lindemann D, Linial ML, Löchelt M, Materniak-Kornas M, Soares MA, Switzer WM. Spumaretroviruses: Updated taxonomy and nomenclature. *Virology*. 2018 Mar;516:158-164. Erratum in: *Virology*. 2019 Feb;528.

King A, Lefkowitz E, Adams M. J., Carstens E. B. *Virus Taxonomy*. [En ligne]. 2011. [consulté le 13 mars 2023]. <https://www.elsevier.com/books/virus-taxonomy/king/978-0-12-384684-6>

Koepke L, Gack MU, Sparrer KM. The antiviral activities of TRIM proteins. *Curr Opin Microbiol*. 2021 Feb;59:50-57.

Kristiansen H, Gad HH, Eskildsen-Larsen S, Despres P, Hartmann R: The oligoadenylate synthetase family: an ancient protein family with multiple antiviral activities. *J Interferon Cytokine Res* 2011, 31:41-47.

Kuiken, C., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J. W., Mullins, J., Wolinsky, S. & Korber, B. (1999). A compilation and analysis of nucleic acid and amino acid sequences. In *Human Retroviruses and AIDS*. Los Alamos, New Mexico: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory.

Kurochkina N, Guha U. SH3 domains: modules of protein–protein interactions. *Biophys Rev*. 2013 Mar;5(1):29–39.

Kvaratskhelia M, Sharma A, Larue RC, Serrao E, Engelman A. Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Research*. 2014 Sep 15;42(16):10209–25.

## [L]

Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Ségéral E, et al. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature*. 2011 Jun;474(7353):654–7.

Lahouassa H, Daddacha W, Hofmann H, Ayinde D, Logue EC, Dragin L, et al. SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol.* 2012 Mar;13(3):223–8.

Lau D, Kwan W, Guatelli J. Role of the Endocytic Pathway in the Counteraction of BST-2 by Human Lentiviral Pathogens. *J Virol.* 2011 Oct;85(19):9834–46.

Le Tortorec A, Neil SJD. Antagonism to and Intracellular Sequestration of Human Tetherin by the Human Immunodeficiency Virus Type 2 Envelope Glycoprotein. *J Virol.* 2009 Nov 15;83(22):11966–78.

Lee SP, Han MK. Zinc Stimulates Mg<sup>2+</sup>-Dependent 3'-Processing Activity of Human Immunodeficiency Virus Type 1 Integrase in Vitro. *Biochemistry.* 1996 Jan 1;35(12):3837–44.

Legrand A, Dahoui C, De La Myre Mory C, Loyer C, Noy K, Guiguet L, et al. The inflammatory and tumor suppressor SAMD9L acts through a Schlafen-like box to restrict HIV and inhibit cell translation in SAAD/ATXPC [Internet]. *Immunology*; 2023 Jan [cited 2023 Mar 13]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2023.01.19.524725>

Leopoldino AM, Canduri F, Cabral H, Junqueira M, de Marqui ABT, Apponi LH, et al. Expression, purification, and circular dichroism analysis of human CDK9. *Protein Expression and Purification.* 2006 Jun;47(2):614–20.

Levin A, Armon-Omer A, Rosenbluh J, Melamed-Book N, Graessmann A, Waigmann E, et al. Inhibition of HIV-1 integrase nuclear import and replication by a peptide bearing integrase putative nuclear localization signal. *Retrovirology.* 2009 Dec;6(1):112.

Levy JA. HIV pathogenesis: 25 years of progress and persistent challenges. *AIDS.* 2009 Jan 14;23(2):147–60.

Li M, Kao E, Gao X, Sandig H, Limmer K, Pavon-Eternod M, et al. Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature.* 2012 Nov;491(7422):125–8.

Li M, Craigie R. Nucleoprotein complex intermediates in HIV-1 integration. *Methods.* 2009 Apr;47(4):237–42.

Li M, Mizuuchi M, Burke TR, Craigie R. Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J.* 2006 Mar 22;25(6):1295–304.

Li YL, Chandrasekaran V, Carter SD, Woodward CL, Christensen DE, Dryden KA, et al. Primate TRIM5 proteins form hexagonal nets on HIV-1 capsids. *eLife.* 2016 Jun 2;5:e16269.

Li YL, Langley CA, Azumaya CM, Echeverria I, Chesarino NM, Emerman M, et al. The structural basis for HIV-1 Vif antagonism of human APOBEC3G. *Nature* [Internet]. 2023 Feb 8 [cited 2023 Mar 13]; Available from: <https://www.nature.com/articles/s41586-023-05779-1>

Liddament MT, Brown WL, Schumacher AJ, Harris RS. APOBEC3F Properties and Hypermutation Preferences Indicate Activity against HIV-1 In Vivo. *Current Biology.* 2004 Aug;14(15):1385–91.

Lin CW, Engelman A. The Barrier-to-Autointegration Factor Is a Component of Functional Human Immunodeficiency Virus Type 1 Preintegration Complexes. *J Virol.* 2003 Apr 15;77(8):5030–6.

Liu S, Koneru PC, Li W, Pathirage C, Engelman AN, Kvaratskhelia M, Musier-Forsyth K. HIV-1 integrase binding to genomic RNA 5'-UTR induces local structural changes in vitro and in virio. *Retrovirology.* 2021 Nov 22;18(1):37.

Liu Z, Pan Q, Ding S, Qian J, Xu F, Zhou J, et al. The Interferon-Inducible MxB Protein Inhibits HIV-1 Infection. *Cell Host & Microbe.* 2013 Oct;14(4):398–410.

Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, et al. Solution Structure of the DNA Binding Domain of HIV-1 Integrase. *Biochemistry.* 1995 Aug 8;34(31):9826–33.

Lu H, Li Z, Xue Y, Zhou Q. Viral–Host Interactions That Control HIV-1 Transcriptional Elongation. *Chem Rev.* 2013 Nov 13;113(11):8567–82.

Lu H, Li Z, Xue Y, Schulze-Gahmen U, Johnson JR, Krogan NJ, et al. AFF1 is a ubiquitous P-TEFb partner to enable Tat extraction of P-TEFb from 7SK snRNP and formation of SECs for HIV transactivation. *Proc Natl Acad Sci USA* [Internet]. 2014 Jan 7 [cited 2023 Mar 13];111(1). Available from: <https://pnas.org/doi/full/10.1073/pnas.1318503111>

## [M]

Ma X, Yang T, Luo Y, Wu L, Jiang Y, Song Z, et al. TRIM28 promotes HIV-1 latency by SUMOylating CDK9 and inhibiting P-TEFb. *eLife*. 2019 Jan 17;8:e42426.

Madison MK, Lawson DQ, Elliott J, Ozantürk AN, Koneru PC, Townsend D, et al. Allosteric HIV-1 Integrase Inhibitors Lead to Premature Degradation of the Viral RNA Genome and Integrase in Target Cells. Sundquist WI, editor. *J Virol*. 2017 Sep;91(17):e00821-17.

Maehigashi T, Ahn S, Kim UI, Lindenberger J, Oo A, Koneru PC, Mahboubi B, Engelman AN, Kvaratskhelia M, Kim K, Kim B. A highly potent and safe pyrrolopyridine-based allosteric HIV-1 integrase inhibitor targeting host LEDGF/p75-integrase interaction site. *PLoS Pathog*. 2021 Jul 22;17(7):e1009671.

Maertens GN, Hare S, Cherepanov P. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature*. 2010 Nov 11;468(7321):326-9.

Magdeldin S, Moser A. Affinity Chromatography: Principles and Applications. *Affinity Chromatography*. [En ligne]. 2012. [consulté le 25 mars 2023]. <http://dx.doi.org/10.5772/39087>

Maignan S, Guilloteau JP, Zhou-Liu Q, Clément-Mella C, Mikol V. Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *J Mol Biol*. 1998 Sep 18;282(2):359-68.

Malim MH, Bieniasz PD. HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harbor Perspectives in Medicine*. 2012 May 1;2(5):a006940–a006940.

Mancebo HSY, Lee G, Flygare J, Tomassini J, Luu P, Zhu Y, et al. P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro. *Genes Dev*. 1997 Oct 15;11(20):2633–44.

Manganaro L, Lusic M, Gutierrez MI, Cereseto A, Del Sal G, Giacca M. Concerted action of cellular JNK and Pin1 restricts HIV-1 genome integration to activated CD4+ T lymphocytes. *Nat Med*. 2010 Mar;16(3):329–33.

Mangus DA, Evans MC, Jacobson A. Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biology*. 2003;4(7):223. doi: 10.1186/gb-2003-4-7-223. Epub 2003 Jul 1.

Maskell DP, Renault L, Serrao E, Lesbats P, Matadeen R, Hare S, et al. Structural basis for retroviral integration into nucleosomes. *Nature*. 2015 Jul;523(7560):366–9.

Masuda T, Kotani O, Yokoyama M, Abe Y, Kawai G, Sato H. Cis-Allosteric Regulation of HIV-1 Reverse Transcriptase by Integrase. *Viruses*. 2022 Dec 21;15(1):31.

Mauro E, Lesbats P, Lapaillerie D, Chaignepain S, Maillot B, Oladosu O, et al. Human H4 tail stimulates HIV-1 integration through binding to the carboxy-terminal domain of integrase. *Nucleic Acids Research*. 2019 Apr 23;47(7):3607–18.

Mitchell RS, Beitzel BF, Schroder ARW, Shinn P, Chen H, Berry CC, et al. Retroviral DNA Integration: ASLV, HIV, and MLV Show Distinct Target Site Preferences. Michael Emerman, editor. *PLoS Biol*. 2004 Aug 17;2(8):e234.

Miyakawa K, Ryo A, Murakami T, Ohba K, Yamaoka S, Fukuda M, et al. BCA2/Rabring7 Promotes Tetherin-Dependent HIV-1 Restriction. Luban J, editor. *PLoS Pathog*. 2009 Dec 18;5(12):e1000700.

Mohammed KD, Topper MB, Muesing MA. Sequential Deletion of the Integrase (Gag-Pol) Carboxyl Terminus Reveals Distinct Phenotypic Classes of Defective HIV-1. *J Virol*. 2011 May 15;85(10):4654–66.

Morita E, Sundquist WI. RETROVIRUS BUDDING. *Annu Rev Cell Dev Biol.* 2004 Nov 1;20(1):395–425.

Mourez T, Simon F, Plantier JC. Non-M variants of human immunodeficiency virus type 1. *Clin Microbiol Rev.* 2013 Jul;26(3):448-61.

Mousnier A, Kubat N, Massias-Simon A, Ségéral E, Rain JC, Benarous R, et al. von Hippel–Lindau binding protein 1-mediated degradation of integrase affects HIV-1 gene expression at a postintegration step. *Proc Natl Acad Sci USA.* 2007 Aug 21;104(34):13615–20.

Mousseau G, Valente ST. Role of Host Factors on the Regulation of Tat-Mediated HIV-1 Transcription. CPD [Internet]. 2017 Nov 2 [cited 2023 Mar 13];23(28). Available from: <http://www.eurekaselect.com/153541/article>

Muniz L, Egloff S, Ughy B, Jády BE, Kiss T. Controlling Cellular P-TEFb Activity by the HIV-1 Transcriptional Transactivator Tat. Cullen BR, editor. *PLoS Pathog.* 2010 Oct 14;6(10):e1001152.

Muniz L, Kiss T, Egloff S. Perturbations de la transcription liées à une dérégulation de P-TEFb : cancer, Sida et hypertrophie cardiaque. *Med Sci (Paris).* 2012 Feb;28(2):200–5.

Mure F, Panthu B, Zanella-Cléon I, Delolme F, Manet E, Ohlmann T, Gruffat H. Epstein-Barr Virus Protein EB2 Stimulates Translation Initiation of mRNAs through Direct Interactions with both Poly(A)-Binding Protein and Eukaryotic Initiation Factor 4G. *J Virol.* 2018 Jan 17;92(3):e01917-17.

Murugesan A, Manoharan M. Dengue Virus. *Emerging and Reemerging Viral Pathogens.* 2020:281–359.

## [N]

Nachtergaele S, He C. The emerging biology of RNA post-transcriptional modifications. *RNA Biology.* 2017 Feb 1;14(2):156–63.

Neil SJD, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature.* 2008 Jan;451(7177):425–30.

Nguyen LH, Espert L, Mechti N, Wilson DM. The Human Interferon- and Estrogen-Regulated ISG20/HEM45 Gene Product Degrades Single-Stranded RNA and DNA in Vitro. *Biochemistry.* 2001 Jun 1;40(24):7174–9.

Nisole S, Stoye JP, Saïb A. TRIM family proteins: retroviral restriction and antiviral defence. *Nat Rev Microbiol.* 2005 Oct;3(10):799-808.

Nowotny M, Yang W. Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *EMBO J.* 2006 May 3;25(9):1924–33.

## [O]

Okumura A, Lu G, Pitha-Rowe I, Pitha PM. Innate antiviral response targets HIV-1 release by the induction of ubiquitin-like protein ISG15. *Proc Natl Acad Sci USA.* 2006 Jan 31;103(5):1440–5.

Ott M, Geyer M, Zhou Q. The control of HIV transcription: keeping RNA polymerase II on track. *Cell Host Microbe.* 2011 Nov 17;10(5):426-35.

## [P]

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. *Journal of Biological Chemistry.* 2004 Nov;279(46):48397–403.

Palais Découverte. Le virus du SIDA. [En ligne]. 2018 [13 mars 2023]. <https://www.palais-decouverte.fr/fr/ressources/les-themes-1c1m/le-virus-du-sida>

Parekh BS, Ou CY, Fonjungo PN, Kalou MB, Rottinghaus E, Puren A, et al. Diagnosis of Human Immunodeficiency Virus Infection. *Clin Microbiol Rev.* 2018 Dec 19;32(1):e00064-18.

Passos DO, Li M, Craigie R, Lyumkis D. Retroviral integrase: Structure, mechanism, and inhibition. In: *The Enzymes* [Internet]. Elsevier; 2021 [cited 2023 Mar 13]. p. 249–300. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S187460472100007X>

Passos DO, Li M, Józwiak IK, Zhao XZ, Santos-Martins D, Yang R, Smith SJ, Jeon Y, Forli S, Hughes SH, Burke TR Jr, Craigie R, Lyumkis D. Structural basis for strand-transfer inhibitor binding to HIV intasomes. *Science*. 2020 Feb 14;367(6479):810-814.

Passos DO, Li M, Yang R, Rebersburg SV, Ghirlando R, Jeon Y, et al. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science*. 2017 Jan 6;355(6320):89–92.

Perez-Caballero D, Zang T, Ebrahimi A, McNatt MW, Gregory DA, Johnson MC, et al. Tetherin Inhibits HIV-1 Release by Directly Tethering Virions to Cells. *Cell*. 2009 Oct;139(3):499–511.

Pincetic A, Kuang Z, Seo EJ, Leis J. The Interferon-Induced Gene ISG15 Blocks Retrovirus Release from Cells Late in the Budding Process. *J Virol*. 2010 May;84(9):4725–36.

Ponchon L, Beauvais G, Nonin-Lecomte S, Dardel F. A generic protocol for the expression and purification of recombinant RNA in *Escherichia coli* using a tRNA scaffold. *Nat Protoc*. 2009 Jun;4(6):947–59.

Ponchon L, Dardel F. Recombinant RNA technology: the tRNA scaffold. *Nat Methods*. 2007 Jul;4(7):571–6.

## [R]

Rampalli S, Kulkarni A, Kumar P, Mogare D, Galande S, Mitra D, Chattopadhyay S. Stimulation of Tat-independent transcriptional processivity from the HIV-1 LTR promoter by matrix attachment regions. *Nucleic Acids Res*. 2003 Jun 15;31(12):3248-56.

Reece J. B., Urry L. A., Cain M. L., et al. *Campbell biology*. 9th Edition. Frenchs Forest, NSW : Pearson, 2010. 1464 p.

Rezende LF, Prasad VR. Nucleoside-analog resistance mutations in HIV-1 reverse transcriptase and their influence on polymerase fidelity and viral mutation rates. *Int J Biochem Cell Biol*. 2004 Sep;36(9):1716-34.

Rice AP. The HIV-1 Tat Protein: Mechanism of Action and Target for HIV-1 Cure Strategies. *Curr Pharm Des*. 2017;23(28):4098-4102.

Rice P, Craigie R, Davies DR. Retroviral integrases and their cousins. *Curr Opin Struct Biol*. 1996 Feb;6(1):76-83.

Richter S, Ping YH, Rana TM. TAR RNA loop: A scaffold for the assembly of a regulatory switch in HIV replication. *Proc Natl Acad Sci USA*. 2002 Jun 11;99(12):7928–33.

Rocchi C, Gouet P, Parissi V, Fiorini F. The C-Terminal Domain of HIV-1 Integrase: A Swiss Army Knife for the Virus? *Viruses*. 2022 Jun 27;14(7):1397.

Rocchi C, Louvat C, Miele AE, Batisse J, Guillon C, Ballut L, et al. The HIV-1 Integrase C-Terminal Domain Induces TAR RNA Structural Changes Promoting Tat Binding. *IJMS*. 2022 Nov 8;23(22):13742.

Rogez S. Structure, réplication et variabilité des virus de l'immunodéficience humaine. *Actualités Pharmaceutiques*. 2022 Sep;61(618):22–6.

Romani N, Gruner S, Brang D, Kämpgen E, Lenz A, Trockenbacher B, Konwalinka G, Fritsch PO, Steinman RM, Schuler G. Proliferating dendritic cell progenitors in human blood. *J Exp Med*. 1994 Jul 1;180(1):83-93.

Rossi E, Meuser ME, Cunanan CJ, Cocklin S. Structure, Function, and Interactions of the HIV-1 Capsid Protein. *Life*. 2021 Jan 29;11(2):100.

Rozina A, Anisenko A, Kikhai T, Silkina M, Gottikh M. Complex Relationships between HIV-1 Integrase and Its Cellular Partners. *IJMS*. 2022 Oct 15;23(20):12341.

Ryoo J, Choi J, Oh C, Kim S, Seo M, Kim SY, et al. The ribonuclease activity of SAMHD1 is required for HIV-1 restriction. *Nat Med*. 2014 Aug;20(8):936–41.

## [S]

Sanchez-Pescador R, Power MD, Barr PJ, Steimer KS, Stempien MM, Brown-Shimer SL, et al. Nucleotide Sequence and Expression of an AIDS-Associated Retrovirus (ARV-2). *Science*. 1985 Feb;227(4686):484–92.

SARTORIUS. Bio-Layer Interferometry (BLI) Label-Free Technology. [En ligne]. 2023. [consulté le 25 mars 2023]. <https://www.sartorius.com/en/applications/life-science-research/label-free-detection/bli-technology>

Sato K, Takeuchi JS, Misawa N, Izumi T, Kobayashi T, Kimura Y, et al. APOBEC3D and APOBEC3F Potently Promote HIV-1 Diversification and Evolution in Humanized Mouse Model. Ross SR, editor. *PLoS Pathog*. 2014 Oct 16;10(10):e1004453.

Shafer RW, Vuitton DA. Highly active antiretroviral therapy (HAART) for the treatment of infection with human immunodeficiency virus type 1. *Biomed Pharmacother*. 1999 Mar;53(2):73-86.

Schoggins JW. Interferon-stimulated genes: roles in viral pathogenesis. *Current Opinion in Virology*. 2014 Jun;6:40–6.

Schröder ARW, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell*. 2002 Aug;110(4):521–9.

Schulze-Gahmen U, Echeverria I, Stjepanovic G, Bai Y, Lu H, Schneidman-Duhovny D, et al. Insights into HIV-1 proviral transcription from integrative structure and dynamics of the Tat:AFF4:P-TEFb:TAR complex. *eLife*. 2016 Oct 12;5:e15910.

Schulze-Gahmen U, Hurley JH. Structural mechanism for HIV-1 TAR loop recognition by Tat and the super elongation complex. *Proc Natl Acad Sci USA*. 2018 Dec 18;115(51):12973–8.

Schulze-Gahmen U, Lu H, Zhou Q, Alber T. AFF4 binding to Tat-P-TEFb indirectly stimulates TAR recognition of super elongation complexes at the HIV promoter. *eLife*. 2014 Apr 24;3:e02375.

Schulze-Gahmen U, Upton H, Birnberg A, Chou S, Krogan NJ, Zhou Q, et al. The AFF4 scaffold binds human P-TEFb adjacent to HIV Tat. *Genes and chromosomes*.

Sedore SC, Byers SA, Biglione S, Price JP, Maury WJ, Price DH. Manipulation of P-TEFb control machinery by HIV: recruitment of P-TEFb from the large form by Tat and binding of HEXIM1 to TAR. *Nucleic Acids Research*. 2007 Jun 18;35(13):4347–58.

Sego TJ, Aponte-Serrano JO, Ferrari Gianlupi J, Heaps SR, Breithaupt K, Bruschi L, et al. A modular framework for multiscale, multicellular, spatiotemporal modeling of acute primary viral infection and immune response in epithelial tissues and its application to drug therapy timing and effectiveness. Peirce SM, editor. *PLoS Comput Biol*. 2020 Dec 21;16(12):e1008451.

Sertznig H, Hillebrand F, Erkelenz S, Schaal H, Widera M. Behind the scenes of HIV-1 replication: Alternative splicing as the dependency factor on the quiet. *Virology*. 2018 Mar;516:176-188.

Shaban NM, Shi K, Lauer KV, Carpenter MA, Richards CM, Salamango D, et al. The Antiviral and Cancer Genomic DNA Deaminase APOBEC3H Is Regulated by an RNA-Mediated Dimerization Mechanism. *Molecular Cell*. 2018 Jan;69(1):75-86.e9.

Sharma A, Slaughter A, Jena N, Feng L, Kessl JJ, Fadel HJ, et al. A New Class of Multimerization Selective Inhibitors of HIV-1 Integrase. Cullen BR, editor. *PLoS Pathog*. 2014 May 29;10(5):e1004171.

Shatkin AJ, Manley JL. The ends of the affair: capping and polyadenylation. *Nat Struct Biol*. 2000 Oct;7(10):838-42.

Shehu-Xhilaga M, Crowe SM, Mak J. Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J Virol*. 2001 Feb;75(4):1834-41.

- Shema Mugisha C, Dinh T, Kumar A, Tenneti K, Eschbach JE, Davis K, et al. Emergence of Compensatory Mutations Reveals the Importance of Electrostatic Interactions between HIV-1 Integrase and Genomic RNA. Jurado K, editor. *mBio*. 2022 Oct 26;13(5):e00431-22.
- Shen Q, Wu C, Freniere C, Tripler TN, Xiong Y. Nuclear Import of HIV-1. *Viruses*. 2021 Nov 8;13(11):2242.
- Shinohara T, Singh DP, Fatma N. LEDGF, a survival factor, activates stress-related genes. *Progress in Retinal and Eye Research*. 2002 May;21(3):341–58.
- Sidaction. VIH/SIDA : Qu'est-ce que c'est ? [En ligne]. [consulté le 13 mars 2023]. <https://www.sidaction.org/vihsida-quest-ce-que-cest>
- Sim G, Kehling AC, Park MS, Secor J, Divoky C, Zhang H, et al. Manganese-dependent microRNA trimming by 3'→5' exonucleases generates 14-nucleotide or shorter tiny RNAs. *Proc Natl Acad Sci USA*. 2022 Dec 20;119(51):e2214335119.
- Smerdon SJ, Jager J, Wang J, Kohlstaedt LA, Chirino AJ, FRIEDMAN JM, et al. Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1.
- Sobhian B, Laguette N, Yatim A, Nakamura M, Levy Y, Kiernan R, et al. HIV-1 Tat Assembles a Multifunctional Transcription Elongation Complex and Stably Associates with the 7SK snRNP. *Molecular Cell*. 2010 May;38(3):439–51.
- Strauss JH, Strauss EG. Viruses Whose Life Cycle Uses Reverse Transcriptase. In: *Viruses and Human Disease* [Internet]. Elsevier; 2008 [cited 2023 Mar 13]. p. 211–259. Available from: <https://linkinghub.elsevier.com/retrieve/pii/B978012373741050009X>
- Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. The cytoplasmic body component TRIM5a restricts HIV-1 infection in Old World monkeys. 2004;427.
- Sundquist WI, Krausslich HG. HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harbor Perspectives in Medicine*. 2012 Jul 1;2(7):a006924–a006924.
- [T]
- Tabah AA, Tardif K, Mansky LM. Anti-HIV-1 activity of Trim 37. *Journal of General Virology*. 2014 Apr 1;95(4):960–7.
- Tabasi M, Nombela I, Janssens J, Lahousse AP, Christ F, Debyser Z. Role of Transportin-SR2 in HIV-1 Nuclear Import. *Viruses*. 2021 May 4;13(5):829.
- Tahirov TH, Babayeva ND, Varzavand K, Cooper JJ, Sedore SC, Price DH. Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature*. 2010 Jun;465(7299):747–51.
- Tahirov TH, Babayeva ND, Varzavand K, Cooper JJ, Sedore SC, Price DH. Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature*. 2010 Jun;465(7299):747–51.
- Tekeste SS, Wilkinson TA, Weiner EM, Xu X, Miller JT, Le Grice SFJ, et al. Interaction between Reverse Transcriptase and Integrase Is Required for Reverse Transcription during HIV-1 Replication. Beemon KL, editor. *J Virol*. 2015 Dec;89(23):12058–69.
- Terreni M, Valentini P, Liverani V, Gutierrez MI, Primio CD, Fenza AD, et al. GCN5-dependent acetylation of HIV-1 integrase enhances viral integration. 2010;
- Teyra J, Huang H, Jain S, Guan X, Dong A, Liu Y, Tempel W, Min J, Tong Y, Kim PM, Bader GD, Sidhu SS. Comprehensive Analysis of the Human SH3 Domain Family Reveals a Wide Variety of Non-canonical Specificities. *Structure*. 2017 Oct 3;25(10):1598-1610.e3.
- Tissot C, Mechti N. Molecular Cloning of a New Interferon-induced Factor That Represses Human Immunodeficiency Virus Type 1 Long Terminal Repeat Expression. *Journal of Biological Chemistry*. 1995 Jun;270(25):14891–8.

Tüngler V, Staroske W, Kind B, Dobrick M, Kretschmer S, Schmidt F, et al. Single-stranded nucleic acids promote SAMHD1 complex formation. *J Mol Med*. 2013 Jun;91(6):759–70.

## [U]

Uchil PD, Quinlan BD, Chan WT, Luna JM, Mothes W. TRIM E3 Ligases Interfere with Early and Late Stages of the Retroviral Life Cycle. Luban J, editor. *PLoS Pathog*. 2008 Feb 1;4(2):e16.

UNAIDS. Prevention of Mother to Child Transmission (PMTCT). 2021. [consulté le 24 mars 2023]. <https://www.usaid.gov/global-health/health-areas/hiv-and-aids/technical-areas/pmtct>

UNAIDS. UNAIDS. [En ligne]. 2023. [consulté le 24 mars 2023]. <https://www.unaids.org>

## [V]

Vaccari M, Poonam P, Franchini G. Phase III HIV vaccine trial in Thailand: a step toward a protective vaccine for HIV. *Expert Rev Vaccines*. 2010 Sep;9(9):997-1005.

Valkov E, Gupta SS, Hare S, Helander A, Roversi P, McClure M, Cherepanov P. Functional and structural characterization of the integrase from the prototype foamy virus. *Nucleic Acids Res*. 2009 Jan;37(1):243-55.

Van Cor-Hosmer SK, Kim DH, Daly MB, Daddacha W, Kim B. Restricted 5'-end gap repair of HIV-1 integration due to limited cellular dNTP concentrations in human primary macrophages. *J Biol Chem*. 2013 Nov 15;288(46):33253-62.

van Gent M, Sparrer KMJ, Gack MU. TRIM Proteins and Their Roles in Antiviral Host Defenses. *Annu Rev Virol*. 2018 Sep 29;5(1):385–405.

van Gent M, Sparrer KMJ, Gack MU. TRIM Proteins and Their Roles in Antiviral Host Defenses. *Annu Rev Virol*. 2018 Sep 29;5(1):385–405.

van Heuvel Y, Schatz S, Rosengarten JF, Stitz J. Infectious RNA: Human Immunodeficiency Virus (HIV) Biology, Therapeutic Intervention, and the Quest for a Vaccine. *Toxins*. 2022 Feb 14;14(2):138.

van Tol S, Hage A, Giraldo M, Bharaj P, Rajsbaum R. The TRIMendous Role of TRIMs in Virus–Host Interactions. *Vaccines*. 2017 Aug 22;5(3):23.

Vansant G, Bruggemans A, Janssens J, Debyser Z. Block-And-Lock Strategies to Cure HIV Infection. *Viruses*. 2020 Jan 10;12(1):84.

vih.org. Recherche vaccinale — Essai thaïlandais RV144 : le droit d’espérer. [En ligne]. 2010. [consulté le 27 mars 2023]. <https://vih.org/20100106/essai-thailandais-rv144-le-droit-desperer/>

vih.org. Vaccination contre le VIH : où en est-on ? [En ligne]. 2021. [consulté le 13 mars 2023]. <https://vih.org/20211129/vaccination-contre-le-vih-ou-en-est-on/>

ViralZone. The Baltimore classification. [En ligne]. [consulté le 13 mars 2023]. <https://viralzone.expasy.org/254>

Vunjak M, Versteeg GA. TRIM proteins. *Current Biology*. 2019 Jan;29(2):R42–4.

## [W]

Wain-Hobson S, Sonigo P, Danos O, Cole S, Alizon M. Nucleotide sequence of the AIDS virus, LAV. *Cell*. 1985 Jan;40(1):9–17.

Wang JY, Ling H, Yang W, Craigie R. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J*. 2001 Dec 17;20(24):7333-43.

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*. 1998 Feb;92(4):451–62.

Weiss CM, Trobaugh DW, Sun C, Lucas TM, Diamond MS, Ryman KD, Klimstra WB. The Interferon-Induced Exonuclease ISG20 Exerts Antiviral Activity through Upregulation of Type I Interferon Response Proteins. *mSphere*. 2018 Sep 19;3(5):e00209-18.

White TE, Brandariz-Nuñez A, Valle-Casuso JC, Amie S, Nguyen L, Kim B, et al. Contribution of SAM and HD domains to retroviral restriction mediated by human SAMHD1. *Virology*. 2013 Feb;436(1):81–90.

Wikipédia. Virus de l'immunodéficience. [En ligne]. 2023. [consulté le 13 mars 2023]. [https://fr.wikipedia.org/wiki/Virus\\_de\\_l%27immunod%C3%A9ficience\\_humaine](https://fr.wikipedia.org/wiki/Virus_de_l%27immunod%C3%A9ficience_humaine)

Wilkinson TA, Januszyk K, Phillips ML, Tekeste SS, Zhang M, Miller JT, et al. Identifying and Characterizing a Functional HIV-1 Reverse Transcriptase-binding Site on Integrase. *Journal of Biological Chemistry*. 2009 Mar;284(12):7931–9.

Winans S, Goff SP. Mutations altering acetylated residues in the CTD of HIV-1 integrase cause defects in proviral transcription at early times after integration of viral DNA. Krausslich HG, editor. *PLoS Pathog*. 2020 Dec 22;16(12):e1009147.

Wu N, Nguyen XN, Wang L, Appourchaux R, Zhang C, Panthou B, et al. The interferon stimulated gene 20 protein (ISG20) is an innate defense antiviral factor that discriminates self versus non-self translation. Dittmann M, editor. *PLoS Pathog*. 2019 Oct 10;15(10):e1008093.

Wu X, Liu H, Xiao H, Conway JA, Hehl E, Kalpana GV, et al. Human Immunodeficiency Virus Type 1 Integrase Protein Promotes Reverse Transcription through Specific Interactions with the Nucleoprotein Reverse Transcription Complex. *J Virol*. 1999 Mar;73(3):2126–35.

Wymant C, Bezemer D, Blanquart F, Ferretti L, Gall A, Hall M, et al. A highly virulent variant of HIV-1 circulating in the Netherlands. 2022;

## [Y]

Yan J, Kaur S, DeLucia M, Hao C, Mehrens J, Wang C, et al. Tetramerization of SAMHD1 Is Required for Biological Activity and Inhibition of HIV Infection. *Journal of Biological Chemistry*. 2013 Apr;288(15):10406–17.

Yang W, Lee JY, Nowotny M. Making and Breaking Nucleic Acids: Two-Mg<sup>2+</sup>-Ion Catalysis and Substrate Specificity. *Molecular Cell*. 2006 Apr;22(1):5–13.

Yang W, Steitz TA. Recombining the structures of HIV integrase, RuvC and RNase H. *Structure*. 1995 Feb;3(2):131–4.

Yang ZN, Mueser TC, Bushman FD, Hyde CC. Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J Mol Biol*. 2000 Feb 18;296(2):535-48.

Yin X, Langer S, Zhang Z, Herbert KM, Yoh S, König R, et al. Sensor Sensibility—HIV-1 and the Innate Immune Response. *Cells*. 2020 Jan 20;9(1):254.

Yin Z, Shi K, Banerjee S, Pandey KK, Bera S, Grandgenett DP, et al. Crystal structure of the Rous sarcoma virus intasome. *Nature*. 2016 Feb;530(7590):362–6.

Yu A, Skorupka KA, Pak AJ, Ganser-Pornillos BK, Pornillos O, Voth GA. TRIM5 $\alpha$  self-assembly and compartmentalization of the HIV-1 viral capsid. *Nat Commun*. 2020 Mar 11;11(1):1307.

## [Z]

Zaitseva L, Cherepanov P, Leyens L, Wilson SJ, Rasaiyaah J, Fassati A. HIV-1 exploits importin 7 to maximize nuclear import of its DNA genome. *Retrovirology*. 2009 Dec;6(1):11.

Zamborlini A, Coiffic A, Beauclair G, Delelis O, Paris J, Koh Y, et al. Impairment of Human Immunodeficiency Virus Type-1 Integrase SUMOylation Correlates with an Early Replication Defect. *Journal of Biological Chemistry*. 2011 Jun;286(23):21013–22.

Zhang F, Landford WN, Ng M, McNatt MW, Bieniasz PD, Hatzioannou T. SIV Nef Proteins Recruit the AP-2 Complex to Antagonize Tetherin and Facilitate Virion Release. Desrosiers RC, editor. PLoS Pathog. 2011 May 19;7(5):e1002039.

Zhang F, Wilson SJ, Landford WC, Virgen B, Gregory D, Johnson MC, et al. Nef Proteins from Simian Immunodeficiency Viruses Are Tetherin Antagonists. Cell Host & Microbe. 2009 Jul;6(1):54–67.

Zhang J, Crumpacker C. HIV UTR, LTR, and Epigenetic Immunity. Viruses. 2022 May 18;14(5):1084.

Zheng R, Jenkins TM, Craigie R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. Proc Natl Acad Sci USA. 1996 Nov 26;93(24):13659–64.

Zheng Y, Yao X. Posttranslational Modifications of HIV-1 Integrase by Various Cellular Proteins during Viral Replication. Viruses. 2013 Jul 16;5(7):1787–801.

Zhou Q, Yik JHN. The Yin and Yang of P-TEFb Regulation: Implications for Human Immunodeficiency Virus Gene Expression and Global Control of Cell Growth and Differentiation. Microbiol Mol Biol Rev. 2006 Sep;70(3):646–59.

Zila V, Margiotta E, Turoňová B, Müller TG, Zimmerli CE, Mattei S, et al. Cone-shaped HIV-1 capsids are transported through intact nuclear pores. Cell. 2021 Feb;184(4):1032-1046.e18.

Zolla-Pazner S, Michael NL, Kim JH. A tale of four studies: HIV vaccine immunogenicity and efficacy in clinical trials. Lancet HIV. 2021 Jul;8(7):e449-e452.

#### Pour les outils bio-informatiques :

BioRender, <https://app.biorender.com>

Clustal Omega, <https://www.ebi.ac.uk/Tools/msa/clustalo/>

ENDscript 2.0, <https://endscript.ibcp.fr/ESPrript/ENDscript/>

PyMoL, <https://pymol.org/2/>

RNAfold web server, <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>

## Contributions scientifiques en lien avec les travaux de thèse

### Articles

- Article de revue :

Deymier S, Louvat C, Fiorini F, Cimarelli A. *ISG20: an enigmatic antiviral RNase targeting multiple viruses*. *FEBS Open Bio*. Accepté en février 2022. <https://doi.org/10.1002/2211-5463.13382>

- Article de recherche en co-premier auteur :

Rocchi C $\S$ , Louvat C $\S$ , Miele AE, Batisse J, Guillon C, Ballut L, Lener D, Negroni M, Ruff M, Gouet P, Fiorini F. *The HIV-1 Integrase C-Terminal Domain Induces TAR RNA Structural Changes Promoting Tat Binding*. *International Journal of Molecular Sciences*. Accepté en novembre 2022. <https://doi.org/10.3390/ijms232213742>

### Communications orales

*The HIV-1 Integrase C-Terminal domain induces TAR RNA structural changes promoting Tat binding*. 27<sup>ème</sup> édition de la journée scientifique de l'Ecole Doctorale Interdisciplinaire Sciences Santé (ED205). En octobre 2022, Lyon.

*The HIV-1 Integrase C-Terminal domain induces TAR RNA structural changes promoting Tat binding*. Séminaire interne de l'Institut de Biologie et Chimie des Protéines. En novembre 2022, Lyon.

Les interactions protéine – ARN au cours de l'infection par le VIH. Séminaire externe de la Structure Fédérative de Recherche – Biosciences. En novembre 2022, Lyon.

### Communications sous forme de poster

*Structural and functional study of the HIV-1 integrase in complex with viral RNA*. 9<sup>ème</sup> Ecole Nationale de Biologie Structurale Intégrative organisée par le Réseau National de Formation en Biologie Structurale Intégrative. En juin 2022, Oléron.

*The HIV-1 Integrase C-Terminal domain induces TAR RNA structural changes promoting Tat binding*. 12<sup>ème</sup> édition du congrès Structure, Intégration, Fonction et Réactivité de l'ARN. En octobre 2022, Bordeaux.

## Annexes

### Détails techniques pour l'électrophorèse d'ARN sur gel de polyacrylamide

Dimension des gels de petite taille : 8 cm x 10 cm (longueur x largeur). Ces gels sont utilisés pour contrôler la qualité des ARN. (Remarque : ce sont ces gels qui sont utilisés pour les électrophorèses de protéines.)

Dimension des gels de taille moyenne : 20 cm x 20 cm. Ces gels sont utilisés pour la purification sur gel d'ARN et pour les tests de retard sur gel.

Dimension des gels de grande taille : 34 cm x 20 cm. Ces gels sont utilisés pour les tests de digestion d'ISG20 (test de processivité, test de protection par la PABPC1).

Dimension des gels de grande largeur : 20 cm x 34 cm. Ces gels sont utilisés pour les tests de retard sur gel et le test de digestion avec les substrats ARN méthylés.

### Composition des tampons de purification

Purification de GST-cycline T1 :

	Lyse	Elution	Stockage
Tris-HCl pH 8	50 mM	50 mM	50 mM
NaCl	100 mM	250 mM	100 mM
Triton X-100	0,1%	0,1%	0,1%
Glycérol	10%	10%	10%
$\beta$ -mercaptoéthanol	2 mM	2 mM	2 mM
MgCl <sub>2</sub>	1 mM		1 mM
ZnSO <sub>4</sub>	1 $\mu$ M		1 $\mu$ M
Glutathion réduit		20 mM	

Purification de his-cycline T1 :

	Lyse	Lavage	Elution	Stockage
Tris-HCl pH 7,5	50 mM	50 mM	50 mM	
HEPES pH 7,5				50 mM
Imidazole	20 mM	20 mM	0,5 M	
NaCl	150 mM	0,5 M	150 mM	500 mM
$\beta$ -mercaptoéthanol	5 mM	5 mM	5 mM	5 mM
Glycérol	20%	20%	20%	10%

Purification de CDK9-his seule, en co-lyse avec his-cycline (conditions native et dénaturante) :

	Lyse (CDK9 seule et co-lyse en condition native)	Lavage (CDK9 seule)	Elution (CDK9 seule et co-lyse en condition native)	Lyse (co-lyse en condition dénaturante)	Elution (co-lyse en condition dénaturante)
Tris-HCl pH 7,5	50 mM	50 mM	50 mM	50 mM	50 mM
Imidazole	20 mM	20 mM	0,5 M	20 mM	0,5 M
NaCl	150 mM	0,5 M	150 mM	1 M	0,5 M
$\beta$ -mercaptoéthanol	5 mM	5 mM	5 mM	2 mM	2 mM
Glycérol	20%	20%	20%	10%	10%
EDTA				0,1 mM	0,1 mM
Urée				2 M	

Purification de his-ISG20wt :

	Lyse	Lavage	Elution IMAC	SEC	Stockage
Tris-HCl pH 7.5	50 mM	50 mM	50 mM	50 mM	50 mM
NaCl	150 mM	500 mM	150 mM	1 M	50 mM
$\beta$ -mercaptoéthanol	5 mM	5 mM	5 mM	2 mM	2 mM
Glycérol	20%	20%	20%	20%	50%
Imidazole	20 mM	20 mM	500 mM		
EDTA				0,1 mM	

Purification de his-ISG20 D94A :

	Lyse	Lavage	Elution IMAC	Héparine A	Héparine B	Stockage
Tris-HCl pH 7.5	50 mM	50 mM	50 mM	50 mM	50 mM	50 mM
NaCl	300 mM	2 M	300 mM	150 mM	1 M	150 mM
$\beta$ -mercaptoéthanol	5 mM	5 mM	5 mM	5 mM	5 mM	2 mM
Glycérol	20%	20%	20%	10%	10%	10%
Imidazole	20 mM	20 mM	500 mM			
EDTA				0,1 mM	0,1 mM	0,1 mM
ATP		2 mM				

Purification de PABPC1-his :

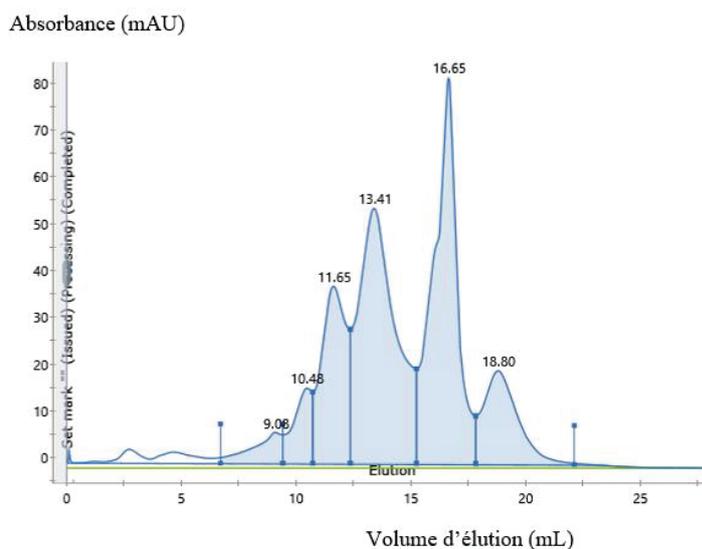
	Lyse	Elution IMAC	SEC	Stockage
Tris-HCl pH 7.5	50 mM	50 mM	50 mM	50 mM
NaCl	300 mM	300 mM	1 M	50 mM
$\beta$ -mercaptoéthanol	5 mM	2 mM	2 mM	2 mM
Glycérol	20%	10%	20%	20%
Imidazole	20 mM	500 mM		
EDTA		0,1 mM	0,1 mM	
MgCl <sub>2</sub>				

Purification de l'ARN chimérique minoTAR :

	Lyse	Elution
Tris-HCl pH 7,5	10 mM	10 mM
MgCl <sub>2</sub>	10 mM	10 mM
NaCl	100 mM	1 M

**Tableau 16.** Composition des tampons de purification. Tous les tampons de purification ont été filtrés à 0,22  $\mu$ m et conservés à 4°C.

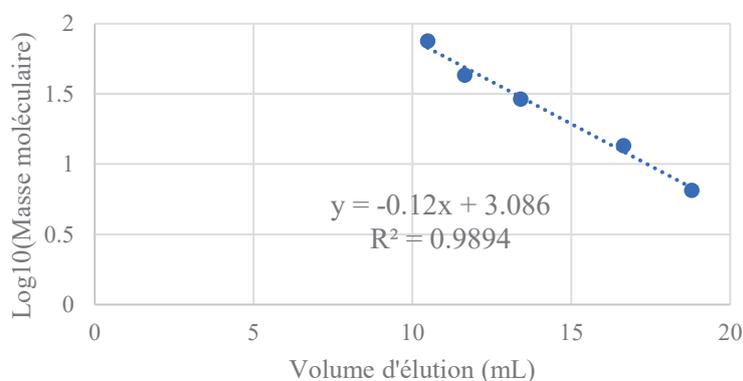
## Calibration de la colonne Superdex 75 Increase 10/300 GL



**Figure 92.** Chromatogramme représentant le profil d'étalonnage sur une colonne Superdex 75 Increase 10/300 GL avec les profils d'élution de plusieurs biomolécules : l'aprotinine, la ribonucléase A, l'anhydrase carbonique, l'ovalbumine, la conalbumine et un aldose. Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'élution en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et le tracé vert correspond au pourcentage de tampon d'élution.

Protéine	Masse moléculaire (kDa)	Log <sub>10</sub> (Masse moléculaire)	Volume d'élution (mL)
Aprotinine	6.5	0.81	18.8
Ribonucléase A	13.5	1.13	16.65
Anhydrase carbonique	29	1.46	13.41
Ovalbumine	43	1.63	11.65
Conalbumine	75	1.88	10.48
Aldose	198	2.30	9.08

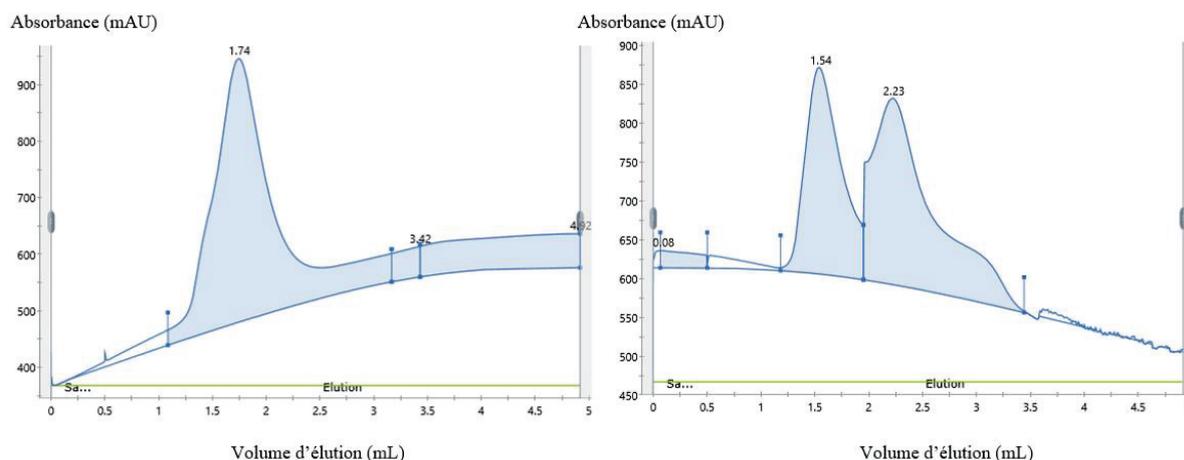
**Tableau 17.** Etalonnage de la colonne Superdex 75 Increase 10/300 GL.



**Figure 93.** Graphique représentant la calibration de la colonne Superdex 75 Increase 10/300 GL et mettant en évidence la relation entre le logarithme décimal de la masse moléculaire et le volume d'élution (mL). Les valeurs expérimentales sont représentées par des points bleus et la droite de tendance linéaire est en pointillé en bleu. L'équation de la droite de tendance et la valeur de  $R^2$  sont

affichées sur le graphique. La valeur expérimentale correspondant à l'aldose a été retirée du graphique car elle est hors-gamme.

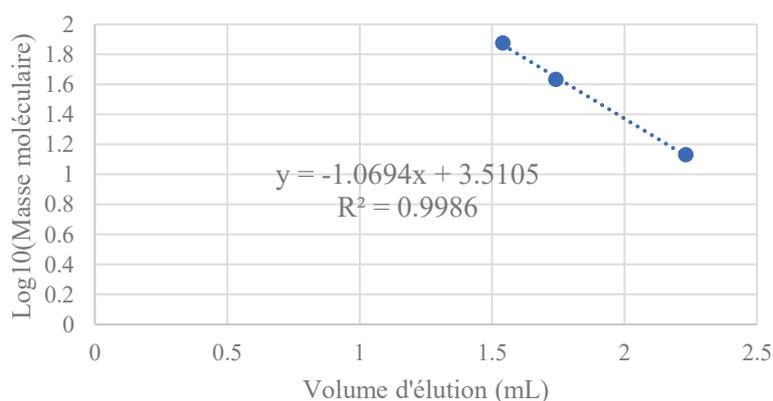
### Calibration de la colonne Superdex 75 Increase 5/150 GL



**Figure 94.** Chromatogrammes représentant le profil d'étalonnage sur une colonne Superdex 75 Increase 5/150 GL avec le profil d'élution de l'ovalbumine (à gauche), de la ribonucléase A et de la conalbumine (à droite). Représentation de l'absorbance en unité arbitraire (mAU) en fonction du volume d'élution en mL. La courbe en bleue correspond à la mesure de l'absorbance à 280 nm et le tracé vert correspond au pourcentage de tampon d'élution.

Protéine	Masse moléculaire (kDa)	Log <sub>10</sub> (Masse moléculaire)	Volume d'élution (mL)
Ribonucléase A	13.5	1.13	2.23
Ovalbumine	43	1.63	1.74
Conalbumine	75	1.88	1.54

**Tableau 18.** Etalonnage de la colonne Superdex 75 Increase 5/150 GL.



**Figure 95.** Graphique représentant la calibration de la colonne Superdex 75 Increase 5/150 GL et mettant en évidence la relation entre le logarithme décimal de la masse moléculaire et le volume d'élution (mL). Les valeurs expérimentales sont représentées par des points bleus et la droite de tendance linéaire est en pointillé en bleu. L'équation de la droite de tendance et la valeur de  $R^2$  sont affichées sur le graphique.

## Résumé

Durant ma thèse, j'ai travaillé sur l'étude moléculaire et structurale de complexes protéines:ARN impliqués dans le cycle de vie du virus de l'immunodéficience humaine de type-1 (VIH-1). Mon but était de mieux comprendre leurs mécanismes d'interaction, afin d'établir de nouvelles stratégies antirétrovirales sur le long terme.

**PARTIE I :** L'intégrase (IN) du VIH-1 est une protéine multifonctionnelle qui intervient à différentes étapes du cycle rétroviral. Plusieurs études ont montré l'importance du domaine C-terminal (CTD) de l'IN dans la transcription provirale et la morphogénèse du VIH-1. Il a été montré que le CTD liait fortement l'ARN TAR du génome viral, qui est impliqué avec la protéine Tat dans l'élongation de la transcription provirale. Au cours de ma thèse, j'ai utilisé des approches de biologie moléculaire, biochimie et biophysique, afin d'étudier les interactions entre le CTD, TAR et Tat. Ces travaux ont conduit à la publication d'un article montrant l'importance des changements de conformations induits par la liaison du CTD à TAR, car ils permettent d'améliorer la liaison de Tat. Mes résultats montrent également une interaction du CTD avec le complexe cycline T1:TAR. Ce dernier résultat suggère que l'IN pourrait intervenir dans la transcription provirale en absence de Tat.

**PARTIE II :** ISG20 (« *interferon-stimulated gene 20 kDa protein* ») est un facteur de restriction induit par l'interféron lors d'une réponse antivirale. Il s'agit d'une exonucléase 3'→5' de la famille des DEDD exonucléases. Les mécanismes d'action et de spécificité d'ISG20 vis-à-vis de ses substrats sont mal connus. J'ai mis en évidence *in vitro* que le résidu catalytique D94 d'ISG20 est critique pour la liaison à l'ARN et que l'enzyme est distributive. Une étude récente a montré que les 2'O-méthylations de l'ARN cellulaire permettaient de le protéger contre ISG20. Ce mode de protection est détourné par certains virus, comme le VIH-1. J'ai étudié et mis en évidence *in vitro* deux autres modes de protection des ARN contre l'activité d'ISG20 : l'addition d'une extrémité 3'-polyadénylée qui permet la liaison de certaines protéines et la structure de l'ARN. Ces résultats permettent de mieux expliquer la protection des ARN cellulaires et de certains ARN viraux à l'activité antivirale d'ISG20.

## Abstract

During my thesis, I worked on the molecular and structural study of protein:RNA complexes involved in the life cycle of human immunodeficiency virus type-1 (HIV-1). My aim was to better understand their interaction mechanisms, in order to establish new long-term antiretroviral strategies.

**PART I:** HIV-1 integrase (IN) is a multifunctional protein that is involved at different stages of the retroviral cycle. Several studies have shown the importance of the C-terminal domain (CTD) of IN in HIV-1 proviral transcription and morphogenesis. The CTD has been shown to strongly bind the TAR RNA of the viral genome which is involved with the Tat protein in the elongation step of proviral transcription. During my thesis, I used molecular biology, biochemistry, and biophysical approaches to study the interactions between CTD, TAR and Tat. This work led to the publication of a paper showing the importance of conformational changes induced by the binding of CTD to TAR, as they enhance the binding of Tat. My results also show an interaction of CTD with the cyclin T1:TAR complex. This latter result suggests that IN may be involved in proviral transcription in the absence of Tat.

**PART II:** ISG20 (interferon-stimulated gene 20 kDa protein) is a restriction factor induced by interferon during an antiviral response. It is a 3'→5' exonuclease of the DEDD exonuclease family. The mechanisms of action and specificity of ISG20 towards its substrates are poorly understood. I have shown *in vitro* that the catalytic residue D94 of ISG20 is critical for RNA binding and that the enzyme is distributive. A recent study showed that 2'O-methylations of cellular RNA protect it from ISG20. This mode of protection is hijacked by some viruses, such as HIV-1. I studied and demonstrated two other modes of RNA protection against ISG20 activity *in vitro*: the addition of a 3'-polyadenylated end that allows the binding of certain proteins and the structure of the RNA. These results help explain the protection of cellular RNA and some viral RNA to the antiviral activity of ISG20.