



HAL
open science

Structure de l'ARN au sein des ribonucléoprotéines des Influenzavirus A

Damien Ferhadian

► **To cite this version:**

Damien Ferhadian. Structure de l'ARN au sein des ribonucléoprotéines des Influenzavirus A. Virologie. Université de Strasbourg, 2018. Français. NNT : 2018STRAJ066 . tel-02918074

HAL Id: tel-02918074

<https://theses.hal.science/tel-02918074>

Submitted on 20 Aug 2020

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

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

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

Institut de Biologie Moléculaire et Cellulaire

UPR9002 : Architecture et Réactivité de l'ARN

THÈSE présentée par :

Damien FERHADIAN

soutenue le **28 Septembre 2018**

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

Sciences de la Vie et de la Santé

Aspects moléculaires et cellulaires de la biologie

**Structure de l'ARN au sein des
ribonucléoprotéines des Influenzavirus A**

THÈSE dirigée par :
Dr Roland Marquet

IBMC, UPR 9002, Strasbourg

RAPPORTEURS :
Dr Thibaut Crepin
Dr Nadia Naffakh

IBS, Grenoble
Institut Pasteur, Paris

EXAMINATEURS :
Dr Sylvie Friant

GMGM, UMR 7156, Strasbourg

Remerciements

Je tiens premièrement à remercier les membres de mon jury d'avoir accepté d'évaluer mon travail de thèse. Je remercie également le ministère de l'enseignement supérieur et de la recherche ainsi que l'agence nationale de la recherche pour le financement de mon doctorat.

Je remercie Philippe, Laurianne et Joanna, de la plateforme de spectrométrie de masse de l'institut, pour les analyses réalisées, ainsi que toutes les personnes qui m'ont aidé au cours de ma thèse. Merci à Laurence, qui a passé une semaine entière sur des données pour extraire des informations importantes, merci à Guillaume et à Philippe pour leur aide précieuse dans la purification de ma protéine, avec les AKTA, les problèmes d'induction et j'en passe.

J'aimerais remercier mon directeur de thèse, Roland, qui a toujours été là dans les moments de doute. Merci pour tes (très) nombreux conseils scientifiques, ton soutien aussi bien pendant les manips que pendant la rédaction de ce manuscrit, et pour les discussions très intéressantes que nous avons eues, aussi bien sur mon projet, les futurs projets que tu as en tête ou encore sur des sujets beaucoup plus larges.

Je remercie l'ensemble de l'équipe pour l'ambiance dans laquelle j'ai pu évoluer pendant ces trois ans. Merci à Valérie pour ses nombreux conseils et sa patience quand nous discutons chimie, merci à Anne pour sa bonne humeur de tous les jours (et de m'avoir accessoirement supporté pendant 3 ans au bureau juste derrière). Merci à Tanja, la caution sérieuse de cette équipe, aux nouveaux arrivants Erwan, Orian et Benjamin qui entretiennent la joie de vivre et le n'importe quoi ambiant du laboratoire. Je remercie également les anciens du laboratoire, Red, Catherine, Marie et Camille, que j'ai eu la chance de rencontrer. Merci à Serena, à son accent ~~espagnol~~ italien chantant, sa bonne humeur contagieuse et sa force de caractère, tu es absolument géniale ! Pour continuer dans l'équipe je remercie évidemment l'unique et l'inégalable Maud, alias petit Wedele, alias ToutanStrudel, avec qui j'ai adoré discuter de tout et de rien pendant 2 ans.

Un merci tout particulier au cycliste buveur de whisky complètement déjanté JC, toujours là pour aider, conseiller, écouter et picoler ! Merci pour tout ce que tu fais au laboratoire ainsi qu'en dehors lors de nos soirées, je ne me serai certainement pas aussi bien intégré dans l'équipe sans toi. Bien évidemment après JC je remercie son compère de toujours Philippe, et son voisin de bureau Guillaume (encore eux), qui, en plus de leur aide dans le domaine scientifique, ont accompagné mes pauses avec des anecdotes et des histoires toujours plus folles (dont la meilleure restera certainement la piscine olympique).

Je remercie les amis avec qui j'ai passé énormément de temps au laboratoire et en dehors. Merci Petitjean pour toutes nos discussions délirantes et totalement déjantées. Merci à Antoine pour sa bonne humeur, son amour de la science et de la musique et pour toutes les discussions que nous avons eues, sur tous les sujets possibles. Enfin, merci à Emma, sans qui ma thèse aurait été beaucoup plus terne. Ta joie de vivre, ta bonne humeur de tous les instants et (soyons honnête) ta maladresse naturelle ont été un allié de taille au cours de ces 3 années. Merci également pour ce voyage inoubliable jusqu'à San Francisco (RIP le poster).

Un grand merci collectif à tous les membres de l'unité que j'ai croisé de manière plus ou moins fréquente et notamment Alexis, Marta, Laura, Paula et Emma D. Merci également à tous les membres du coin café qui m'ont permis de bien commencer chaque journée.

Je remercie également ma famille, et plus particulièrement mes parents, qui ont toujours été derrière moi à tous les niveaux et à chaque instant. Sans eux je ne serai certainement pas où j'en suis aujourd'hui. Merci pour votre confiance, votre soutien et tout ce que vous m'avez apporté pendant ces années.

Et pour finir, je remercie la personne la plus importante dans ma vie depuis plus d'un an et demi, Elodie. Merci pour tout, ta bonne humeur, ton soutien, nos longues discussions quotidiennes, et tout ce que tu m'apportes jour après jour.

Sommaire

Abréviations

I. Introduction.....	1
A. Classification et généralités.....	1
B. Les IVA.....	2
1. <i>Nomenclature des IVA.....</i>	2
2. <i>Evolution des IVA.....</i>	3
3. <i>Structure de la particule virale.....</i>	5
4. <i>La RNPv.....</i>	6
1) <i>Le complexe polymérase.....</i>	8
2) <i>La protéine NP.....</i>	9
3) <i>L'ARN viral.....</i>	11
5. <i>Le cycle viral.....</i>	13
1) <i>Entrée dans la cellule.....</i>	14
2) <i>Import nucléaire.....</i>	16
3) <i>Transcription et export des ARNm.....</i>	16
4) <i>Import nucléaire et réplication.....</i>	18
5) <i>Export des RNPv.....</i>	19
6) <i>Transport cytoplasmique.....</i>	21
7) <i>Assemblage et bourgeonnement.....</i>	22
C. Génome segmenté et empaquetage.....	23
1. <i>L'empaquetage aléatoire chez les virus segmentés : cas de l'IBDV.....</i>	25
2. <i>L'empaquetage sélectif chez les virus segmentés : plusieurs exemples.....</i>	25
1) <i>Les Retroviridae, dont le Virus de l'Immunodéficience Humaine (VIH).....</i>	25
2) <i>Les Cystoviridae : l'exemple du phage $\phi 6$.....</i>	26
3) <i>Les Reoviridae : deux cas similaires.....</i>	26
3. <i>L'empaquetage des IVA, un mécanisme sélectif.....</i>	28
1) <i>Génétique et biochimie.....</i>	28
2) <i>Des signaux d'empaquetage identifiés.....</i>	28
3) <i>Organisation des RNPv dans la particule virale.....</i>	29
4) <i>Mise en évidence d'un réseau d'interaction entre ARNv.....</i>	29
5) <i>Le dernier modèle proposé.....</i>	30
6) <i>Dernières études sur l'empaquetage des Influenza virus.....</i>	30
D. Rôle de la structure des ARN chez les IVA.....	31
1. <i>Dans l'empaquetage.....</i>	31
2. <i>Rôle des motifs structuraux dans les ARN des IVA : article de revue.....</i>	32
E. Les protéines chaperonnes de l'ARN.....	33
1. <i>Généralités.....</i>	33
2. <i>Les protéines chaperonnes chez les virus à ARN.....</i>	34
F. Les techniques d'étude de la structure de l'ARN.....	35
1. <i>Les méthodes biophysiques.....</i>	35
2. <i>Les méthodes biochimiques.....</i>	37
II. Problématique & sujet de recherche.....	40

III. Matériel & méthodes.....42

A. Transcription <i>in vitro</i>	42
B. Production et purification de la protéine NP	43
1. <i>Induction de la protéine</i>	44
1) Premier protocole d'induction de la protéine NP (souche PR8)	44
2) Protocole final d'induction de la protéine NP (souche WSN)	45
2. <i>Purification de la protéine</i>	45
1) Premier protocole de purification de la protéine NP (souche PR8).....	45
2) Protocole final de purification de la protéine NP (souche WSN)	46
3) Suite du protocole après dialyse.....	47
C. Retard de migration sur gel	48
D. SHAPE <i>in vitro</i>	48
1. <i>Etape de modification</i>	48
2. <i>Etape de transcription inverse</i>	49
E. Marquage radioactif d'amorces	50
F. SHAPE <i>in viro</i>	51
1. <i>Production et purification des particules virales</i>	51
2. <i>Modification de l'ARN</i>	51
3. <i>Protocole d'icSHAPE</i>	52
4. <i>Protocole de SHAPE-MaP</i>	55

III. Résultats57

A. Production et purification de la protéine NP	577
1. <i>Souche PR8</i>	577
2. <i>Souche WSN</i>	622
B. Mise au point du protocole de SHAPE <i>in vitro</i>	68
C. Cartographie chimique <i>in viro</i>	72
1. <i>Mise au point de la modification</i>	72
2. <i>Mise en place du protocole icSHAPE</i>	77
3. <i>Mise en place du protocole de SHAPE-MaP</i>	79
D. Résultats SHAPE <i>in vitro</i>	82
1. <i>Comparaison des structures entre les souches PR8 et WSN</i>	82
2. <i>Activité chaperon de la protéine NP</i>	85
3. <i>Sites de fixation de la protéine NP</i>	93

IV. Conclusion & Discussion.....105

V. Perspectives112

Bibliographie

Annexes

Abréviations

10X, 5X, 2X, 1X
1M7

concentré dix, cinq, deux et une fois
1-méthyl-7-nitroisatoïc anhydride

A :

Å
A, T, G, C, U
AcNH₄
ADN
ADNc
ADP
Ai-N₃
AMV
ARN
ARNc
ARNdb
ARNm
ARNr
ARNss
ARNt
ARNv
AcNa

Angström
adénosine, thymine, guanosine, cytosine, uridine
acétate d'ammonium
acide désoxyribonucléique
ADN complémentaire
adénosine diphosphate
N-azidobutyl-anhydride isatoïque
Avian Myeloblastosis Virus
acide ribonucléique
ARN complémentaire
ARN double brin
ARN messenger
ARN ribosomique
ARN simple brin
ARN de transfert
ARN viral
acétate de sodium

B :

BTV
BzCN

Blue Tongue Virus
benzoyl cyanide

C :

°C
µCi
CaCl₂
CDC
CLiP
cm, nm
CMCT
CRM1
cryo-EM
CSE1L
CSP
CTD

degré Celsius
microcurie
chlorure de calcium
Centers for Disease Control
crosslinking immunoprecipitation
centimètre, nanomètre
1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate
chromosome maintenance region 1
cryo-microscopie électronique
chromosome segregaton 1-like protein
cold shock protein
domaine C-terminal

D :

ddGTP
DDX19
DIS
DLS
DMS
DMSO
dNTP
DO
DTT
EDTA

dideoxyguanosine-5'-triphosphate
DEAD-box RNA helicase 19
dimerization initiation site
dynamic light scattering
diméthylsulfate
diméthylsulfoxyde
désoxy-nucléotide triphosphate
densité optique
dithiothréitol
éthylènediaminetétraacétique

E :

eIF2
EMDB
EtOH

initiation factor 2 chez les eucaryotes
electron microscopy data bank
ethanol

F :

FISH

fluorescence in situ hybridization

G :

µg, ng
g

microgramme, nanogramme
unité de gravitation

GDP	guanosine diphosphate
GTP	guanosine triphosphate
<u>H :</u>	
6-His	tag 6 histidine
h	heure
HA	hémagglutinine
HCV	Virus de l'Hépatite C
HEPES	acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
HITS-CLiP	<i>high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation</i>
hnRNP	<i>heterogeneous nuclear ribonucleoproteins</i>
HRB	<i>HIV-1 rev binding protein</i>
<u>I :</u>	
IBDV	<i>Infectious Bursal Disease Virus</i>
icSHAPE	<i>in cell SHAPE</i>
ICTV	<i>international committee on taxonomy of viruses</i>
ICV	<i>irregularly coated vesicles</i>
IMAC	<i>immobilized metal affinity chromatography</i>
IPTG	isopropyl β -D-1-thiogalactopyranoside
IVA	Influenzavirus A
IVB	Influenzavirus B
IVC	Influenzavirus C
IVD	Influenzavirus D
<u>K :</u>	
KCl	chlorure de potassium
kDa	kilodalton
kJ	kiloJoule
<u>L :</u>	
LB	Luria Bertani
<u>M :</u>	
M, mM, μ M, nM	molaire, milimolaire, micromolaire, nanomolaire
M1	protéine de matrice 1
M2	protéine de matrice 2
MALDI-TOF	<i>matrix assisted laser desorption ionisation - time of flight</i>
MaP	<i>mutational profiling</i>
MCM	<i>minichromosome maintenance</i>
MgCl ₂	chlorure de magnésium
min	minute
ml, μ l	millilitre, microlitre
MnCl ₂	chlorure de manganèse
MOI	<i>multiplicity of infection</i>
MTOC	<i>microtubule organizing center</i>
<u>N :</u>	
NA	neuraminidase
NaCl	chlorure de sodium
NaOH	hydroxyde de sodium - soude
NCp7	nucléocapside du VIH-1
NEP	<i>nuclear export protein</i>
NES	<i>nuclear export signal</i>
Ni-NTA	<i>nickel-nitrilotriacetic acid</i>
NLS	<i>nuclear localization signal</i>
NMIA	N-méthylisatoïc anhydride
NP	nucléoprotéine
NS1	protéine non structurale 1
NS2	protéine non structurale 2
NXF1	<i>nuclear RNA export factor 1</i>
<u>O :</u>	
OMS	organisation mondiale de la santé

		<u>P:</u>
p/v	poids/volume	
PA	protéine acide	
PARS	<i>parallel analysis of RNA structure</i>	
pb	paire de base	
PB1	protéine basique 1	
PB2	protéine basique 2	
PBS	<i>primer binding site</i>	
PCR	<i>polymerase chain reaction</i>	
PEG8000	polyéthylène glycol 8000	
PFU	<i>plaque forming unit</i>	
pH	potentiel hydrogène	
PNK	polynucléotide kinase	
		<u>Q:</u>
qPCR	PCR quantitative	
		<u>R:</u>
Rab	<i>ras-related in brain</i>	
RanBP5	<i>ran-binding Protein 5</i>	
RanGAP	<i>ras-related nuclear GTPase-activating protein</i>	
RCC1	<i>regulator of chromosome condensaton</i>	
RE	réticulum endoplasmique	
RMN	résonance magnétique nucléaire	
RNPv	ribonucléoprotéine virale	
RRE	<i>rev response element</i>	
		<u>S:</u>
SDS	sodium dodécyl sulfate	
SDS-PAGE	<i>sodium dodecyl sulfate–polyacrylamide gel electrophoresis</i>	
SHAPE	<i>selective 2'-hydroxyl acylation analyzed by primer extension</i>	
SHAPE-MaP	<i>SHAPE-Mutational profiling</i>	
SL1	<i>stem loop 1</i>	
SPLASH	<i>sequencing of psoralen crosslinked, ligated, and selected hybrids</i>	
SSIV	superscript IV	
STAU1	<i>staufen double-stranded RNA binding protein 1</i>	
		<u>T:</u>
TAR	<i>trans-activation response element</i>	
Tat	<i>trans-activator of transcription</i>	
tat-SF1	<i>tat-specific factor 1</i>	
TB	tris borate	
TBE	tris borate EDTA	
TBM	tris borate MgCl ₂	
Tris-HCl	tris hydrochloride	
		<u>U:</u>
5'UTR	<i>5' untranslated region</i>	
U	unité	
U12	séquence unique de 12 nucléotides en 3' de l'ARNv	
U13	séquence unique de 13 nucléotides en 5' de l'ARNv	
UAP56	<i>ATP-dependant RNA helicase 56</i>	
UV	ultraviolet	
		<u>V:</u>
V	Volt	
v/v	volume/volume	
Vif	<i>viral infectivity factor</i>	
VIH	Virus de l'Immunodéficience Humaine	
VLP	<i>virus-like particle</i>	
vol	volume	
		<u>Y:</u>
YB-1	<i>Y-box binding protein 1</i>	

INTRODUCTION

Introduction

A. Classification et généralités

Les virus appartenant à la famille des *Orthomyxoviridae* (**Figure 1**), définie par l'ICTV (*International Committee on Taxonomy of Viruses*), présentent un génome segmenté, composé de 6 à 8 molécules d'ARN simple brin (ARNss) de polarité négative appelés ARN viraux (ARNv). Les membres de cette famille sont en mesure d'infecter une variété d'hôtes, depuis la tique jusqu'aux mammifères en passant par les oiseaux ou encore les saumons. Depuis 2016, sept genres font partie de cette famille avec la caractérisation d'un nouveau genre d'Influenzavirus [91].

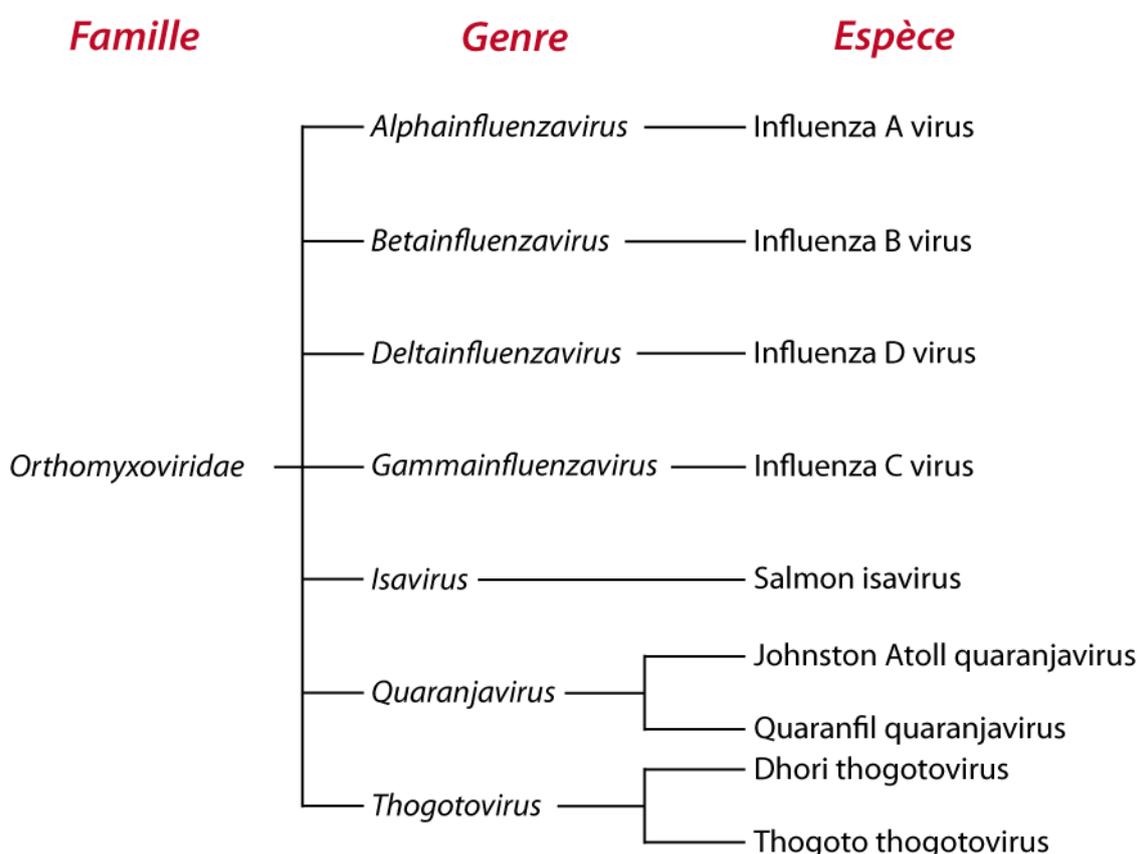


Figure 1 : Classification de la famille des *Orthomyxoviridae*.

Les différents genres sont les *Alphainfluenzavirus* (ou IVA), les *Betainfluenzavirus* (IVB) et les *Gammainfluenzavirus* (IVC), capables d'infecter les humains, les *Deltainfluenzavirus* (IVD), pouvant infecter le bétail et les porcs, les *Isavirus* infectant les saumons, les *Togotovirus*, des arbovirus capables d'infecter les insectes tels que les

moustiques et les tiques, et les *Quaranjavirus* infectant majoritairement les oiseaux et les arthropodes.

Les virus les plus étudiés de cette famille sont les Influenzavirus, appelés plus communément virus de la grippe et classés dans les différents genres en fonction de leurs protéines NP et M1, décrites ultérieurement, et plus précisément de leur propriété antigéniques [245]. D'après les CDC (*Centers for Disease Control*), les IVD ne sont pas connus pour infecter l'Homme, tandis que les IVC, retrouvés majoritairement chez les humains, ne causent que de légères maladies respiratoires et ne causent pas d'épidémies (pour revue sur les IVC voir [172]). Tout comme pour les IVC, les IVB ne semblent pas présenter de réservoir animal bien qu'un isolat récupéré sur phoque ait été rapporté, mais ce dernier ne montrait pas d'adaptation à son hôte. Les IVB ne présentent pas de potentiel pandémique, mais ils provoquent régulièrement des épidémies saisonnières importantes et sont donc aujourd'hui un composant important des vaccins (pour revue sur les IVB voir [110]). Le dernier genre Influenzavirus, *Alphainfluenzavirus*, est le plus étudié car il est à l'origine à la fois d'épidémies de grippe saisonnière et de pandémies et panzooties, présentant une mortalité et une morbidité très importantes. Les IVA ont pour réservoirs naturels les oiseaux aquatiques, mais ils sont en capacité d'infecter de nombreux mammifères, notamment le porc et l'Homme. En France, l'institut Santé Publique France estime à 13 000 le nombre de décès liés à la grippe au cours de la saison grippale 2017-2018 [20], et d'après l'OMS (Organisation Mondiale de la Santé), il y a chaque année dans le monde 3 à 5 millions de cas entraînant 250 000 à 500 000 décès [183]. Ces virus restent donc, malgré les nombreux moyens de lutte mis en place, un problème de santé publique majeur.

Le reste de ce manuscrit s'intéressera majoritairement aux IVA, sur lesquels mon travail de thèse s'est focalisé.

B. Les IVA

1. Nomenclature des IVA

Les IVA sont classés en différents sous-types en fonction de la nature de leurs glycoprotéines de surface, antigènes majeurs, l'hémagglutinine (HA) et la neuraminidase (NA). Ces sous-types sont notés sous la forme HXNX où H représente

un des 18 sous-types de la protéine HA (H1 à H18) et N un des 11 sous-types de la protéine NA (N1 à N11). Toutes les combinaisons possibles n'ont pas été détectées, et certains sous-types ne sont présents que chez certains organismes. Les virus H17-H18 et N10-N11 n'ont par exemple été observés que chez les chauves-souris [246,247]. La majorité des autres sous-types est retrouvées chez les oiseaux, tandis qu'à l'inverse, seuls les sous-types H1, H2, H3, N1 et N2 ont été observés chez l'Homme. Les pandémies jusqu'alors observées n'ont été causées que par des virus H1N1 (exemple de la grippe de 1918, aussi appelée grippe espagnole), H2N2 (exemple de la grippe asiatique de 1957) et H3N2 (exemple de la grippe de Hong Kong de 1968).

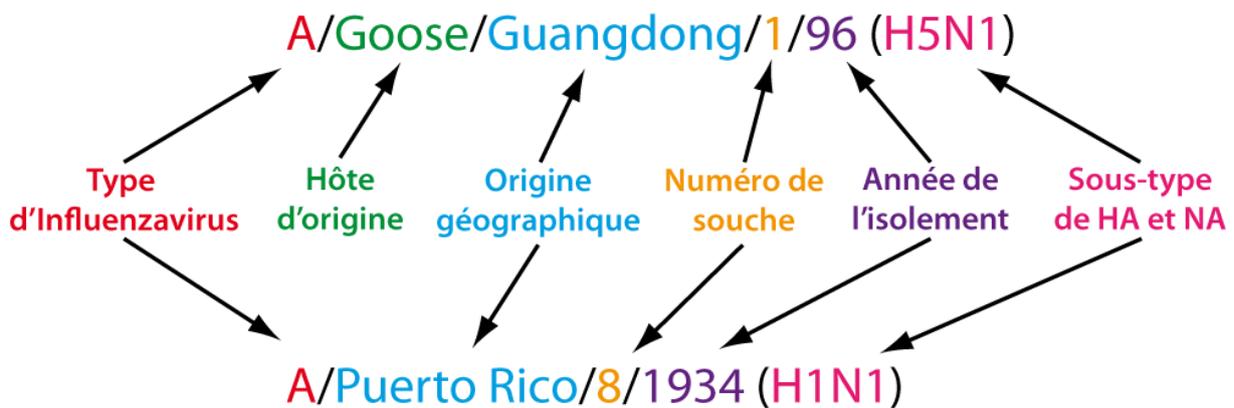


Figure 2 : Nomenclature des IVA définie par l’OMS avec l’exemple d’un virus humain (bas) et aviaire (haut).

Depuis 1980, l’OMS a établi une nomenclature très précise pour l’annotation de chaque souche d’Influenzavirus [268]. Cette annotation (**Figure 2**) contient le type viral (A, B ou C), l’hôte d’origine dans lequel la souche a été isolée (sauf si isolée chez l’Homme), l’origine géographique, le numéro de la souche, l’année d’isolement de la souche et, seulement pour les IVA, le sous-type viral sous forme HXNX.

2. Evolution des IVA

Deux mécanismes moléculaires sont à l’origine de l’évolution des IVA, la dérive antigénique (*antigenic drift*) et la cassure antigénique (*antigenic shift*), leur permettant d’échapper au système immunitaire de l’hôte.

La dérive antigénique est intrinsèquement liée à la faible fidélité de l’ARN polymérase virale qui ne possède pas d’activité de relecture, contrairement à certaines polymérases cellulaires. Cette incapacité à corriger les erreurs introduites au cours de la réplication virale se caractérise par un taux de mutations de l’ordre de 10^{-3} à 10^{-4}

mutations par site et par an [35,87]. En moyenne, cela correspond à l'introduction d'une substitution dans l'ensemble du génome viral par cycle réplcatif [55]. Si l'on s'intéresse à l'effet de ces substitutions sur les deux antigènes majeurs des IVA, les protéines de surface HA et NA, elles sont majoritairement neutres et ne modifient pas leurs propriétés antigéniques. Cependant, certaines mutations peuvent induire un changement de conformation des domaines antigéniques de ces protéines, menant à une perte de reconnaissance de ces dernières par les anticorps de l'hôte et donc à l'échappement du virus au système immunitaire [27]. Cette dérive antigénique explique en partie pourquoi les vaccins peuvent se montrer inefficaces d'une année sur l'autre alors que les mêmes souches circulent [244].

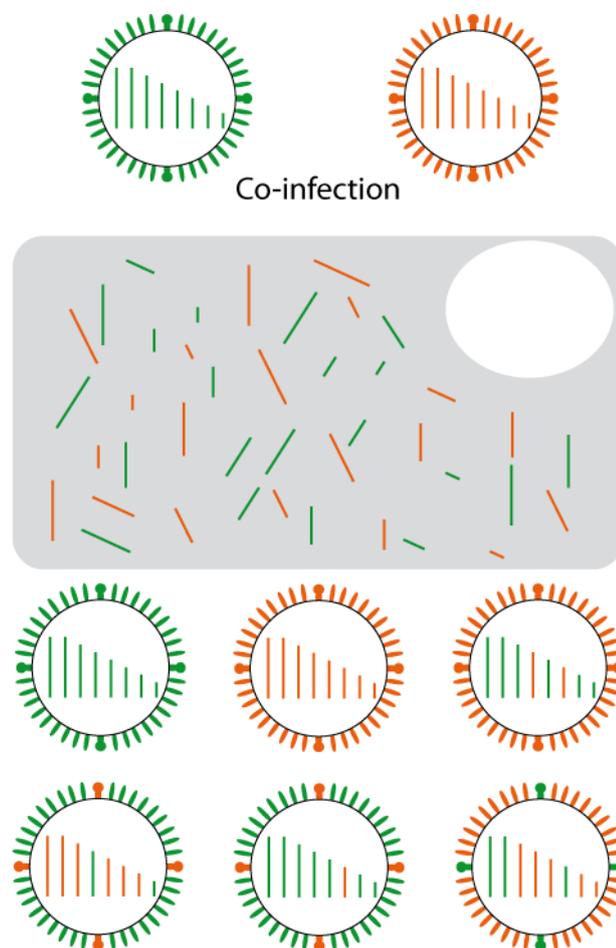


Figure 3 : Principe de réassortiment génétique chez les IVA. La co-infection d'une même cellule par deux virus différents permet la production de virus hybrides.

La cassure antigénique (**Figure 3**) est un mécanisme radicalement différent, car elle permet la formation de particules virales qui sont très différentes de celles qui sont à l'origine de l'infection. Ce mécanisme se déroule lors de l'infection d'une même cellule par deux virus différents. Les particules virales nouvellement produites peuvent

alors contenir un génome hybride comprenant des segments appartenant aux deux virus parentaux. Ce mécanisme de réassortiment génétique permet au virus d'évoluer rapidement (pour revue voir [243]), et est même à l'origine des pandémies grippales (pour revue voir [85]). En effet, quand des segments HA ou NA d'origine porcine ou aviaire sont empaquetés par des virus capables d'infecter l'Homme, la diffusion du virus est beaucoup plus rapide et globale car la population humaine est immunologiquement naïve vis-à-vis de ces antigènes.

3. Structure de la particule virale

Les IVA sont pléomorphes, avec des particules pouvant être sphériques, avec un diamètre moyen de 100 nm [70], ou filamenteuses, avec un diamètre moyen de 100 nm et une longueur pouvant atteindre 20 μm [23,171]. L'architecture sphérique de la particule virale semble être associée aux souches cultivées en laboratoire tandis que les formes allongées seraient fréquemment retrouvées dans les isolats prélevés sur les patients [40,118].

La particule virale comprend une enveloppe lipidique d'origine cellulaire acquise lors du processus de bourgeonnement marquant la fin du cycle viral. Les glycoprotéines virales HA et NA sont ancrées dans cette enveloppe respectivement sous forme de trimères et de tétramères (**Figure 4**). Les protéines HA sont le plus représentées à la surface de la particule virale, mais le rapport avec les protéines NA diffère selon la souche étudiée, et généralement les protéines HA sont 5 à 8 fois plus abondantes [78,89]. Chaque sous-unité du trimère de HA est composée de deux polypeptides résultant d'un clivage du précurseur HA₀ traduit au cours du cycle viral, et responsable du mécanisme de fusion membranaire nécessaire pour l'entrée du virus dans la cellule (pour revue sur HA voir [72]). Le rôle de NA est de cliver les acides sialiques présents à la surface des glycoprotéines de la membrane plasmique et probablement à la surface des virions, empêchant ainsi la réinfection de la cellule par les virions néoformés et l'agrégation de ces derniers (pour revue sur NA voir [72]). La protéine virale M2 (protéine de matrice 2), troisième composant de cette enveloppe virale [131] est un canal ionique qui permet l'entrée d'ions à l'intérieur de la particule virale (pour revue sur M2 voir [196]).

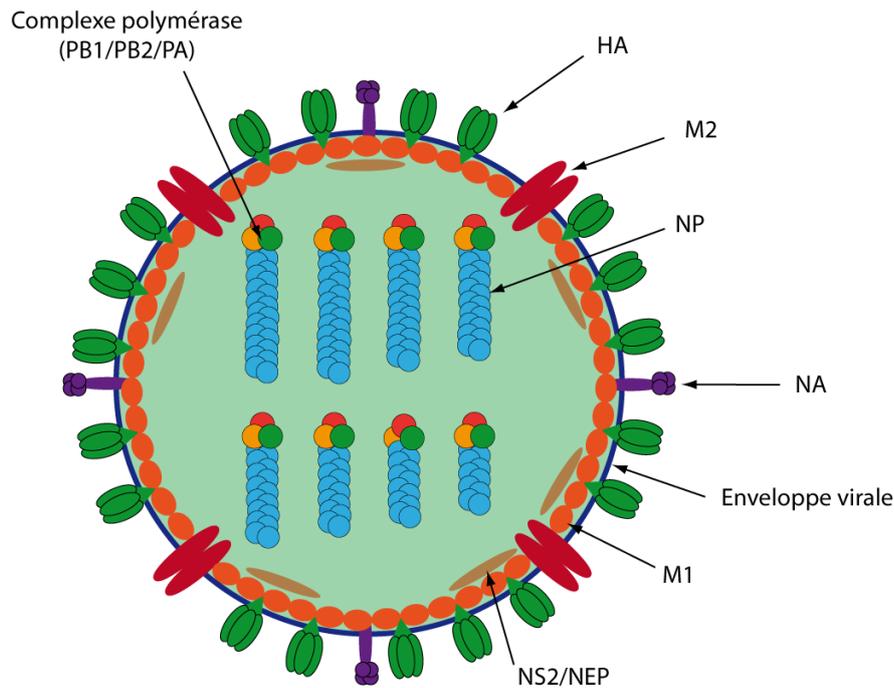


Figure 4 : Structure schématique de la particule virale des IVA.

Sous l'enveloppe virale se trouve la protéine de matrice 1 (M1), formant des interactions avec HA, NA et M2 [4]. Cette protéine possède trois domaines importants pour sa localisation, deux d'entre eux sont des NLS (*nuclear localization signal*) et permettent son import dans le noyau [49,279] tandis que le dernier est un NES (*nuclear export signal*) nécessaire à son export dans le cytoplasme [24]. Les rôles de cette protéine dans le cycle viral des IVA seront discutés ultérieurement. La protéine NS2 (protéine non structurale 2), également appelée NEP (*nuclear export protein*), est aussi retrouvée dans les particules virales [211] et serait associée à la protéine M1 [277]. Cette protéine possède un NES [108,185] dont la fonction sera discutée ultérieurement (pour revue sur NS2/NEP voir [190]). Au centre de la particule se trouve le génome des IVA sous la forme de huit ribonucléoprotéines virales (RNPv).

4. La RNPv

La RNPv (**Figure 5**) représente l'unité indépendante de réplication et de transcription des virus Influenza. Les IVA forment huit RNPv de tailles différentes en fonction de l'ARN qu'elles contiennent, chacun codant au moins une protéine virale. Elles sont composées d'un ARNv de polarité négative, associé à un complexe polymérase hétérotrimérique au niveau de ses extrémités 5' et 3' qui interagissent entre

elles, tandis que le reste de la molécule repose sur un squelette protéique formé par la nucléoprotéine (NP).

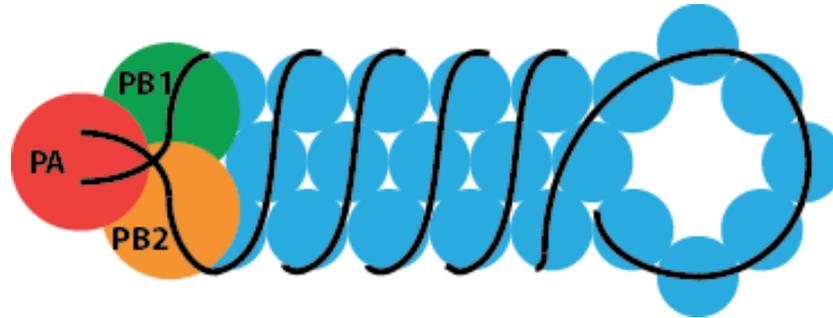


Figure 5 : Représentation de la RNPv des IVA. L'ARN viral (en noir) est recouvert de la protéine NP (en bleu) à l'exception des extrémités 5' et 3' associées au complexe polymérase hétérotrimérique.

En microscopie électronique les RNPv forment une hélice dont la longueur varie de 30 à 120 nm, avec un diamètre invariable de 15 nm [42,200]. Les deux extrémités de la RNPv sont bien distinctes, l'une contenant une boucle tandis que l'autre contient les extrémités 3' et 5' de l'ARNv [42,99] auxquelles est lié le complexe polymérase hétérotrimérique [66]. Les premières études stœchiométriques en microscopie électronique ont estimé, en se basant sur la taille des RNPv et leur composition relative en ARN et NP, qu'une protéine est retrouvée en moyenne tous les vingt nucléotides [42]. Cette estimation a été confirmée à nouveau à l'aide de la microscopie électronique [186] et par spectrométrie de masse [101]. Ces trois études dénombrent les protéines NP et calculent, en fonction de la taille de l'ARN étudié, le nombre de nucléotides par NP. Plus récemment, une étude a montré que la fixation d'une protéine NP sur l'ARNv est en moyenne restreinte à 12 nucléotides et que chaque site de fixation de la NP est espacé d'environ 25 nucléotides [272]. Des études de microscopie électronique plus récentes ont permis de construire de nouveaux modèles 3D du squelette protéique de la RNPv mais ne permettent pas de visualiser l'ARN au sein de ces structures [7,163]. Les deux modèles présentent une structure en double hélice avec des interactions NP-NP intra et inter-brins. Cependant, un des modèles montre une hélice droite contenant en moyenne cinq NP par tour d'hélice et par brin avec un angle de rotation entre chaque protéine de $73,9^\circ$ [163] tandis que le second modèle (**Figure 6**) représente une hélice gauche contenant six protéines par tour d'hélice et par brin avec un angle de rotation entre chaque NP compris entre -57 et -64° [7].

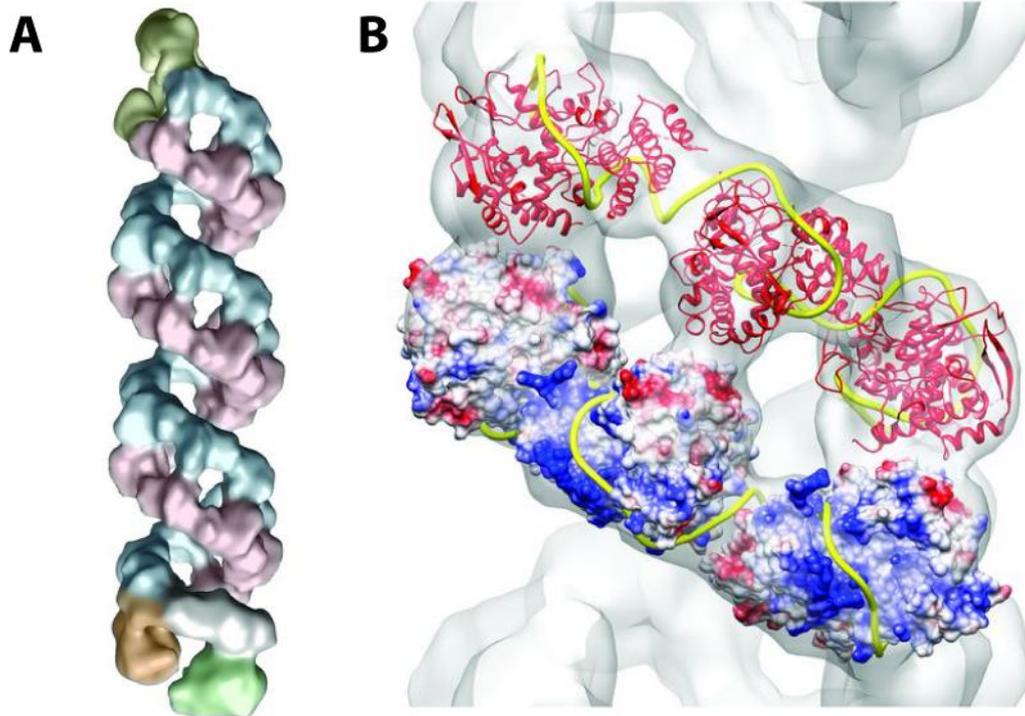


Figure 6 : Modèle de la structure de la RNPv (adapté de [7]). (A) Reconstruction complète du segment 8 avec le complexe polymérase (en bas) et boucle terminale (en haut). (B) Modèle de fixation de l'ARN (en jaune) au sein de la structure hélicoïdale de la RNPv. Un des brins de NP est présenté selon les potentiels de surface (rouge pour négatif et bleu pour positif) et le second sous la représentation en ruban.

1) Le complexe polymérase

Le complexe polymérase des IVA est composé de trois protéines différentes, PB1 (protéine basique 1), PB2 (protéine basique 2) et PA (protéine acide), chacune jouant un rôle bien défini lors des étapes de transcription et réplication du génome viral qui seront décrites ultérieurement. La structure de ce complexe, associé aux extrémités 5' et 3' de l'ARN viral formant le promoteur viral, est décrite dans la revue au §I.,D.,2. La structure du complexe seul, de 250 kDa, a été élucidée récemment pour un IVA de chauve-souris en association avec le promoteur de l'ARNv [194], un IVB humain en association avec le promoteur de l'ARNc [194] et un IVC humain en absence d'ARN [93]. La structure du complexe polymérase d'un IVA humain a également été élucidée, cependant incomplète car sans la partie C-terminale de la sous-unité PB2 [29]. Ces structures montrent de nombreuses interactions entre les sous-unités le composant, notamment entre PB1, PA (domaine C-terminal) et PB2 (domaine N-terminal) pour former le cœur de la polymérase ainsi qu'entre PB2 et PA. Avant ces premières

structures complètes du complexe, seules les structures de certains domaines des sous-unités composant ce complexe avait été élucidées [52,86,92,237,241].

2) La protéine NP

La nucléoprotéine virale, d'une taille de 55 kDa, est capable d'oligomériser afin de former un squelette hélicoïdal qui reste stable en absence d'ARN [7,163,178,218,278]. Cette capacité à oligomériser dépend de l'état de phosphorylation de la protéine [36,102,249]. L'ARNv doit s'enrouler vraisemblablement autour de cette structure [150] laissant les bases accessibles aux solvants et avec une affinité de 20 nM [11,120]. Les structures à haute résolution de protéines NP de souches H1N1 [278] et H5N1 [178] ont été élucidées et montrent une organisation similaire avec deux domaines principaux (la tête et le corps) et une boucle terminale flexible (**Figure 7A**) qui, en s'insérant dans le corps d'une protéine NP adjacente, permet les interactions NP-NP [7,28,41,178,278]. La capacité de la NP à fixer l'ARN dépend de son extrémité N-terminale [3,122], et les structures cristallographiques mettent en évidence dans cette région un sillon reliant la tête et le corps de la protéine (**Figure 7B**) et présentant de nombreux résidus basiques qui sont importants pour la fixation de l'ARN [178,278]. Cependant, des études plus anciennes ont mis en évidence d'autres régions tout au long de la protéine importantes pour la fixation de l'ARN [61,159].

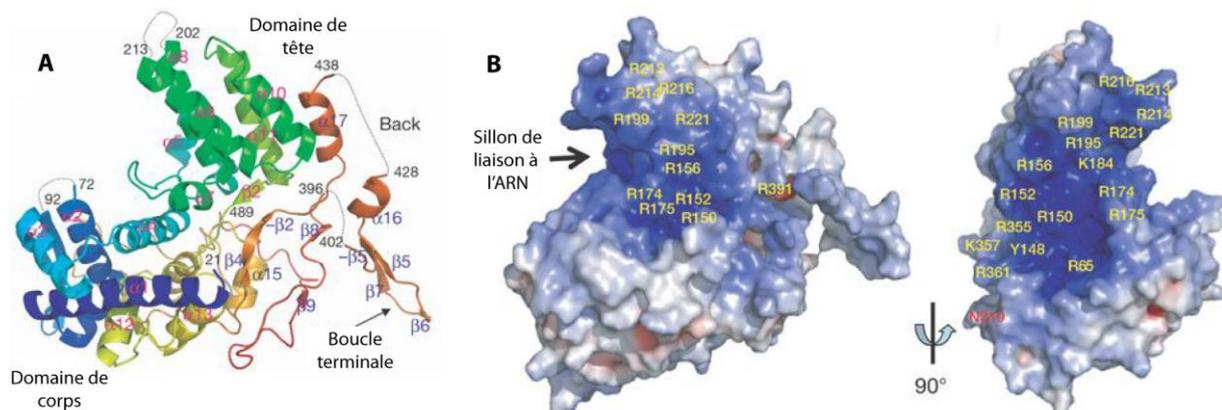


Figure 7 : Structure de la protéine NP de la souche WSN des IVA (adapté de [278]). (A) Structure cristallographique déterminée à une résolution de 3,2 Å. (B) Distribution des potentiels électrostatiques. En rouge et bleu sont représentés respectivement les potentiels négatifs et positifs.

Trois NLS ont été identifiés au sein de la NP des IVA. Le premier à avoir été identifié est un NLS non conventionnel, localisé dans la région N-terminale de la protéine [259]. Un second NLS bipartite a par la suite été découvert, entre les acides aminés 198 et 216

[262]. Cependant, les analyses structurales suggèrent que les deux régions qui composent ce NLS bipartite sont trop proches spatialement pour permettre une activité efficace. Le NLS non conventionnel est le signal majeur permettant l'import de la protéine NP dans le noyau [44,184]. Un troisième NLS, localisé entre les résidus 320 et 400, a été proposé mais son activité n'a pas encore été caractérisée [21]. De plus, la NP semble posséder un signal de localisation nucléaire présent entre les nucléotides 336-345, qui permettrait de maintenir la protéine dans le noyau pour la formation des nouvelles RNPv [50].

La fixation de la NP à de petits oligonucléotides et dans différentes conditions salines a été analysée par microscopie électronique [125] et montre que sans ARN, la NP est monomérique dans des conditions de faible salinité (50 mM NaCl) et trimérique dans des conditions de salinité plus élevées (150 et 300 mM NaCl), ce qui confirme des résultats précédemment publiés [242]. La fixation de la protéine monomérique à des oligonucléotides de six nucléotides et plus montre que la NP oligomérisse et forme principalement des trimères. La fixation de la protéine trimérique à des oligonucléotides de diverses tailles montre que la NP reste trimérique, sauf quand des oligonucléotides de douze nucléotides sont utilisés avec la protéine stockée à 150 mM NaCl. Dans ce cas, et si les oligonucléotides possèdent un fluorophore terminal, on observe par microscopie électronique des structures oligomériques qui ressemblent à des RNPv. Ces structures commencent à se former une heure après le début de l'incubation [125]. Des oligoribonucléotides d'une cinquantaine de nucléotides de long en présence de protéine NP monomérique ou trimérique montrent les mêmes résultats. En effet, la NP monomérique a tendance à oligomériser environ une heure après incubation avec l'ARN, tandis que la NP trimérique reste dans le même état d'oligomérisation [36]. L'utilisation d'un ARNv de pleine longueur pour la reconstitution d'un complexe ARN-NP *in vitro* montre que les ribonucléoprotéines formées se comportent similairement à des RNPv purifiées, ce qui indique qu'il est possible de reconstituer des RNPv *in vitro* [274].

Au sein des RNPv, la NP ne fixe pas seulement l'ARN mais interagit également avec le complexe polymérase hétérotrimérique, et notamment les sous-unités PB1 et PB2, cette dernière présentant même deux domaines d'interaction avec la NP [13,41,201]. Les résidus importants pour ces interactions sont localisés au niveau de la tête de la protéine et semblent également être liés à la capacité de la NP de promouvoir la

synthèse d'ARN par la polymérase virale [146]. Plus récemment, une étude a démontré une interaction entre la NP et la sous-unité PA du complexe polymérase, bien que ce ne soit pas la protéine PA entière qui ait été utilisée [252].

3) L'ARN viral

Le composant central de la RNPv est l'ARNv de polarité négative [187], ce qui indique que la première étape du cycle viral, une fois le génome libéré dans la cellule hôte, est le passage à un ARN de polarité positive. Les huit ARNv des IVA partagent tous la même organisation (**Figure 8**), avec des séquences communes aux extrémités 5' et 3', longues respectivement de 13 et 12 nucléotides, appelées Uni13 (U13) et Uni12 (U12) [226].

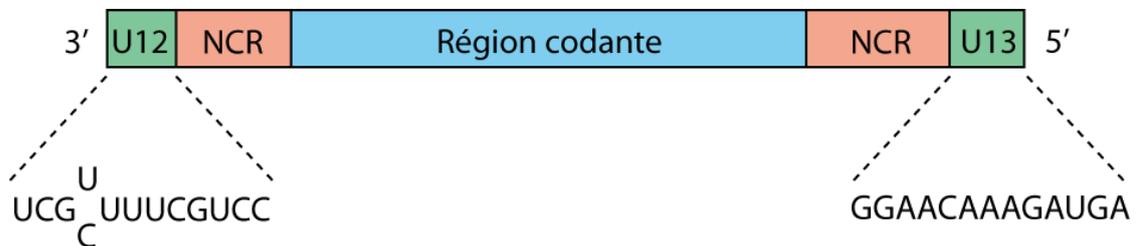


Figure 8 : Représentation de l'organisation commune des huit segments génomiques des IVA. La région codante, permettant l'expression d'une ou plusieurs protéines est flanquée d'une région non codante de longueur variable selon le segment. Les deux séquences uniques U12 et U13 sont localisées respectivement en 3' et 5' des ARNv.

Ces deux régions sont le point de fixation de l'ARN polymérase virale à l'ARN et forment le promoteur viral nécessaire pour les étapes de transcription et de réplication [17].

Au sein de la RNPv, ces deux régions sont en interactions, mais, bien que complémentaires, pas totalement appariées. La structure du promoteur a d'ailleurs longtemps été un point de débat (**Figure 9A**) dans la littérature (pour revue voir [65]) jusqu'à ce que la structure de la polymérase virale en complexe avec le promoteur soit élucidée (**Figure 9B**) et révèle la structure du promoteur au sein de la RNPv [194].

Directement accolées aux séquences U12 et U13 se trouvent des régions non codantes, dont la longueur diffère à chacune des extrémités de chacun des ARNv. Ces régions non codantes contiennent une partie des signaux nécessaires à l’empaquetage du virus en fin de cycle viral et qui seront discutés ultérieurement.

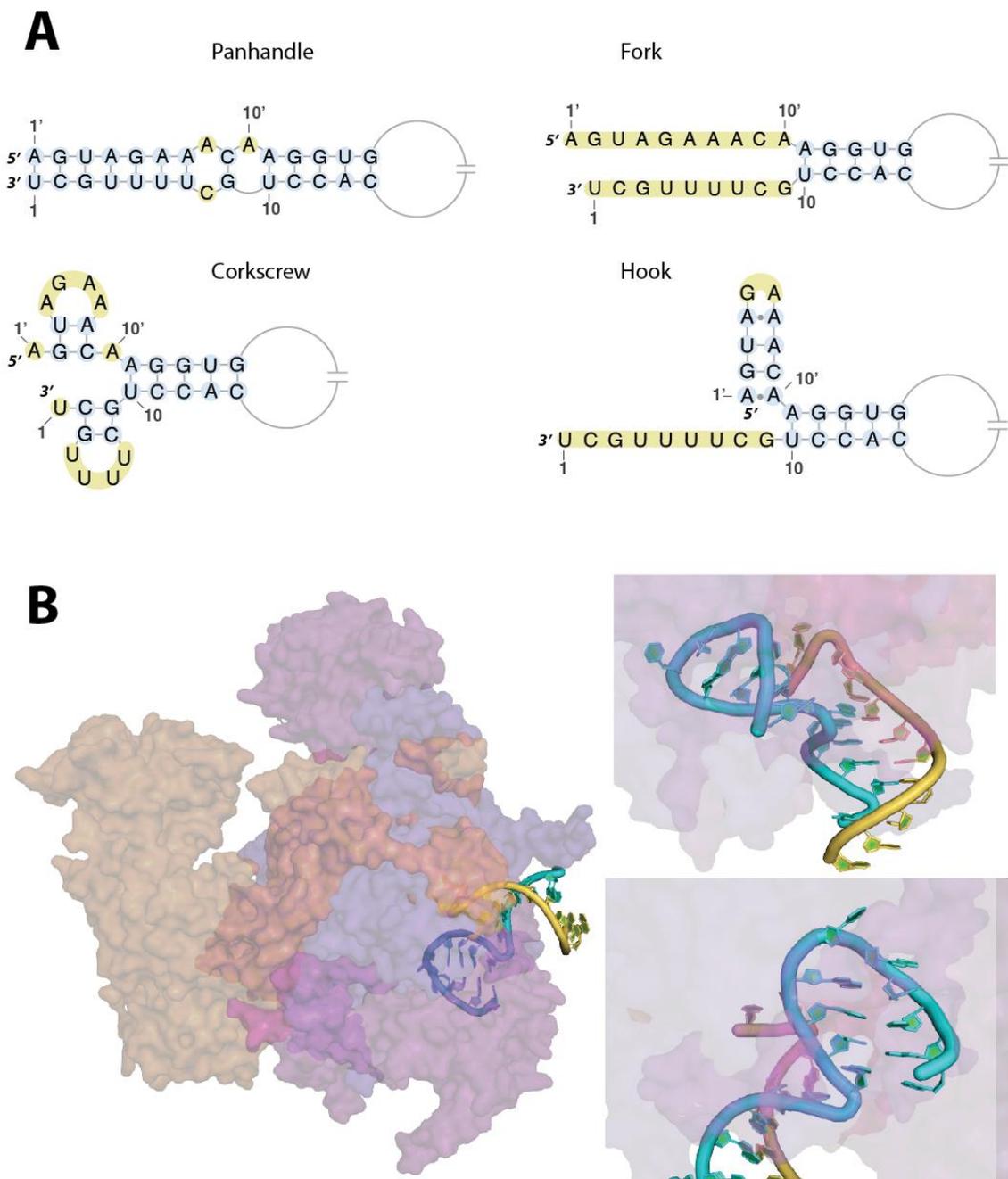


Figure 9 : (A) Représentation des différentes structures du promoteur des ARNv proposées au cours du temps. Les nucléotides non appariés sont surlignés en jaune. (B) Structure du promoteur des ARNv associé au complexe polymérase déterminée par cristallographie aux rayons X. Cette figure a été dessinée à l’aide du logiciel PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) en utilisant les coordonnées obtenues de [194]. Les extrémités 5’ et 3’ de l’ARN sont représentées respectivement en cyan et en jaune. A gauche est représentée la vue générale du complexe et à droite sont représentées deux vues rapprochées permettant d’observer la conformation « hook ».

Enfin, au centre de l'ARNv se trouve la région codante, qui permet la synthèse d'au moins une protéine virale (**Figure 10**). Dans le cas où plusieurs protéines sont codées par la région codante d'un même ARNv, différents mécanismes sont mis en place, tels que le *leaky scanning*, le décalage de cadre de lecture ou encore l'épissage alternatif (pour revue voir [77,251]).

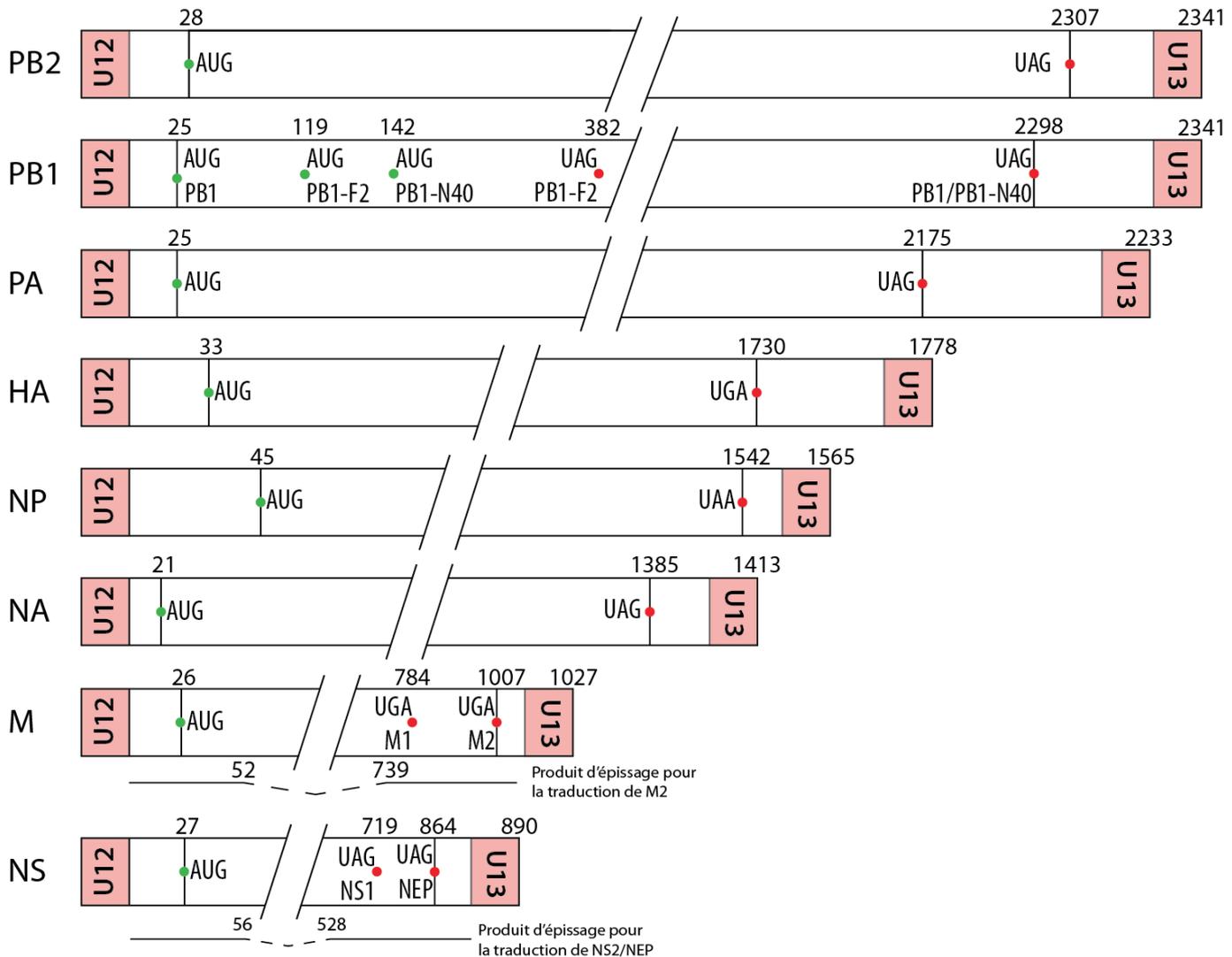


Figure 10 : Représentation schématique des huit ARNv des IVA. Les points verts représentent les codons d'initiation AUG des différentes protéines tandis que les points rouges représentent les codons STOP. Les ARNm épissés sont représentés sous les ARN M et NS. Les ARN négatifs sont numérotés de 3' vers 5'.

5. Le cycle viral

Les virus à ARN simple brin, de polarité positive ou négative, réalisent en général les étapes de transcription et de réplication de leur cycle viral dans le cytoplasme de la cellule infectée. Chez les IVA, ces étapes s'effectuent au sein du noyau des cellules

infectées (**Figure 11**). Une carte interactive, qui recense un grand nombre de facteurs cellulaires et viraux impliqués dans ce cycle et dans la réponse cellulaire associée à l'infection aux IVA, est disponible depuis quelques années [153].

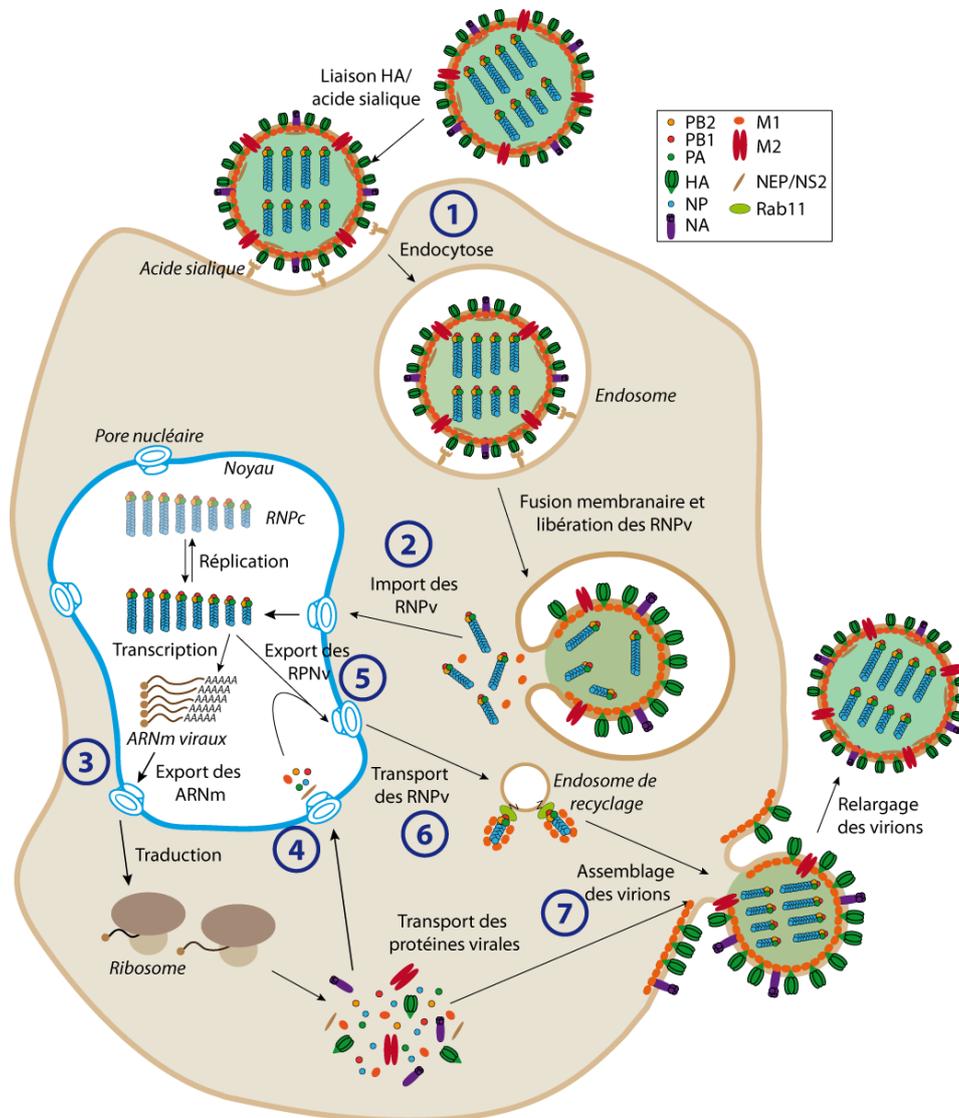


Figure 11 : Cycle viral des IVA. Les étapes numérotées sont décrites dans cette section.

1) Entrée dans la cellule

L'entrée dans la cellule des virus Influenza est réalisée suite à la reconnaissance et fixation de la protéine HA, présente à la surface de la particule virale sous forme d'homotrimer, aux acides sialiques présents sur les glycoprotéines et glycolipides ancrés dans la membrane plasmique [230]. L'affinité de reconnaissance entre la protéine HA et l'acide sialique dépend de la nature des liaisons ($\alpha(2,3)$ ou $\alpha(2,6)$) entre les acides sialiques et les glucides. Les liaisons $\alpha(2,6)$ sont préférentiellement

reconnues par les virus humains tandis que les liaisons $\alpha(2,3)$ sont préférentiellement reconnues par les virus équin et aviaires. Les virus porcins sont capables de reconnaître ces deux types de liaisons [230], expliquant ainsi la capacité des cochons à servir d'incubateur aux virus humains et aviaires, permettant la production de nouveaux virus pathogènes par réassortiment génétique. L'entrée dans la cellule suite à cette fixation peut être réalisée de plusieurs façons : historiquement, l'endocytose clathrine-dépendante est le premier mécanisme à avoir été décrit [152,191,282] avant qu'il soit démontré que les virus Influenza sont également capables d'utiliser la voie d'endocytose clathrine-indépendante [219,229]. Plus récemment, c'est la macropinocytose qui a été identifiée comme voie alternative utilisée par les IVA [256]. Ces différents mécanismes d'entrée dans la cellule mènent tous à une internalisation de la particule virale qui exploite le système de transport de la cellule [127]. Les endosomes précoces contenant le virus sont localisés à la périphérie de la cellule et transportés de façon actine-dépendante. Pendant la phase de maturation des endosomes précoces en endosomes tardifs, les vésicules contenant les particules virales sont transportées jusqu'à la région périnucléaire le long des microtubules. L'environnement acide des endosomes tardifs permet la fusion HA-dépendante de la membrane virale avec celle de l'endosome [47,143,266]. Le pH optimal pour la fusion des deux enveloppes est sous-type-dépendant, et il a été récemment démontré que pour un même sous-type les protéines HA d'isolats humains nécessitent un pH plus bas que les protéines HA aviaires [71]. L'environnement acide de l'endosome induit un changement conformationnel de la protéine HA, qui expose le peptide de fusion proche de la membrane de l'endosome [22,25,34]. Le peptide de fusion est alors inséré dans la membrane de l'endosome, rapprochant ainsi la membrane virale [57,248,263]. La fusion membranaire est mise en place par des changements conformationnels simultanés de plusieurs trimères de HA [147], et cette fusion permet la formation d'un pore au travers desquels les RNPv seront relâchées dans le cytoplasme. Pour cela, dans un premier temps les interactions entre la protéine M1 et les RNPv au sein de la particule virale sont déstabilisées. La protéine M2, en tant que canal ionique, crée un influx de protons depuis l'endosome après acidification de ce dernier [265], induisant une baisse de pH au sein de la particule virale et la rupture des interactions M1-RNPv, vraisemblablement suite à un changement conformationnel de M1 [287].

2) Import nucléaire

Une fois les RNPv libérées dans le cytosol, elles sont importées activement dans le noyau, lieu de la réplication virale, *via* la voie de l'importine α . Il a été montré que les RNPv seules, micro-injectées dans la cellule étaient capables de pénétrer dans le noyau [115] alors que l'ARN viral seul en était incapable, l'apport de NP étant nécessaire [184]. Bien que tous les composants protéiques des RNPv (PB1, PB2, PA et NP) possèdent un NLS, seul le NLS non-conventionnel de la NP décrit précédemment semble principalement entrer en jeu [44,184]. Les karyophérines α_1 , α_3 et α_5 sont les importines principales impliquées dans l'import des RNPv [160,184,241,259] ainsi que la protéine CSE1L (*chromosome segregation 1-like protein*), responsable de l'export des karyophérines dans le cytoplasme [123]. Les NLS des sous-unités du complexe polymérase seront utilisés plus tardivement dans le cycle viral, lors de l'import nucléaire des protéines après leur synthèse par la machinerie traductionnelle cellulaire.

3) Transcription et export des ARN messagers

Les RNPv importées dans le noyau après entrée du virus dans la cellule sont à la fois utilisées pour la transcription (**Figure 12A**) des ARN messagers (ARNm) et pour la synthèse des ARN complémentaires (ARNc), nécessaires à la réplication du génome viral (**Figure 12B**). La transcription des ARNm viraux est un processus indépendant de la synthèse de nouvelles protéines virales [145], suggérant que ce sont les complexes polymérase présents aux extrémités des RNPv qui en sont responsables. Le mécanisme de « *cap-snatching* », réalisé par le complexe polymérase viral, permet le vol de la coiffe des ARNm cellulaires à laquelle reste associés quelques nucléotides. Ces nucléotides seront utilisés comme amorce pour la synthèse des ARNm viraux [15,197,198]. L'ARN polymérase virale interagit avec le domaine C-terminal (CTD) de l'ARN polymérase II sous sa forme phosphorylée [63,151], ce qui lui permet de reconnaître la coiffe des ARNm cellulaires néoformés *via* sa sous-unité PB2 [86]. Le clivage 10 à 13 nucléotides en aval de la coiffe est réalisée par l'activité endonucléase de la sous-unité PA, [52,284]. L'amorce ainsi obtenue permet d'initier la transcription des ARNm viraux, réalisée par l'activité ARN polymérase ARN dépendante de la sous-unité PB1 du complexe polymérase [14,194,207]. L'étape de transcription se termine par un mécanisme de bégaiement de la polymérase virale sur une séquence oligoU qui permet l'ajout d'une queue polyA avant que la polymérase ne se décroche [202,213].

Certains transcrits sont épissés pour permettre l'expression des protéines virales NS2/NEP [129] et M2 [107,130]. Une troisième protéine issue de l'épissage de l'ARNm NS a été proposée mais n'est pas trouvée dans toutes les souches [223].

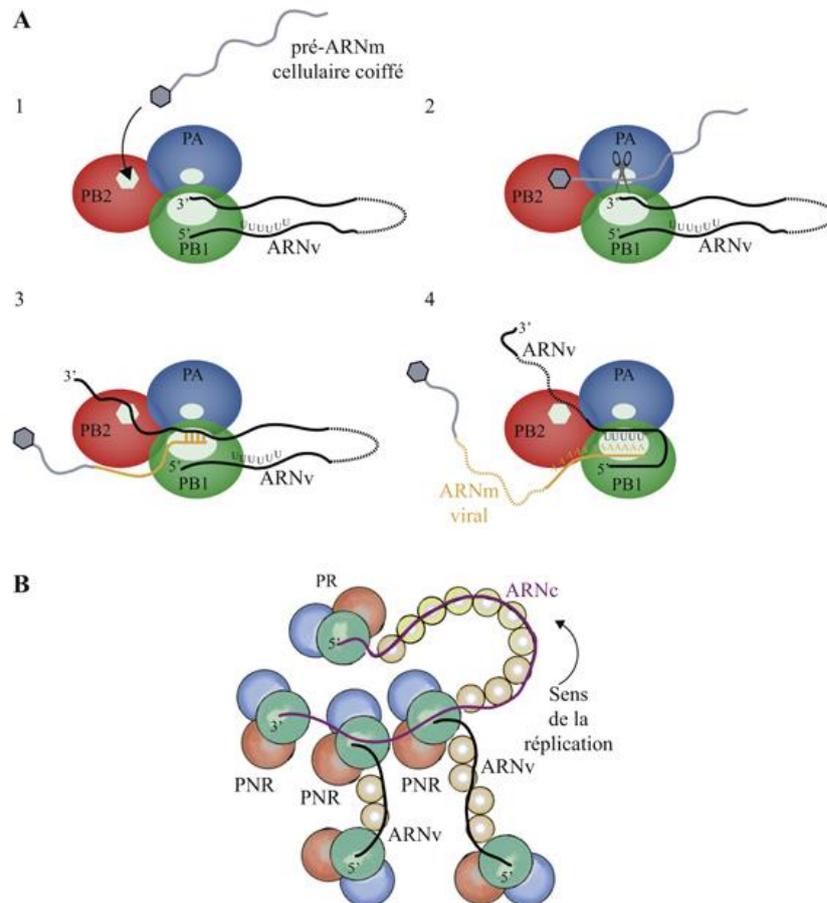


Figure 12 : Etapes de transcription et de répllication du génome des IVA (d'après [37]). **(A)** Transcription coiffe-dépendante des RNPv. (1) Reconnaissance, par la sous-unité PB2 du complexe polymérase viral, d'un pré-ARNm cellulaire qui (2) est clivé, par la sous-unité PA, 10-15 nucléotides en aval de la coiffe. (3) Ce court fragment coiffé sert d'amorce à la transcription des ARNm viraux qui (4) sont polyadénylés à la suite du bégaiement de la polymérase sur une séquence polyU. **(B)** Réplication des RNPv. L'ARNc (en violet), synthétisé à partir de l'ARNv, est associé à la NP et au complexe polymérase viral et est utilisé comme matrice pour la synthèse des ARNv (noir) qui formeront les RNPv après leur association avec le complexe polymérase et la protéine NP.

Les ARNm viraux sont ensuite exportés dans le cytoplasme afin de permettre la synthèse des protéines virales nécessaires à la poursuite du cycle viral. Alors que l'export nucléaire des ARNm cellulaires est couplé à l'épissage (pour revue voir [206,269]) et fait intervenir le facteur NXF1 (*nuclear RNA export factor 1*) qui forme un hétérodimère avec la protéine P15, la majorité des ARN viraux ne sont pas épissés et les virus Influenza ont donc dû évoluer afin d'exporter leur ARNm par un autre mécanisme. L'export des ARNm viraux dépend à la fois de l'interaction de la

polymérase virale avec le domaine CTD de l'ARN polymérase II cellulaire [6] et de l'activité de cette enzyme [253]. La protéine cellulaire DDX19 (*DEAD-box RNA helicase 19*) est également un facteur cellulaire essentiel pour l'export des ARNm viraux en se fixant, de façon non spécifique, sur le squelette sucre-phosphate de l'ARN après recrutement par le complexe polymérase [54]. La voie d'export NXF1 a été déterminée comme voie principale utilisée par les ARNm viraux [88,260] bien que plus récemment il a été montré que les ARNm codant les protéines du complexe polymérase (PA, PB1 et PB2) emprunteraient une voie non classique pour leur export [132]. Une fois dans le cytoplasme, les ARNm viraux coiffés et polyadénylés sont reconnus par la machinerie cellulaire et traduits comme les ARNm cellulaires.

4) Import nucléaire et réplication

Plusieurs protéines des IVA possèdent un ou plusieurs NLS et peuvent donc être importées dans le noyau après leur synthèse afin d'assurer la réplication du génome viral, la formation des RNPv et leur export. L'import de ces protéines peut se faire de différentes façons (voir revue [137]). Les protéines du complexe polymérase PB1 et PA forment un dimère et interagissent avec RanBP5 (*Ran-binding protein 5*) [51,67,104,239] tandis que NEP est importée indépendamment de la présence d'un NLS, par diffusion passive à travers le pore nucléaire [73].

La réplication se compose de deux phases : premièrement, un ARNc de polarité positive et correspondant à la copie exacte de l'ARNv, non coiffé et non polyadénylé, est synthétisé. Cet ARNc est ensuite utilisé pour la synthèse d'un nouvel ARNv qui formera une nouvelle RNPv. Contrairement à l'étape de transcription, la réplication du génome viral est réalisée sans amorce, grâce à la présence d'une extrémité 5' triphosphate chez les ARNc et ARNv [254]. De plus, cette étape de réplication peut être réalisée avec une polymérase soluble, apportée en *trans*, qui n'est donc pas celle fixée aux extrémités 3' et 5' des RNPv qui entrent dans le noyau [111,281]. L'ARNc synthétisé est recouvert de NP et fixé par la polymérase virale de la même manière que l'ARNv [281]. Des études structurales récentes ont permis de mettre en évidence des différences importantes entre l'initiation de la synthèse des ARNc et ARNv liées à la structure des promoteurs de ces deux molécules (pour revue voir [195]).

L'import des protéines du complexe polymérase et de la NP nouvellement synthétisées permet d'augmenter l'efficacité de la réplication en stabilisant les ARNc synthétisés et en empêchant leur dégradation par les nucléases cellulaires [255]. Cette

stabilisation des ARNc permettrait de réguler l'équilibre transcription/réplication, mécanisme encore mal connu. L'importance de certains facteurs viraux (NS1, NEP, NP etc...) ainsi que de facteurs cellulaires tels que MCM (*minichromosome maintenance*), UAP56 (*ATP-dependant RNA helicase 56*) ou tat-SF1 (*tat-specific factor 1*) a été proposée (pour revue voir [208]). Des petits ARN viraux (22 à 27 nucléotides) correspondant à l'extrémité 5' de chaque segment ont également été détectés et sont suspectés interagir avec le complexe polymérase afin d'assurer la transition entre transcription et réplication [193]. Cependant, le modèle selon lequel la transcription serait réalisée en *cis* par la polymérase fixée aux RNPv, alors que la réplication utiliserait en *trans* la polymérase nouvellement produite [111], est le plus populaire dans la littérature. Cependant, ce modèle a été remis en question par une étude suggérant que la polymérase apportée en *trans* n'aurait pour but que d'activer la polymérase présente au niveau des RNPv [281].

5) Export des RNPv

En plus de son rôle dans la régulation du passage de la transcription à la régulation, il semble que le facteur cellulaire UAP56 stimule la formation de trimères de NP pour permettre l'assemblage des RNPv [100]. Une fois formées dans le noyau, elles doivent être exportées dans le cytoplasme (**Figure 13**) et amenées à la membrane plasmique pour permettre leur empaquetage et la formation de nouvelles particules virales. L'utilisation de Leptomycin B, inhibiteur de la voie d'export CRM1 (*chromosome maintenance region 1*) a montré une accumulation de RNPv dans le noyau, indiquant que cette voie est utilisée par le virus pour l'export nucléaire des RNPv [62,142,261]. Dans cette voie d'export, la protéine CRM1 reconnaît les protéines cargos possédant un NES riche en lysine et s'associe avec Ran-GTP pour permettre le transport du noyau vers le cytoplasme. Une fois dans le cytosol, la protéine RanGAP (*Ras-related nuclear GTPase-activating protein*) hydrolyse la molécule de GTP (guanosine triphosphate), facilitant la dissociation du complexe et relâchant la protéine cargo. Ran-GDP est alors relocalisée dans le noyau, où elle est prise en charge par la protéine RCC1 (*regulator of chromosome condensation 1*) qui échange le GDP (guanosine diphosphate) contre du GTP.

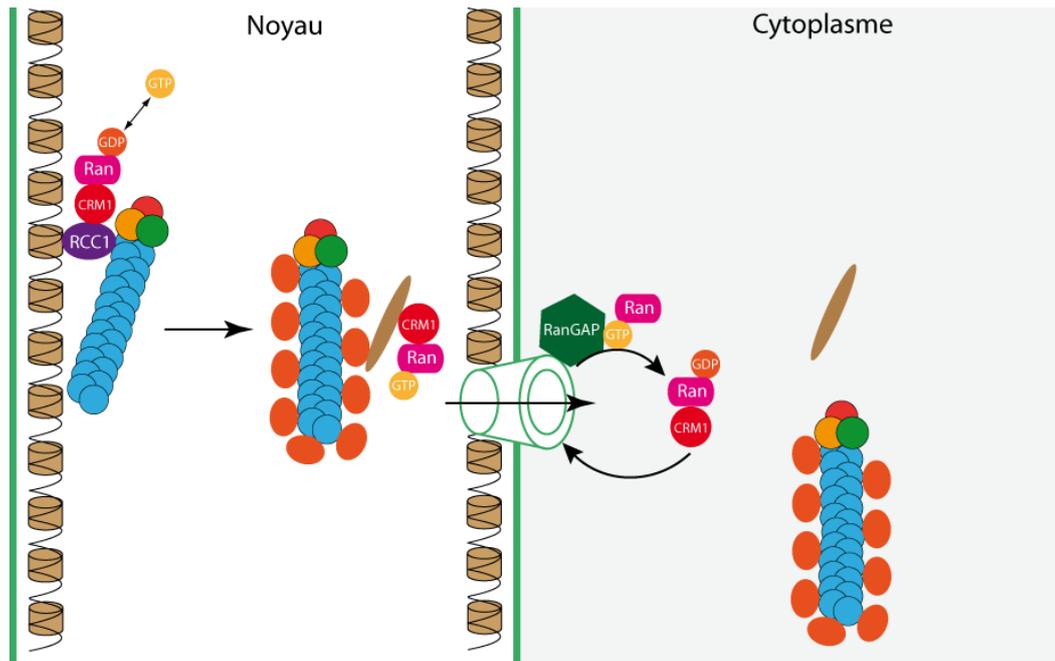


Figure 13 : Mécanisme d'export des RNPv. Les RNPv, en complexe avec le facteur cellulaire RCC1, sont associées à la chromatine dans le noyau. La protéine virale M1 (en orange) s'associe à la RNPv et permet l'interaction avec NS2/NEP (en marron) pour l'export à travers le pore nucléaire.

Plus récemment, il a été démontré que les RNPv sont fixées à la chromatine et forment un complexe avec RCC1 ainsi qu'avec CRM1 [30]. Les RNPv, seules, ne sont pas en capacité d'être exportées dans le cytoplasme, malgré la présence de plusieurs NES dans la NP [176,283], le fait qu'elle interagisse avec CRM1 *in vitro* [62] et sa capacité à être exportée dans le cytoplasme en absence d'autres protéines virales [176,267]. Le mécanisme d'export des RNPv dans le cytoplasme est dépendant de la présence de M1 et de NS2-NEP [19,149]. La protéine M1 doit être associée au RNPv pour permettre leur export [221], mais ce n'est pas elle, malgré la présence d'un NES, qui fait le lien directement avec CRM1 [24]. Plusieurs hypothèses peuvent être émises suite à ces résultats : M1 pourrait n'être qu'un intermédiaire pour connecter les RNPv à un facteur capable de lier CRM1 ou encore, M1 étant capable de fixer la chromatine [30,288], elle pourrait abolir l'interaction entre la chromatine et les RNPv. De plus, l'interaction entre les RNPv et M1 permettrait d'éviter à nouveau l'import nucléaire des RNPv exportées, en masquant le NLS de la protéine NP [267]. Contrairement à la protéine M1, la protéine NS2-NEP est capable de lier CRM1 [177] *via* deux motifs NES. Aucune interaction directe entre NS2-NEP et les RNPv n'a été caractérisée à ce jour chez les IAV [2,105,185,212] contrairement aux IBV [105], alors que des études ont montré que NS2-NEP et M1 peuvent interagir [2,228]. L'interaction entre M1 et les

RNPv ayant été précédemment démontrée, il semble que M1 sert d'intermédiaire entre NS2-NEP et les RNPv pour l'export nucléaire de ces dernières pas la voie CRM1. Cependant, un nouveau modèle a été proposé plus récemment, où NS2-NEP serait en mesure de lier PB1, mais les expériences n'ont pas été réalisées avec la protéine NS2-NEP entière [18]. Savoir si les RNPv sont exportées dans le cytoplasme individuellement ou en groupes reste une interrogation majeure pour ce mécanisme bien que des résultats suggèrent un export de plusieurs RNPv en même temps [38]. En outre, il a été montré que plusieurs RNPv pouvaient s'associer dans le cytoplasme en périphérie du noyau [128].

6) Transport cytoplasmique

Une fois dans le cytoplasme, les RNPv doivent atteindre le lieu de bourgeonnement des particules virales qui se situe au niveau de la membrane plasmique. Le transport cytoplasmique des RNPv se fait grâce au réseau de microtubules. Rapidement après leur export du noyau, elles s'accumulent au niveau du MTOC (*microtubule organizing center*) [5,165], proche du noyau puis colocalisent avec le réseau de microtubule en direction de la membrane plasmique [165,166]. De plus, les RNPv colocalisent avec la protéine Rab11 (*ras-related in brain*), marqueur des endosomes de recyclage et possédant une activité GTPase nécessaire au trafic des RNPv [5,8,38,59,166]. Cette relation entre Rab11 et les RNPv s'explique par la capacité du complexe polymérase à interagir avec Rab11 [166], et plus particulièrement la protéine PB2 [5]. Cette interaction spécifique entre les RNPv et Rab11 pourrait servir de plateforme qui concentrerait les RNPv et permettrait alors de mettre en place des interactions préalables à la formation du complexe des huit segments utilisé pour le mécanisme d'empaquetage sélectif (pour revue voir [79]). Cependant, bien que la protéine virale M1 soit importante pour l'export nucléaire des RNPv et empêche leur réimport, aucune étude n'a formellement démontré que le transport des RNPv au niveau des vésicules Rab11 était réalisé en présence de M1. Une étude récente a d'ailleurs montré par microscopie une co-localisation entre Rab11 et NP mais non M1 [26]. Par ailleurs, la SUMOylation de la protéine M1 semble jouer un rôle important dans les phases d'assemblage de la particule virale, son inhibition provoquant une accumulation des ARNv dans le cytoplasme [273]. De plus, l'absence de SUMOylation sur les régions de la protéine M1 permettant l'interaction avec les RNPv a pour effet de diminuer cette

capacité d'interaction [273]. Ensemble, ces résultats suggèrent un rôle de M1 dans le transport cytoplasmique des RNPv, mais vraisemblablement indirect.

D'autres protéines cellulaires semblent jouer un rôle dans le transport des RNPv jusqu'à la membrane plasmique. YB-1 (*Y-box binding protein 1*) est supposée faciliter l'accès aux vésicules Rab11 pour les RNPv [114], STAU1 (*staufen double-stranded RNA binding protein 1*) est associé aux RNPv mais son rôle reste encore flou [139], tandis que la déplétion de HRB (*HIV-1 rev binding protein*), capable d'interagir avec NEP [185], induit la rétention des RNPv dans la région périnucléaire [60].

Cependant, un modèle alternatif de transport des RNPv à la membrane plasmique a été récemment proposé suite à l'observation que les RNPv et la protéine Rab11 sont associés à un réticulum endoplasmique (RE) profondément remanié par l'infection virale, qui s'étend jusqu'à la membrane plasmique [148]. Les RNPv nouvellement exportées seraient recrutées et intégrées avec Rab11 dans de nouvelles vésicules, appelées ICV (*irregularly coated vesicles*), formées à la surface de ce RE. Ces vésicules seraient alors transportées jusqu'à la membrane plasmique avant de relâcher les RNPv pour permettre leur empaquetage. Ce modèle confirme le rôle de Rab11 dans le transport des RNPv mais fait intervenir le réticulum endoplasmique et non plus les endosomes de recyclage.

7) Assemblage et bourgeonnement

Une fois arrivées à la membrane plasmique, les RNPv doivent être décrochées de Rab11, cette protéine n'étant pas retrouvée dans les particules virales [227]. Plusieurs études ont montré que le facteur déterminant pour le recrutement à la membrane plasmique des RNPv est la queue cytoplasmique de la protéine M2 [32,109,154] qui interagit directement avec la protéine M1 [32].

Bien que le mécanisme exact d'assemblage des particules virales ne soit pas connu, plusieurs études ont mis en évidence l'importance de nombreuses protéines virales. Les premières expériences ont montré que l'expression de la protéine HA seule permettait la formation et la libération de VLP (*virus-like particle*) [33], de même qu'avec les protéines NA [126,280], M1 [82,258] et M2 [33] seules mais dans une moindre mesure. Cette activité serait due à la capacité de ces protéines à modifier la courbure de la membrane plasmique. Cependant, lors d'une infection virale, la seule présence de HA ne suffit pas à former des particules virales, M2 est indispensable

[32,109,154,155,214]. L'initiation du bourgeonnement est réalisée par les protéines HA et NA, toutes deux localisées au niveau des radeaux lipidiques, bien qu'il ait été montré que HA n'était pas indispensable à cette étape [192]. La protéine M1 est capable de lier les queues cytoplasmiques des protéines HA et NA (pour revue voir [222]), et par son interaction avec NP, permet le recrutement des RNPv au lieu de bourgeonnement. M1 recrute également la protéine M2 qui stabilise le site de bourgeonnement [32,84,155] avant d'être positionnée à l'interface entre l'enveloppe lipidique du virion et la membrane plasmique [215]. M2 va alors utiliser sa propre capacité à modifier la courbure de la membrane pour permettre la dissociation de ces deux membranes et le relargage de la particule virale [215]. Cependant, cette étape ne garantit pas obligatoirement le détachement du virion de la cellule étant donné que les trimères de HA à la surface de la particule virale sont capables d'interagir avec les acides sialiques à la surface de la cellule. Le clivage de ces acides sialiques est réalisé par la protéine NA ce qui permet la libération finale des virions [188,189].

C. Génome segmenté et empaquetage

Bien que la segmentation du génome viral apporte un avantage évolutif important, permettant le mécanisme de cassure antigénique, cette segmentation est une contrainte majeure pour l'étape d'empaquetage qui se déroule en fin de cycle viral. En effet, pour qu'un virus soit infectieux, il doit être comprendre au moins une copie de chaque segment. Le mécanisme d'empaquetage se doit donc d'être le plus efficace possible pour éviter qu'un trop grand nombre de particules virales non infectieuses soient produites et bloquent le cycle d'infection suivant. Ce problème peut être résolu assez facilement lors d'une infection avec une charge virale importante grâce à un mécanisme de *trans*-complémentation. Ce mécanisme permet, lors d'une co-infection, de combiner le matériel génétique de deux particules virales défectives afin de réaliser un cycle viral productif. Cependant, chez l'Homme, la transmission des IVA se fait généralement par les aérosols, soit avec une charge virale très faible. La possibilité d'une *trans*-complémentation est donc peu probable et renforce l'idée selon laquelle le processus d'empaquetage doit permettre à un maximum de particules virales de contenir un jeu complet de huit segments.

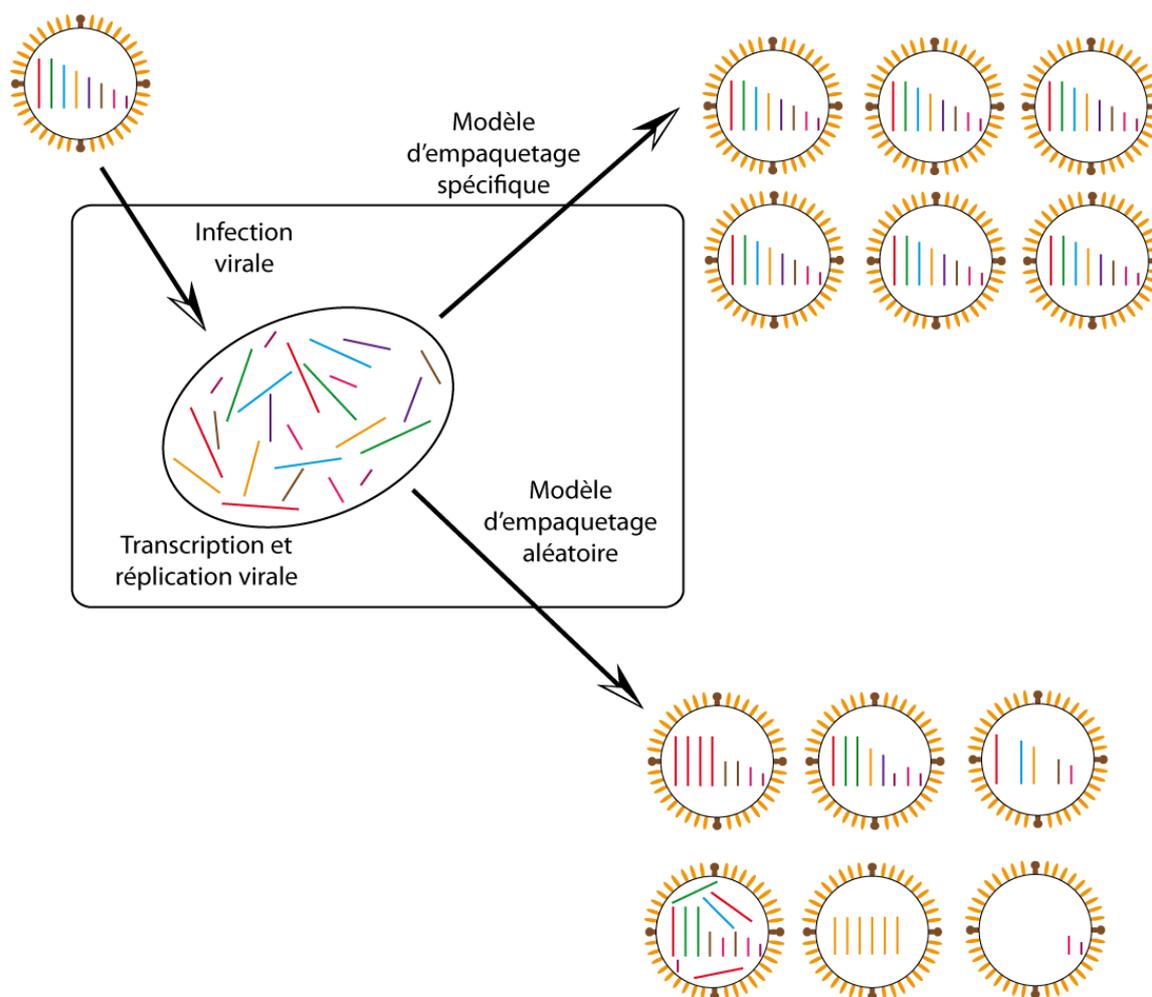


Figure 14 : Représentation schématique des modèles aléatoires et spécifiques initialement proposés pour l’empaquetage des IVA.

Dans la littérature scientifique, deux modèles ont longtemps été opposés pour tenter d’expliquer le mécanisme d’empaquetage des RNPv au sein des IVA (**Figure 14**). Le premier modèle, appelé modèle aléatoire et proposé en 1962 [97], suppose une discrimination entre les ARNv et les autres ARN présents dans le cytoplasme (ARNc, ARNm viraux, ARN cellulaires). Ce modèle assure donc l’empaquetage d’ARN viraux mais ne permet de contrôler ni la quantité de segments ni leur nature. La très grande majorité des particules virales produites seraient alors non infectieuses et l’empaquetage d’un jeu complet de huit segments ne reposerait que sur le hasard. Le second modèle [199], appelé sélectif, suppose qu’en plus de la discrimination entre les ARNv d’une part et les ARN viraux de polarité positive et les ARN cellulaires d’autre part, la machinerie d’empaquetage est capable de discriminer les huit ARNv. Ce modèle implique la présence sur chaque ARNv de séquences spécifiques d’empaquetage. Les nombreuses données accumulées depuis plusieurs années par

différentes équipes à travers le monde sont en faveur de ce mécanisme sélectif et seront discutées ultérieurement.

Les *Orthomyxoviridae*, les IVA en tête, ne sont pas les seuls virus à posséder un génome segmenté, et certains mécanismes d'empaquetage ont été élucidés, permettant d'émettre des hypothèses quant à celui des IVA.

1. L'empaquetage aléatoire chez les virus segmentés : cas de l'IBDV

Le virus de la bursite infectieuse (IBDV), qui fait partie de la famille des *Birnaviridae*, composée de virus à ARN double brin dont le génome est segmenté, incorpore les segments génomiques dans les particules virales de manière aléatoire [168]. Le génome de ce virus est composé de deux segments, A et B, et sa stratégie consiste en l'empaquetage de plusieurs segments, jusqu'à 4 [141], pour augmenter la probabilité d'empaqueter les deux segments qui composent son génome. Cependant, cet empaquetage aléatoire n'empêche pas la présence de signaux spécifiques dans les régions 5' et 3' non codantes des ARN viraux et est réalisé sans interactions entre les deux segments [168]. La présence de structures similaires dans les régions 3' et 5' non codantes de ces deux segments [169] suggère que ces structures pourraient être les signaux responsables de l'empaquetage, bien que la région de l'ARN A codant la protéine VP3 semble être un élément déterminant de l'empaquetage de ce segment [170].

Ce mécanisme est efficace pour un virus dont le génome n'est segmenté qu'en deux parties, cependant pour les IVA ce sont huit segments qui doivent être empaquetés dans la particule virale. Une telle stratégie chez ces virus impliquerait l'empaquetage d'un nombre très important de RNPv pour s'assurer de la présence d'un génome complet.

2. L'empaquetage sélectif chez les virus segmentés : plusieurs exemples

1) Les *Retroviridae*, dont le Virus de l'Immunodéficience Humaine (VIH)

L'exemple des rétrovirus est particulier car leur génome n'est pas, à proprement parlé, segmenté. Tous les gènes sont présents sur une même molécule d'ARN, néanmoins lors de l'empaquetage c'est un dimère d'ARN génomique qui est encapsidé. Cette dimérisation est spécifique et implique l'interaction entre les deux copies d'une

structure en tige-boucle présente dans la région 5' non traduite des deux molécules d'ARN génomique (pour revue voir [56]). Cette même structure est spécifiquement reconnue par la polyprotéine virale Gag, et cette interaction de très haute affinité joue un rôle central dans la sélection de l'ARN génomique du VIH, son transport jusqu'à la membrane plasmique et son encapsidation [1,144,234].

2) Les *Cystoviridae* : l'exemple du phage $\phi 6$

La famille des *Cystoviridae* regroupe des bactériophages possédant un génome à ARNdb segmenté. Ces segments sont au nombre de trois et sont appelés S (small), M (medium) et L (large) d'après leur taille. Le phage $\phi 6$ est l'espèce de référence de cette famille, sur laquelle les études se sont concentrées pour étudier le mécanisme d'empaquetage. L'empaquetage de ces segments est typique des virus à ARNdb, où un ARN simple brin est encapsidé et le second brin est synthétisé par le complexe polymérase à l'intérieur de la particule, dans le but d'éviter une reconnaissance du génome ARNdb par les défenses immunitaires de la cellule. Dans le cas du phage $\phi 6$, il a été clairement démontré que des séquences d'environ 200 nucléotides, appelées pac, spécifiques à chaque fragment, situées 50 nucléotides en aval de l'extrémité 5', sont nécessaires pour l'encapsidation des ARNs viraux. Cette distance est cruciale pour l'encapsidation, sa réduction entraînant une perte de l'efficacité d'empaquetage [203]. Chez le phage $\phi 6$, l'encapsidation des 3 segments d'ARN génomique est réalisée de manière séquentielle : l'ARN S est incorporé en premier dans la particule virale, puis l'ARN M et enfin l'ARN L, avec des changements de conformations du complexe polymérase à l'intérieur de la particule virale entre chaque étape. Ce n'est qu'une fois que tous les segments sont empaquetés que la polymérase virale est activée et synthétise les brins complémentaires (pour revue sur l'empaquetage du phage $\phi 6$ voir [161]).

3) Les *Reoviridae* : deux cas similaires

La famille des *Reoviridae* regroupe des virus dont le génome est composé de 9 à 12 segments d'ARNdb. Au sein de cette famille, le mécanisme d'empaquetage a été élucidé dans deux cas distincts.

Le BTV (*Blue Tongue Virus*) est transmis par les insectes et est responsable d'une mortalité élevée notamment chez les moutons. Il appartient au genre *Orbivirus*, et sont composés de 10 segments génomiques (S1 à S10). Une étude utilisant le retard de

migration sur gel, réalisée sur des ARN transcrits *in vitro*, a montré que les segments S5 et S10 interagissent entre eux, de même que les segments S1 et S7, les séquences d'interaction étant localisées aux extrémités 5' et 3' des ARN [16]. Ces extrémités 5' et 3' non codantes sont complémentaires, de tailles différentes entre chaque segment mais présentent une séquence de 6 nucléotides identiques aux deux extrémités. Il a récemment été montré que ces régions sont responsables du recrutement des différents fragments qui interagissent entre eux pour former un complexe qui sera encapsidé. Ce complexe serait formé de façon ordonnée : le plus petit ARN S10 formerait un complexe avec les autres petits segments S9, S8 et S7. Ce complexe interagirait dans un deuxième temps avec les fragments de taille intermédiaire S6, S5 et S4, puis finalement avec les plus grands fragments S3, S2 et S1. C'est ensuite autour de ce complexe que la capsid se formerait, et une fois tous les segments présents dans cette capsid sous forme simple brin, la polymérase virale réaliserait la synthèse du second brin (pour revue sur l'empaquetage du BTV voir [217]).

Le genre *Rotavirus* regroupe des virus dont le génome est composé de 11 segments et qui sont capables d'infecter les jeunes mammifères (ils sont les responsables principaux des gastro-entérites chez les enfants) et les oiseaux. Il avait été proposé que les signaux spécifiques d'empaquetage de ces ARN étaient présents au sein des extrémités 5' et 3' non traduites, où plusieurs structures avaient été prédites bio-informatiquement (pour revue voir [156]). De même, une interaction entre deux segments (NSP2 et NSP3) a été prédite [238]. Plus récemment, l'équipe de Polly Roy, qui a déjà élucidé le mécanisme d'empaquetage du BTV, a démontré que la région 3' non traduite semble être essentielle pour l'empaquetage de ces virus. En étudiant les plus petits segments du génome, cette équipe a montré *in vitro* que des interactions entre ces segments existent, et que leur abolition a pour effet d'inhiber la réplication virale en culture cellulaire [64]. Cet effet pourrait passer par une déstabilisation des structures présentes au sein de la région 3' non traduite. Il est donc envisageable que le mécanisme d'empaquetage des Rotavirus ressemble fortement à celui du BTV, où des interactions spécifiques entre les segments génomiques forment un complexe qui est encapsidé par la suite (pour revue sur les modèles d'empaquetage des Rotavirus voir [157]).

3. L'empaquetage des IVA, un mécanisme sélectif

Depuis plusieurs années, de nombreuses équipes s'intéressent au mécanisme d'empaquetage des IVA, en vue de mieux comprendre cette étape et, à plus long terme, de pouvoir inhiber ce processus essentiel à la formation de nouvelles particules virales infectieuses. Plusieurs résultats au cours de cette période ont permis de privilégier l'hypothèse de l'empaquetage sélectif au détriment du modèle aléatoire, à commencer par le fait qu'environ 1 à 10% des particules produites sont infectieuses, une proportion élevée comparées aux prédictions du modèle aléatoire [103].

1) Génétique et biochimie

Des expériences de migration électrophorétique ont mis en évidence la présence en quantités équimolaires des huit segments génomiques du virus de la grippe, malgré le fait qu'ils soient exprimés dans des quantités différentes dans la cellule [12,90,158,231]. Il a de plus été montré qu'une seule copie de chaque ARNv est empaqueté dans une particule virale [39], en accord avec des résultats de compétition d'empaquetage de deux pseudos-ARNv, différant dans la région codante mais présentant les mêmes régions non codantes [106].

Plus récemment, Moreira *et coll.* ont mis en évidence l'existence d'un code nucléoprotéique contrôlant l'empaquetage spécifique des RNPv au sein des IVA [167]. Dans cette étude, on observe que le remplacement de certains résidus de la protéine NP d'une souche par une autre modifie la nature des segments incorporés dans les particules virales. En condition sauvage, chaque segment est incorporé de façon égale, tandis que dans les conditions mutées certains segments sont retrouvés plus fréquemment que d'autres. De plus, en fonction de la région de la NP dans laquelle se trouve la mutation, ce ne sont pas les mêmes segments qui sont retrouvés plus fréquemment. Ces résultats suggèrent que certains acides aminés de la NP ont des rôles spécifiques, ces résidus identifiés comme importants pour la coordination de l'empaquetage de plusieurs segments n'ayant pas d'impact sur l'activité de la polymérase et la synthèse de l'ARN.

2) Des signaux d'empaquetage identifiés

L'étude des différentes régions des ARNv a permis de caractériser le rôle de chacune d'entre elles. Ainsi, les régions U12 et U13, seules séquences communes entre tous les

segments, servent à discriminer les ARNv des autres ARN présents dans la cellule. La discrimination des ARNv entre eux fait intervenir les régions non codantes, spécifiques de chaque segment et flanquant la région codante. Ces régions sont en effet suffisantes pour l'empaquetage dans les virions d'ARNv chimères, codant un gène rapporteur flanqué par les régions non codantes des différents segments. Il faut cependant noter que pour la plupart des segments, l'efficacité d'empaquetage de ces ARN chimères est faible en présence des régions non codantes uniquement. L'ajout des extrémités 5' et 3' des régions codantes (la taille de ces extrémités dépend du segment) améliore fortement l'efficacité d'empaquetage du gène rapporteur (pour revue voir [77]).

3) Organisation des RNPv dans la particule virale

Plusieurs études de microscopie électronique ont mis en évidence une organisation conservée en « 7+1 », où une RNPv centrale est entourée de 7 autres [68,69,75,89,181]. Certaines de ces études font également état de contacts entre les différentes RNPv, que ce soit au niveau d'une plateforme située à un pôle de la particule virale [68,69] ou tout au long des RNPv [182]. Les données réunies jusqu'à maintenant suggèrent que les contacts entre les RNPv sont médiés par des interactions ARN-ARN [164], mais il n'est pas possible d'exclure d'autres types d'interactions, faisant intervenir des protéines, cellulaires ou virales, et des signaux d'empaquetage présents sur les ARN.

4) Mise en évidence d'un réseau d'interaction entre ARNv

Des interactions entre les différents ARNv ont été identifiées *in vitro* par retard de migration sur gel, en co-incubant ou co-transcrivant chaque ARNv avec chacun des autres ARNv, puis en analysant les complexes par électrophorèse. Ainsi, des réseaux d'interaction ont été identifiés, dans lesquels chaque segment interagit au moins avec un partenaire [69,75]. Il est intéressant de noter qu'en fonction de la souche étudiée ce réseau n'est pas le même. Ainsi, dans une souche aviaire H5N2 le segment 2 (PB1) interagit avec les segments 3 (PA), 4 (HA) et 6 (NA) [75] alors que dans une souche humaine H3N2 le segment 2 n'interagit qu'avec le segment 6 [69]. Les régions d'interactions entre les segments 2 et 8 du virus H5N2 ont été déterminées au nucléotide près *in vitro*. En culture cellulaire l'abolition de cette interaction conduit à des virus présentant des défauts de réplication liés à des défauts d'empaquetage du génome [76]. Ces résultats démontrent l'importance d'une interaction entre ARNv dans l'empaquetage du génome viral.

5) Le dernier modèle proposé

Dans une revue récente consacrée à l'emballage des IVA, Giese *et coll.* ont proposé un modèle dans lequel une première discrimination entre les RNPv a lieu en amont de l'étape de transport à la membrane. Des RNPv interagiraient entre elles afin d'être acheminées à la membrane par les vésicules Rab11-positives sous forme d'assemblages partiels ('*sub-bundle*'). Ces regroupements de RNPv se rencontreraient au niveau de la zone de bourgeonnement et fusionneraient afin de former l'assemblage ('*bundle*') de huit RNPv qui est emballé dans la nouvelle particule virale [79]. Ce modèle s'appuie sur des expériences de FISH *in cellula* qui montrent la co-localisation de certaines RNPv, après leur export nucléaire et au moment de leur transport par les vésicules Rab11-positives [38,128,250].

6) Dernières études sur l'emballage des Influenza virus

Il a récemment été montré que les particules virales ont tendance à emballer exactement huit segments, même si un des segments génomique est absent [180]. Dans cette étude, des particules virales sont produites *via* l'utilisation d'un système de génétique inverse, où des cellules sont transfectées avec huit plasmides permettant l'expression de chacun des ARNv (sous contrôle d'un promoteur polI) et des ARNm associés (sous contrôle d'un promoteur polII). Avec cette technique, des particules virales ont été produites en l'absence du segment HA et l'analyse de la composition des virions néoformés montre que seulement 17% des particules virales ont incorporé sept segments, les autres en ayant incorporé huit. De façon intéressante, les RNPv sont organisées au sein de la particule virale selon les modèles « 6+1 » et « 7+1 » respectivement. Le huitième segment incorporé, après séquençage à haut débit, correspond aux ARN ribosomiques (ARNr) 18S et 28S, bien que ces derniers ne sont pas retrouvés dans les particules sauvages [180]. De plus, ces ARNr semblent former des RNPv dans la particule virale, mais les mécanismes moléculaires responsables de cette incorporation sont encore inconnus.

Une deuxième étude, sur les IVC et IVD, montre des résultats similaires. Ces deux virus ont un génome composé de seulement sept segments [91]. L'analyse par microscopie électronique, montre que 70% des particules virales produites par ces deux virus comportent huit segments, également organisés selon la conformation « 7+1 » [174]. Le mécanisme d'emballage de huit segments serait donc une

caractéristique conservée au sein des Influenza virus, indépendamment du nombre de segments qui composent leurs génomes.

Il a cependant été précédemment montré, pour les IVA et IVB, que la capacité d'empaqueter efficacement huit segments est dépendante de la souche. Parmi celles étudiées, toutes présentent une très grande majorité de particules virales avec huit segments, mais chez certaines sont retrouvées des particules composées de cinq, six ou sept segments [175].

D. Rôle de la structure des ARN chez les IVA

1. Dans l'empaquetage

Il est donc maintenant admis, au vu des nombreux résultats présentés précédemment, que le mécanisme d'empaquetage des IVA est sélectif. Cette sélection fait intervenir des contacts entre les RNPv, qui sont vraisemblablement dus à des interactions ARN-ARN comme les différents réseaux d'interaction déterminés le suggèrent. L'ARN au sein de la RNPv est en contact avec la polymérase virale et la nucléoprotéine, sa structure est donc un élément clef dans sa capacité à interagir avec les autres molécules environnantes. Bien que la NP s'associe avec l'ARN sur toute sa longueur, certaines régions doivent rester accessibles afin que ces interactions soient possibles. Des études récentes ont effectivement montré que la NP ne se fixe pas uniformément au long de l'ARN, mais qu'il existe des régions riches en NP et d'autres plus libres [133]. Ce profil de fixation de la NP est spécifique de chaque souche virale, bien que certaines régions présentent un profil semblable. Par ailleurs, les régions fixées par la NP semblent être enrichie en guanine et appauvrie en uracile, ce qui pourrait expliquer comment certaines régions restent en simple brin [133]. Ces résultats remettent en cause le modèle de l'architecture de la RNPv utilisée dans la littérature depuis de nombreuses années.

Plus récemment, William *et coll.* ont montré qu'une protéine NP interagit en moyenne avec 12 nucléotides, avec un espacement d'environ 25 nucléotides entre deux sites d'interaction, ce qui est cohérent avec l'idée d'une fixation de la protéine non uniforme le long de l'ARN [272]. Ces régions qui n'interagissent pas avec la NP pourraient former des structures secondaires potentiellement impliquées dans les interactions ARN-ARN et donc importantes pour l'étape d'empaquetage [272].

Ces deux études se sont intéressées au côté protéique de la RNPv, là où une autre étude a regardé la structure de l'ARNv [45]. D'après les résultats obtenus, l'ARNv présent au sein des RNPv est plus ouvert, moins structuré que l'ARNv déprotéinisé, ce qui avait déjà été proposé il y a plusieurs années [11] et fait écho aux résultats précédemment présentés, avec des régions simples brin de l'ARN qui seraient impliquées dans des interactions entre RNPv. La méthode SPLASH (*sequencing of psoralen crosslinked, ligated and selected hybrids*) a permis de proposer un réseau d'interaction entre tous les ARNv de la souche WSN, certaines de ces interactions étant identifiées au nucléotide près. Cependant, il semble peu probable que des interactions puissent s'étendre sur 30 à 100 nucléotides comme proposé [45], étant donné les résultats exposés précédemment quant à la fixation de la NP le long de l'ARNv [272].

D'autres étapes du cycle viral des IVA font également intervenir la structure de certaines régions spécifiques.

2. Rôle des motifs structuraux dans les ARN des IVA : article de revue



Structural and Functional Motifs in Influenza Virus RNAs

Damien Ferhadian, Maud Contrant, Anne Printz-Schweigert, Redmond P. Smyth, Jean-Christophe Paillart and Roland Marquet*

CNRS – UPR 9002, Architecture et Réactivité de l'ARN, IBMC, Université de Strasbourg, Strasbourg, France

Influenza A viruses (IAV) are responsible for recurrent influenza epidemics and occasional devastating pandemics in humans and animals. They belong to the *Orthomyxoviridae* family and their genome consists of eight (–) sense viral RNA (vRNA) segments of different lengths coding for at least 11 viral proteins. A heterotrimeric polymerase complex is bound to the promoter consisting of the 13 5'-terminal and 12 3'-terminal nucleotides of each vRNA, while internal parts of the vRNAs are associated with multiple copies of the viral nucleoprotein (NP), thus forming ribonucleoproteins (vRNP). Transcription and replication of vRNAs result in viral mRNAs (vmRNAs) and complementary RNAs (cRNAs), respectively. Complementary RNAs are the exact positive copies of vRNAs; they also form ribonucleoproteins (cRNPs) and are intermediate templates in the vRNA amplification process. On the contrary, vmRNAs have a 5' cap snatched from cellular mRNAs and a 3' polyA tail, both gained by the viral polymerase complex. Hence, unlike vRNAs and cRNAs, vmRNAs do not have a terminal promoter able to recruit the viral polymerase. Furthermore, synthesis of at least two viral proteins requires vmRNA splicing. Except for extensive analysis of the viral promoter structure and function and a few, mostly bioinformatics, studies addressing the vRNA and vmRNA structure, structural studies of the influenza A vRNAs, cRNAs, and vmRNAs are still in their infancy. The recent crystal structures of the influenza polymerase heterotrimeric complex drastically improved our understanding of the replication and transcription processes. The vRNA structure has been mainly studied *in vitro* using RNA probing, but its structure has been very recently studied within native vRNPs using crosslinking and RNA probing coupled to next generation RNA sequencing. Concerning vmRNAs, most studies focused on the segment M and NS splice sites and several structures initially predicted by bioinformatics analysis have now been validated experimentally and their role in the viral life cycle demonstrated. This review aims to compile the structural motifs found in the different RNA classes (vRNA, cRNA, and vmRNA) of influenza viruses and their function in the viral replication cycle.

OPEN ACCESS

Edited by:

David Gilmer,
Université de Strasbourg, France

Reviewed by:

Levon Abrahamyan,
Université de Montréal, Canada
Isabelle Imbert,
Aix-Marseille Université, France

*Correspondence:

Roland Marquet
r.marquet@ibmc-cnrs.unistra.fr

Specialty section:

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

Received: 23 January 2018

Accepted: 12 March 2018

Published: 29 March 2018

Citation:

Ferhadian D, Contrant M,
Printz-Schweigert A, Smyth RP,
Paillart J-C and Marquet R (2018)
Structural and Functional Motifs
in Influenza Virus RNAs.
Front. Microbiol. 9:559.
doi: 10.3389/fmicb.2018.00559

Keywords: influenza, influenza A virus, RNA, RNA structure, promoter, vRNA, CRNA

INTRODUCTION

Influenza viruses belong to the *Orthomyxoviridae* family that comprises seven genera: *Alphainfluenzavirus*, *Betainfluenzavirus*, *Gammainfluenzavirus*, and *Deltainfluenzavirus*, as well as *Isavirus* (which infects salmon), *Thogotovirus* (transmitted by arthropods), and *Quaranjavirus* (which predominantly infects insects and birds). Among this family, the major threat for public health is the *Alphainfluenzavirus* genera. Influenza B and C

share common ancestry with *Influenza A virus* (IAV). However, *Influenza B* (IBV) and *C* (ICV) viruses mainly infect human and are genetically less diverse. IBV, like IAV, has 8 genome segments coding for at least 11 proteins (**Figure 1A**), whereas Influenza C has only 7 genome segments coding for at least 9 proteins. The segments are either numbered from the longest to the shortest or named after the main protein they code for, e.g., in influenza A viruses, segments 1–8 are also named PB2 (Polymerase Basic 2), PB1 (Polymerase Basic 1), PA (Polymerase Acid), HA (HemAgglutinin), NP (nucleoprotein), NA (NeurAminidase), M (Matrix), and NS (Non-Structural), respectively.

Each segment constitutes an individual replication unit coding for at least one essential protein and form viral ribonucleoproteins (vRNPs). They are composed of a viral RNA (vRNA) of negative polarity, a heterotrimeric polymerase complex bound at the paired 5' and 3' termini of the vRNA and multiple copies of NP (Eisfeld et al., 2015).

After viral entry in the host cell, vRNPs are imported in the nucleus, where transcription and replication take place. During these steps, two different classes of RNA of positive polarity are synthesized: complementary RNA (cRNA), which is the exact copy of vRNA, and viral messenger RNA (vmRNA) (Eisfeld et al., 2015). The latter one is synthesized using a capped primer obtained by cap-snatching, a mechanism during which capped cellular mRNAs are recognized and cleaved by the polymerase complex (Eisfeld et al., 2015). On the contrary, cRNA are synthesized *de novo* and are used as template for producing new vRNA molecules that will be packaged.

The RNA secondary structure is biologically important for many RNA viruses, governing key processes, including viral replication, RNA packaging, RNA editing, and mRNA splicing regulation. Influenza viruses are no exception and several mechanisms during their life cycle are controlled by RNA structure. Steps such as splicing, the switch between transcription and replication, vRNA packaging and recognition by the host immune system have been proposed to be under the control of RNA structure. Only few things are known about the structure of the vRNAs and even less about the structure–function relationships, but things are starting to change dramatically thanks in part to new methodologies exploiting the potential of next generation RNA sequencing.

This review aimed to compile the structural motifs found in the different RNA classes of influenza viruses and their function in the viral replication cycle. A previous review was published many years ago and focused on the differences and similarities between the RNA structure in IAV and the genome structure of other RNA viruses (Gulyaev et al., 2010).

INFLUENZA A VIRUS

Negative Polarity RNA (vRNA)

The Terminal Promoter

For convenience, in this review, the numeration of nucleotides forming the promoter will be as follows: from the 5' terminus, nucleotides will be numbered as X', while from the 3' terminus, they will be numbered as X.

Historical background

The structure of the vRNA promoter has been a controversial topic in the literature, and as many as four different models have been proposed (**Figure 1B**). They all share some features such as base pairing between certain parts of the 3' and 5' ends but are quite different when it comes to the structure of the first nucleotides at each end. Hsu et al. (1987) observed that the vRNA formed circular structures after light induced-RNA crosslinking with psoralen and proposed that these structures are the result of a panhandle structure (**Figure 1B-1**) arising from sequence complementarity between the 3' and 5' vRNA ends. They proposed that this panhandle structure would be found only within vRNPs, and that no interaction between the two ends would take place in 'naked' vRNAs. The same year, Honda et al. (1987) also proposed base pairing between the 3' and 5' ends of the vRNA in their structural model of the RNP core. In contrast, based on RNA probing experiments, Baudin et al. (1994) proposed that within the panhandle, the 3' and 5' ends are not paired (**Figure 1B-2**). Such a pairing would occur in naked vRNAs but the addition of NP would melt the structure. The same year, Fodor et al. (1994) kept the idea of a pairing between the two ends forming the promoter but proposed the 'fork model' for the transcription initiation. In this model, base pairing takes place from nucleotide 13 at the 3' end and 11' at the 5' end, while the vRNA ends are single stranded (Fodor et al., 1994). Flick et al. (1996) proposed a third model called 'corkscrew model' (**Figure 1B-3**). In this model, two short stem-loops are formed within the 3' (nucleotides 1–9) and 5' ends (nucleotides 1' to 9') and a five base-pair 'duplex region' is formed between both ends from nucleotides 11' and 10 to nucleotides 13' and 12 (Flick et al., 1996). Moreover, it was proposed that conservation of nucleotides at the 3' and 5' ends could be extended by up to 4 nucleotides in a segment-specific way (Desselberger et al., 1980), thus allowing an extension of the duplex region of the promoter. An important structural feature observed by Flick et al. (1996) in this promoter is that nucleotide A10' needs to be unpaired and forms a flexible junction. When a U residue was inserted in the 3' end to form a base pair with A10', transcription was impaired (Flick et al., 1996). The unpaired nucleotide A10' is found in all the proposed structures, except for the one from Hsu et al. (1987) where it is nucleotide A11' that is unpaired. However, using a reporter assay, Flick et al. (1996) showed that A11' is paired with U10 and A10' is unpaired. The functions of this specific structural feature will be discussed below.

Crystallographic structure of the polymerase complex and functional implications

The vRNA promoter of IAV is used for two main mechanisms, transcription (which leads to the synthesis of capped and polyadenylated vmRNA) and replication (synthesis of cRNA which will be used as template for vRNA synthesis). The fork model described earlier was proposed in order to explain the initiation of the transcription, where the polymerase complex first binds the single-stranded 5' end of the vRNA and then the single-stranded 3' end (Fodor et al., 1995). However, a recent crystal structure of the vRNA promoter bound to the polymerase complex of bat IAV showed that this is not the conformation

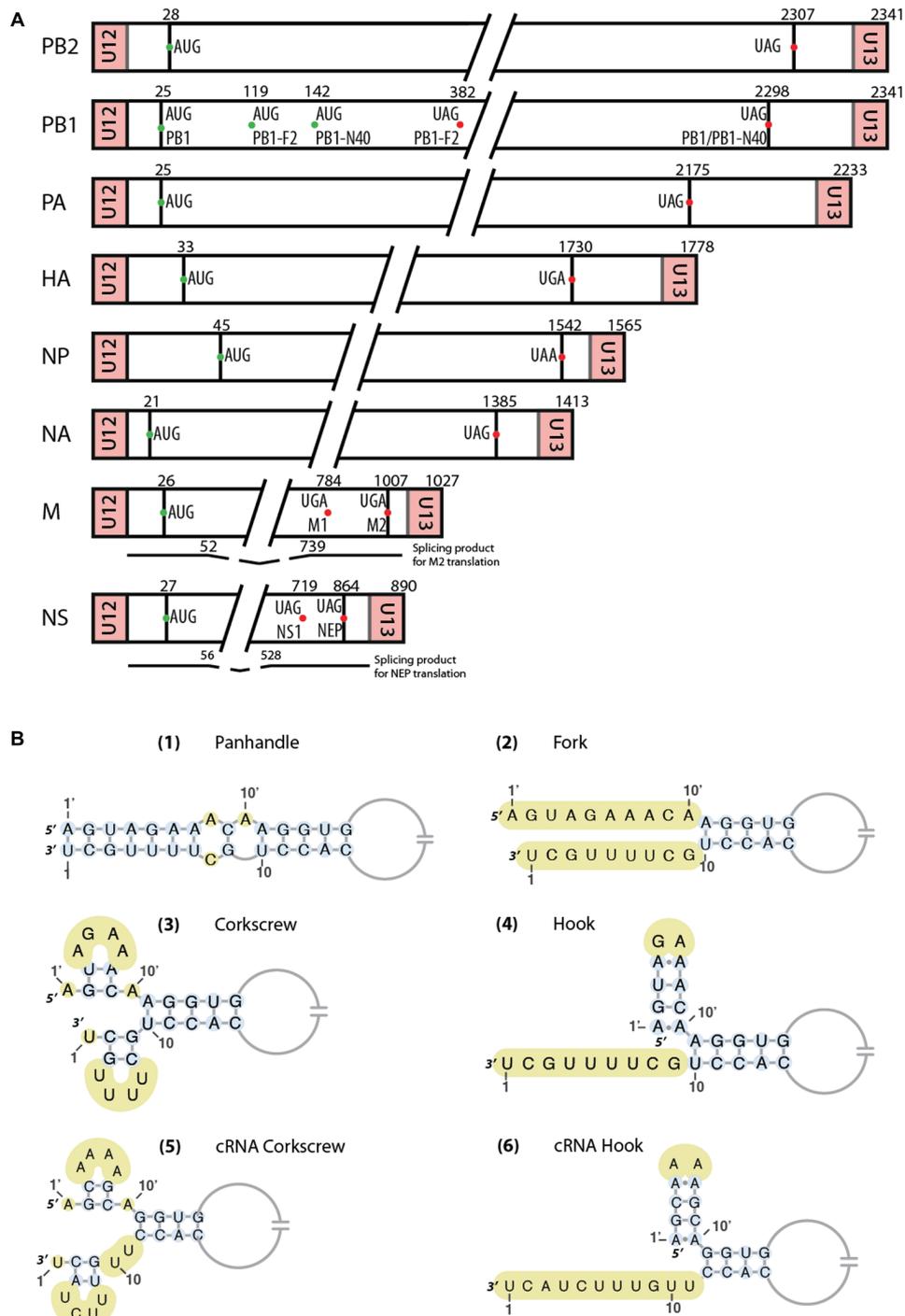
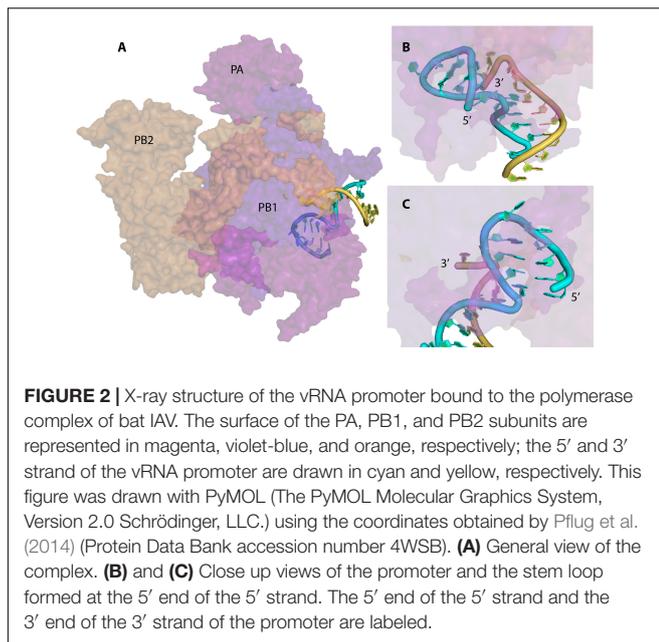


FIGURE 1 | (A) Schematic representation of the eight RNA segments of Influenza A viruses. AUG start codons of the different proteins are represented as green dots, while stop codons are indicated by red dots. Regions between U12/U13 and start codons correspond to the untranslated regions (UTRs). Below the M and NS segments are represented the spliced mRNAs. Important nucleotides positions are indicated. **(B)** Different vRNA promoter structure models proposed over time. Unpaired nucleotides are highlighted in yellow. In the ‘hook structure’ (4,6), two canonical base-pairs are flanked by two non-canonical A-A base-pairs.

adopted by the promoter (Pflug et al., 2014). In this complex, the promoter adopts a structure similar to the ‘corkscrew’ model, except that only the 5’ strand forms a stem-loop, while the

3’ strand is single stranded and enters into the polymerase active site (Figures 1B-4, 2). The 5’ stem-loop (or ‘hook’) is formed by two canonical base-pairs flanked at both sides by non-canonical



A-A base pairs, and this compact structure binds in a pocket at the interface of the PA and PB1 polymerase subunits (Pflug et al., 2014). These results are in agreement with the literature (Hsu et al., 1987; Baudin et al., 1994; Fodor et al., 1994), as it seems that the promoter folds as a panhandle (**Figure 1B-1**) only in the absence of the polymerase. Accordingly, Noble et al. (2011) showed using biophysical analysis that in solution and in the absence of polymerase complex the 5' and 3' vRNA ends adopt a panhandle structure. Kierzek's team reached the same conclusion by performing RNA chemical probing on 'naked' full-length NS and M vRNAs (Lenartowicz et al., 2016; Ruszkowska et al., 2016). However, the general consensus is that the IAV promoter is never free during the viral life cycle, and thus the biological function of this extended panhandle structure (**Figure 1B-1**) is unclear. When the polymerase complex is bound to the promoter, a structural rearrangement occurs with the formation of a hook in the 5' end proximal region, leaving the 3' end single stranded (Pflug et al., 2014). The distal region of both strands remains unchanged and forms a RNA duplex between both strands (**Figure 2**).

As mentioned earlier, the duplex region of the promoter can be extended in a segment-specific way. In the case of the HA segment, this duplex can be extended by forming a base pair between nucleotides 13 and 14'. Wang et al. (2016) analyzed the functional importance of nucleotides 13–15 and 14'–16' (where only nucleotides 13–14' form a base-pair). They found that the 13–14' base pair of the duplex is important for the promoter activity, that base-pairing between nucleotides 14 and 15' enhances this activity, and that nucleotides 15 and 16' do not play any role in promoter activity. Importantly, mutations of the 13–14' base pair and of nucleotide 14 showed a reduction of packaging efficiency not only of the HA segment but also of the PB1, PB2, and PA segments (Wang et al., 2016). This strongly suggests that the segment-specific duplex extension is involved

in transcription/replication and, at least in one case, in vRNA packaging.

Another step of the transcription mechanism that implicates the promoter structure is the endonuclease activity of the polymerase complex. This activity is involved in the cap-snatching mechanism, where the polymerase complex 'steals' the 5' capped end of cellular mRNAs that will be used as primer for the viral mRNA synthesis. Leahy et al. (2001a) determined the importance of the structure of the 5' and 3' ends for the endonuclease activity. They showed, using mutagenic analysis, that the 5' stem of the corkscrew model is essential for this activity as well as the RNA duplex formed by the association of the 3' and 5' strands (**Figure 1B-3**) (Leahy et al., 2001a). Of note, these structural elements also exist in the hook structure observed in the crystal structure of the polymerase bound to the polymerase (**Figure 1B-4**) (Pflug et al., 2014). Using the same approach, Leahy et al. (2001b) showed that some mutations in the 3' end disrupted the endonuclease activity, suggesting the presence of a 3' hairpin loop. They proposed a model in which the polymerase complex binds the 5' hairpin loop and induces a conformational change to form the 3' hairpin loop that will be needed for the cap-snatching mechanism. Importantly, comparison of the X-ray structure of the polymerase complexes of IAV and IBV suggests that capped snatching requires rotation of the PB2 cap-binding domain to direct the capped primer first towards the endonuclease domain in the PA subunit and then into the polymerase active site located in the PB1 subunit (Reich et al., 2014).

During viral transcription, polyadenylation of vmRNAs is also regulated by the vRNA promoter structure. Indeed, Luo et al. (1991) showed that the promoter needs to be structured (3' and 5' end paired) to allow the stuttering of the polymerase on the uridine stretch that is found 15–17 nucleotides from the 3' end, and hence addition of the polyA tail to vmRNA by the transcribing viral polymerase. Indeed, template translocation is likely prevented by tight binding of the 5' hook to the polymerase (Pflug et al., 2017).

The results discussed above show that the vRNA promoter structure is essential for the whole transcription process, from the beginning (cap-snatching) to the end (polyA tailing). In addition, the switch between transcription and replication of the vRNA could also be due, at least partly, to the promoter structure. Indeed, even though the IAV promoter is highly conserved across viral strains and genomic segments, it comes in two variants. The promoter of the three segments coding for the polymerase subunits have a C residue at position 4 (C4 promoter), while the promoter of the five remaining segments has a U residue at this position (U4 promoter). Using C4 and U4 promoters, Lee and Seong (1998) showed that the C4 RNA promoter promotes replication and downregulates transcription, and that these two promoters fold differentially (Lee et al., 2003). However, their NMR structural study (Lee et al., 2003) was performed on short RNA oligonucleotides in the absence of proteins and its biological relevance might be limited. Nevertheless, in the crystal structure of the IBV polymerase, the U4 residue in the 3' strand makes specific contacts with Val133 and Arg135 of the PB1 subunit (Reich et al., 2014), and hence binding to C4 and U4 promoters

is likely slightly different, which might have an effect on the transcription/replication switch.

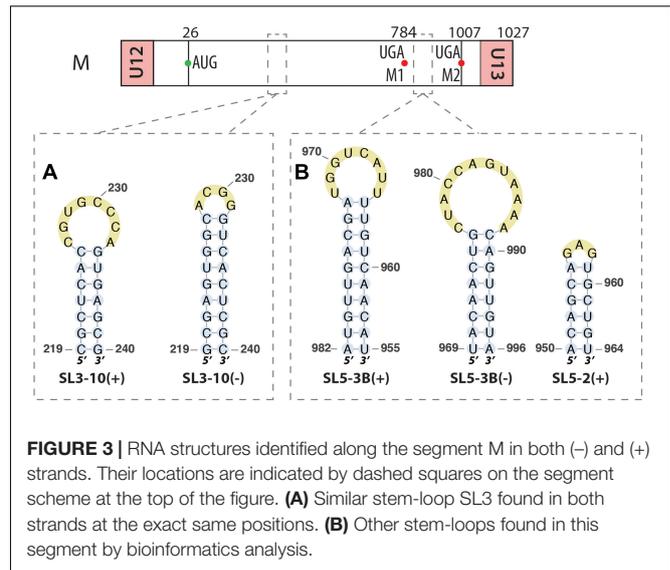
Finally, another mechanism influenced by the vRNA promoter is the recognition of IAV by the host innate immune system. Liu et al. (2015) analyzed the relation between the vRNA promoter of IAV and RIG-I activation. They first demonstrated that transfection of cells with vRNA induced production of IFN in a strain-independent but vRNA dose-dependent manner and that RIG-I is the major actor in the IFN response against IAV. They then mapped the region of the vRNA that is important for RIG-I activation using truncated constructions and showed that the panhandle structure formed by the promoter *in vitro* is sufficient to induce IFN production and more importantly that the degree of complementarity in the proximal step of the promoter is critical for RIG-I activation (Liu et al., 2015). Similarly, Anchisi et al. (2016) observed that the presence of mismatches in the vRNA promoter decreased formation of the vRNA-RIG-I complex and also reduced IFN- β production. Interestingly, the nucleotide composition of the WT promoter region appears to be optimized for a good balance between vRNA and vmRNA synthesis and for limiting RIG-I activation. Lee et al. (2016) further investigated the role of two distinct features of the promoter, the 5'-PPP moiety proposed to be an agonist of RIG-I (Rehwinkel et al., 2010) and the bending at the 5'-terminal stem, in its recognition by RIG-I. They concluded that the helical bend is more important for the interaction with RIG-I than the 5'-PPP moiety (Lee et al., 2016).

The vRNA Internal Region

Until very recently not much was known about the structure of vRNAs, apart from the promoter region. However, several studies addressed this question using *in vitro* or/and bioinformatics approaches or tried to identify regions that are important for the packaging.

Gulyaev et al. (2016) performed a bioinformatics study of the putative conserved secondary structures that might impose evolutionary constraints on the HA vRNA. Covariation analysis indicated that the structured domains in HA vRNA are mostly subtype specific. Several local structured domains that are not essential for viral replication but contribute to viral fitness were identified and were proposed to play a role in virus evolution and re-assortment (Gulyaev et al., 2016). Of note, a similar study on the NP segment identified a functional pseudoknot structure in the packaging signal region of the NP vRNA (Gulyaev et al., 2014).

A bioinformatics study of the RNA structure of segment M in both (+) and (-) strands indicated that the first structure found on both strands is similar, consisting of a stem-loop (nucleotides 219–240), named SL3-10, the difference being that on the (+) strand the loop is 8 nts long, while it is only 4 nts long in (-) strand, but with a longer stem (Figure 3A) (Kobayashi et al., 2016). The same type of structures were found around positions 967–994 for both strands, named SL5-3B, with a second stem loop just before in the (+) strand at positions 950–964, named SL5-2 (Figure 3B) (Kobayashi et al., 2016). Mutations disrupting SL3-10 and SL5-3B structures impaired viral replication and SL3-10 mutants also produced more defective particles, leading to

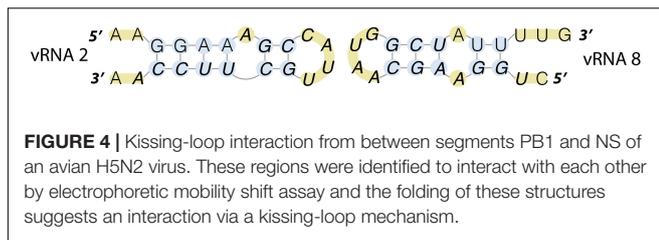


the hypothesis that these secondary structures are important for packaging (Kobayashi et al., 2016).

Secondary structure models of the NS and M fragments based on RNA chemical mapping experiments and structure predictions were recently published (Lenartowicz et al., 2016; Ruszkowska et al., 2016). They showed that these vRNAs are highly structured with many bulges and internal loops. However, since these structures were obtained *in vitro* in the absence of NP protein and polymerase complex, they most likely do not reflect the actual vRNA structure found within vRNPs in the infected cells or in the viral particles. Nevertheless, the presence of multiple single stranded regions suggests that they could constitute potential interaction sites between the different vRNAs leading to an efficient and specific packaging (Gerber et al., 2014).

Such interactions have been described *in vitro* for a human H3N2 (Fournier et al., 2012a,b) and an H5N2 strain (Gavazzi et al., 2013a). It was observed, using an electrophoretic mobility shift assay, that the 8 vRNAs of each of these strains form a network of interactions, but unexpectedly, these networks are strain specific (Fournier et al., 2012a,b; Gavazzi et al., 2013a). These interactions may be involved in the specific packaging of the influenza virus genome. Indeed, the interaction between segments PB1 and NS of the avian H5N2 virus that was identified *in vitro* has actually been shown to be required for optimal packaging into viral particles (Gavazzi et al., 2013b). The interacting sequences can fold into weakly stable stem-loops, suggesting that the interaction between these two RNAs might initiate via a kissing-loop mechanism (Figure 4). To date this remains the only inter-vRNA interaction that has been unambiguously demonstrated to play a role in packaging.

Very recently, additional interactions between the vRNA segments of an H1N1 strain have been proposed using SPLASH (Sequencing of Psoralen crosslinked, Ligated, And Selected Hybrids) (Dadonaite et al., unpublished). Most segments could potentially interact with at least two other vRNAs, and NA and NS are the vRNAs that form the fewest interactions.



Mutations aiming at disrupting the proposed intermolecular interactions resulted in packaging defects (Dadonaite et al., unpublished); however, in the absence of *trans*-compensatory mutants the observed defects could also be attributed to alteration of functionally important intramolecular structures. In the same study, SHAPE-MaP experiments (Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension and Mutational Profiling) were performed to compare the flexibility profile of each vRNA nucleotide *in vitro* and *ex vitro*, after removal of the vRNP-associated proteins (Dadonaite et al., unpublished). The results suggest that the structure is more flexible *in vitro* with uneven NP binding along the genome. Interestingly, a recent HITS-CLIP (High Throughput Sequencing of RNA isolated by CrossLinking ImmunoPrecipitation) study showed that NP does not bind evenly along the vRNAs: in each vRNA, there are zones of low NP density (Lee et al., 2017), which might be involved in intermolecular base-pairing. These results were confirmed by a very recent PAR-CLIP study, which also showed that on average a NP monomer binds 12 vRNA nucleotides and that there is a mean distance of 25 nucleotides between adjacent NP binding sites (Williams et al., 2018). Moreover, these authors identified 24 regions with low NP binding, representing about 10% of the genome, with some of them expected to form secondary structures. Mutations designed to disrupt these structures were introduced in segments PB2, PB1, NP, and NS and resulted in packaging defects, whereas mutations preserving the predicted vRNA structure had no effect on vRNA packaging and viral replication (Williams et al., 2018). These results strongly suggest that the predicted vRNA structures within the low NP-binding regions are important for the coordinated packaging of a complete set of eight IAV segments.

Altogether, these recent studies suggest that both intermolecular vRNA-vRNA structures and intramolecular vRNA structures in NP-devoided regions play a major role in the selective packaging process of influenza viruses.

Positive Polarity RNA

While most of the structural elements that will be described below are specific to cRNA or mRNA, there is a common feature shared by these two molecules. In 2013, a study investigated binding of the NS1 protein to RNA and its specificity using SELEX (Systematic Evolution of Ligands by Exponential enrichment) experiments (Marc et al., 2013). It was shown that both the sequence and the structure were important. The protein recognizes the AGCAAAG motif that is present in both mRNA and cRNA in all IAV segments (Marc et al., 2013). Moreover,

this binding depends on the structure of this motif, since minor changes such as deletion of a 1-nt bulge is sufficient to highly reduce binding (Marc et al., 2013). This structural feature might be important in the viral life cycle for allowing NS1 to play its roles in translation and nucleocytoplasmic transport.

Complementary RNA (cRNA)

Unlike vmRNAs, cRNAs have a terminal promoter that is required for vRNA synthesis and is assumed to be organized in the same way as the vRNA promoter. However, due to the imperfect complementarity between the 5' and 3' RNA termini, the vRNA and cRNA promoters must be slightly different. Comparing the NMR structure of short vRNA and cRNA promoters in the absence of polymerase, Park et al. (2003) observed that the terminal region of the cRNA panhandle is extremely unstable. However, by analogy with the vRNA promoter, it is likely that the cRNA promoter does not adapt a panhandle structure when bound to the polymerase complex (see section "The Terminal Promoter").

In the historical corkscrew model of the vRNA promoter, nucleotide A10' is single stranded and forms a flexible hinge between the short 5' stem-loop and the distal stem (**Figure 1B-3**), whereas in the cRNA promoter A10' can form a base pair with U11, while U10 is unpaired (**Figure 1B-5**). It was proposed that these differences between the promoters induce a different recognition by the polymerase complex (Flick et al., 1996) and allow discrimination between vRNA and the cRNA (Tchatalbachev et al., 2001). Indeed, while both cRNA and vRNA form RNPs, only the latter is found within virions. Based on the biological activity of promoter mutants, Tchatalbachev et al. (2001) suggested that an unpaired nucleotide at position 10' is required for RNP export and hence packaging; they therefore proposed an interaction between M1 and the polymerase complex rather than NP to explain their results (Tchatalbachev et al., 2001). However, in contrast with these results, a recent study reported that vRNPs and cRNPs are both efficiently exported from the nucleus, but only vRNPs traffic to the plasma membrane, indicating that discrimination occurs during cytoplasmic transport (Chaimayo et al., 2017).

If we consider the X-ray crystallographic structure of the vRNA promoter in complex with the viral polymerase (Pflug et al., 2014; Reich et al., 2014), the cRNA promoter can form the same 5' hook structure, including the two non-canonical A-A base pairs (**Figure 1B-6**) that plays a key role in the recognition of the polymerase complex. However, the distal stem of the cRNA promoter is one base pair shorter than its vRNA counterpart, its 3' single-stranded region is one nucleotide longer (due to the unpaired U11), and the single-stranded nucleotides 3, 5, and 8 are different. Interestingly, nucleotides at positions 3 and 8 interact in a base-specific manner with the viral polymerase (Reich et al., 2014) and one can therefore expect that the vRNA and cRNA promoters interact with the polymerase complex in slightly different ways. This is corroborated by the fact that nucleotides 2–5 and 7–9 of the cRNA promoter are essential for activity (Crow et al., 2004).

These authors also found that the 5' hairpin, but not the 3' hairpin, of the corkscrew model is required for the activity of the cRNA promoter, and overall their results are in good agreement with the recently visualized hook model (Crow et al., 2004).

In 2012, a study looked at Global Ordered RNA Structure (GORS) in the whole Influenza genome (both strand orientations). GORS revealed 'excess' thermodynamic stability of wild-type RNA sequences versus random RNA of the same composition. This study demonstrated that GORS only exist in the (+) strand of segments PB2, NP, and NS (Priore et al., 2012) and that these GORS evolved in a host specific way. At the moment, the biological significance of these GORS remains to be demonstrated.

Viral Messenger RNA (vmRNA)

The vRNA and cRNA promoters are conserved in all segments of all influenza strains and thus their structure attracted a lot of attention. However, influenza A vmRNAs might also contain segment-specific structural motifs that regulate processes such as translation, export, or splicing. The later process, which is required for production of the NS2 and M2 proteins, has been the subject of several studies.

Segment NS

Gulyaev et al. (2007) focused on the segment NS vmRNA. Using a bioinformatics approach, they identified a region conserved between different strains of IAV and also in an IBV strain (nucleotides 524–574 for IAV and 716–767 for IBV). This region is predicted to form a hairpin structure, which is more stable in the IAV strains that emerged recently (Gulyaev et al., 2007). They also found that this region can adopt an alternative pseudoknot structure (Figure 5). These results were supported by an independent bioinformatics analysis (Moss et al., 2011). Interestingly, native gel electrophoresis revealed no evidence of the pseudoknot conformation in a strain that presents the stabilized hairpin structure. Considering the location of this structure in the segment NS pre-mRNA (Figure 5), it was suggested that the equilibrium between the two structures could regulate splicing. Indeed, a single mutation predicted to induce misfolding of this RNA region strongly reduced production of the NS2/NEP mRNA and attenuated viral replication, whereas introduction of a second site mutation predicted to restore RNA folding restored splicing and viral replication, demonstrating the importance of the folding of this RNA region for splicing regulation (Jiang et al., 2016).

Ilyinskii et al. (2009) analyzed a region located near the 5' end (nucleotides 81–148) of the same vmRNA. They found that this region folds into a multi-branch structure (Figure 6A). Disruption of this secondary structure reduced transcription of the NS1 vmRNA. However, this structure was questioned by Moss et al. (2011) who proposed an alternative four loop hairpin model for this region (Figure 6B), again using a bioinformatics approach. Priore et al. (2013a) used structure prediction software to discover a third structure that was predicted to fold as a single extended hairpin (Figure 6C). They used RNA chemical probing and isoenergetic microarray techniques to confirm that this

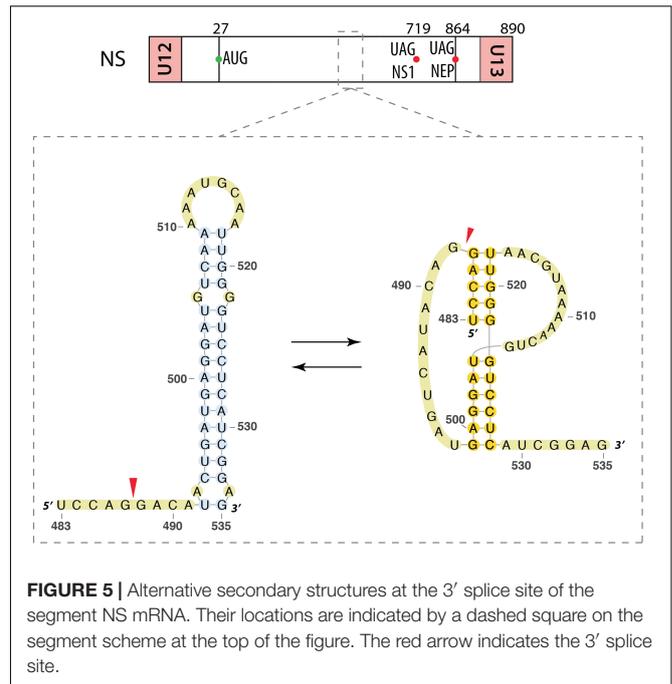
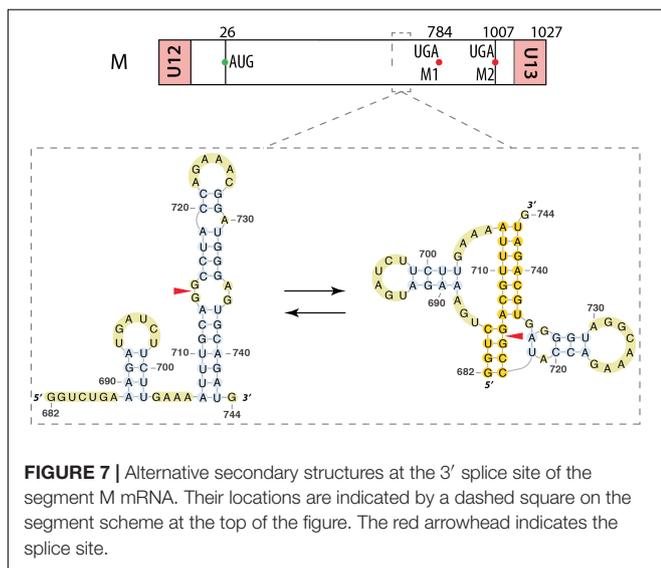
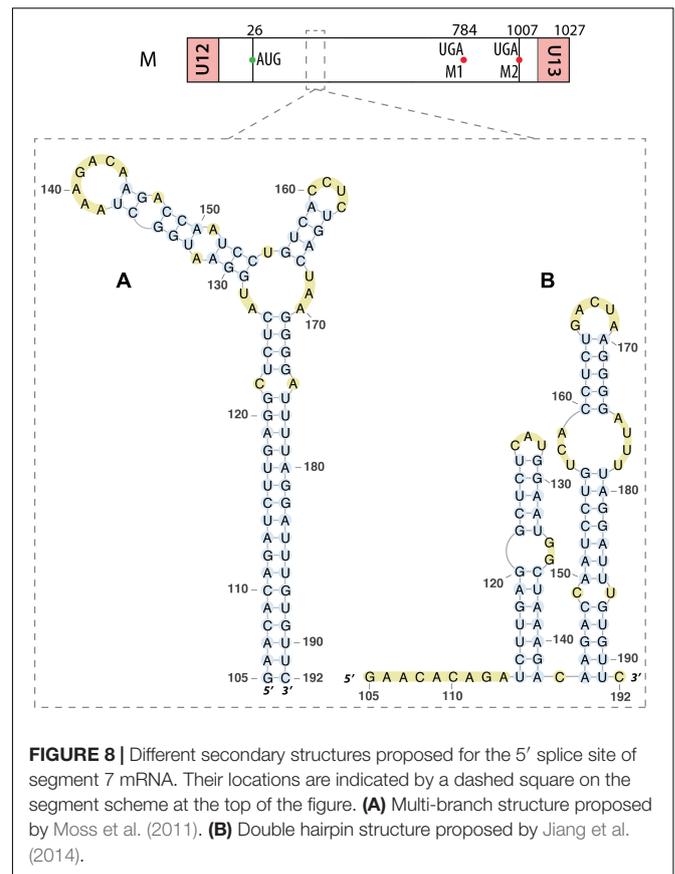
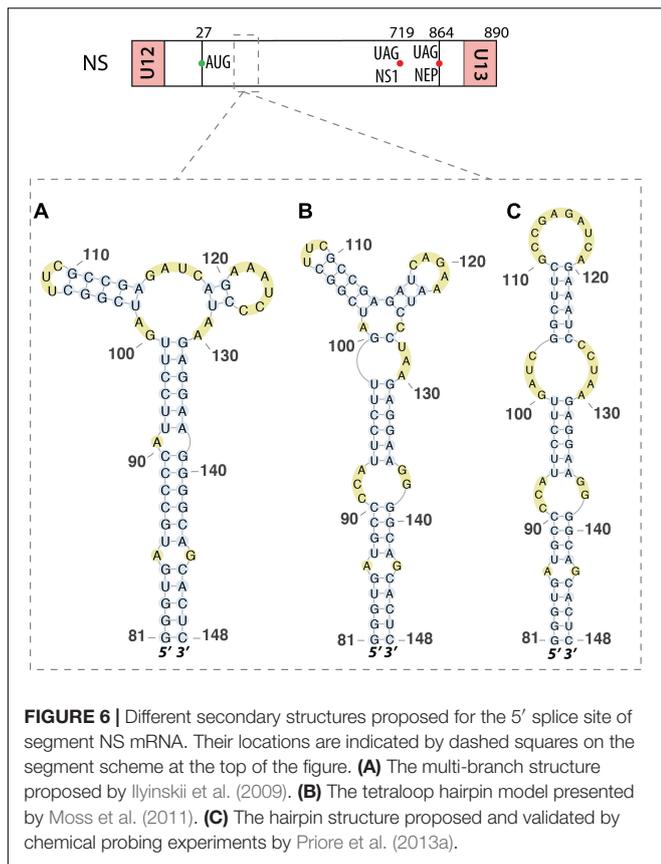


FIGURE 5 | Alternative secondary structures at the 3' splice site of the segment NS mRNA. Their locations are indicated by a dashed square on the segment scheme at the top of the figure. The red arrow indicates the 3' splice site.

hairpin structure exists in the mRNA *in vitro* (Priore et al., 2013a). A recent bioinformatics prediction confirmed earlier findings in this region, with an extended hairpin structure conserved among all human and non-human IAVs and derived from the A/Brevig_Mission/1/1918 (H1N1) strain (Vasin et al., 2016). This study also predicted a hairpin structure derived from the same pandemic H1N1₁₉₁₈ strain between nucleotides 497 and 564 of this RNA (Vasin et al., 2016).

Segment M

Segment M also requires a splicing event to allow expression the M2 ion channel protein. Interestingly, the region surrounding the 3' splice site of this segment (nucleotides 714–715) displays the same structural feature as the 3' splice site of segment NS, namely an equilibrium between a pseudoknot and a double hairpin structure that was first proposed by bioinformatics analysis (Figure 7) (Moss et al., 2011) and supported by native gel electrophoresis, chemical and enzymatic RNA probing, and oligonucleotide binding (Moss et al., 2012a), as well as further sequence analysis (Moss et al., 2012b). As for segment NS, the equilibrium between these structures is thought to play a role in the splicing regulation. The structure of the larger hairpin was determined by NMR and, due to its unique structural features, including a base-triple, it was suggested to constitute a potential therapeutic target (Chen et al., 2015). A region downstream of the 5' splice site of the segment M pre-mRNA (nucleotides 105–192) was also investigated. A multi-branch structure model was predicted by bioinformatics (Figure 8A) (Moss et al., 2011). This region can also adopt an extended hairpin conformation (Figure 8B) but chemical and enzymatic RNA mapping, mutagenesis and NMR all support the multi-branch structure (Jiang et al., 2014). However, mutants designed to induce misfolding of the RNA structure of the regions close to



2015). Indeed, Soszynska-Jozwiak et al. (2015) proposed that this region folds into a multi-branch structure with three hairpins (**Figure 9A**), that they slightly revised based on chemical and enzymatic probing experiments (**Figure 9B**).

Segment PB1

Another pseudoknot was predicted on the PB1 vRNA (nucleotides 65–126); it contains the start codon of two alternative polypeptides encoded by this segment, PB1-F2 and N40. This structure was predicted using bioinformatics (Moss et al., 2011) and confirmed using chemical probing experiments, although another structure was also proposed in this region consisting of two hairpins separated by a single stranded region (Priore et al., 2015). The pseudoknot structure is observed in presence of Mg²⁺ and has been proposed to regulate translation of PB1-F2 and N40.

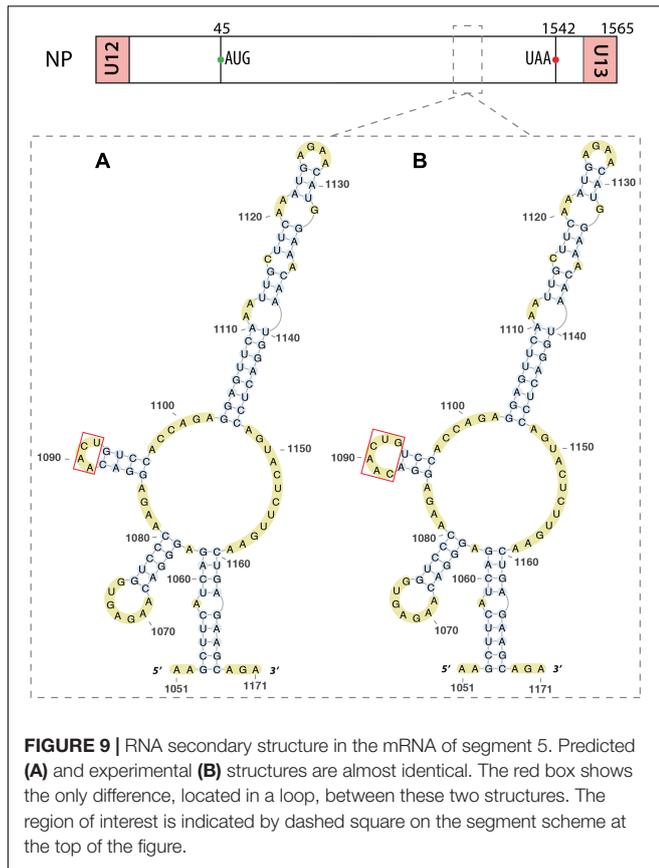
INFLUENZA B AND C VIRUSES

The polymerase complex of Influenza B virus bound to the vRNA promoter was crystalized and its X-ray structure was determined. Indeed, the IBV polymerase binds the vRNA promoter in the same way as the IAV polymerase and the IBV vRNA promoter also adopts the hook conformation when bound to the viral polymerase (Reich et al., 2014). Regarding ICV, only the apo polymerase complex was crystalized, and a comparison of the

the 3' splice site and downstream of the 5' splice site have little effect on the production of the M1 and M2 proteins, even though they have a small effect on viral replication (Jiang et al., 2016).

Segment NP

Within the NP vRNA, nucleotides 1051–1171 were predicted to be structured (Moss et al., 2011; Soszynska-Jozwiak et al.,



structures of the influenza polymerases with and without a bound vRNA promoter suggests that the apo polymerase is in a “closed-state” that is activated upon binding of the promoter (Hengrung et al., 2015).

To determine possible structural constraints acting on Influenza viruses, protein-coding sequences of IBV were used to perform a survey for conserved secondary structure in the (+) and the (–) RNAs. The GORS analysis revealed similarities between IAV and IBV and GORS have been found in both RNA orientations in NP, NS, and PB2 segments (Priore et al., 2013b). In general, influenza B RNA motifs are more stable in the (+) RNA than in the (–) RNA, except for the HA fragment. Some differences also exist between IAV and IAB as segments M and PB1 display GORS only in IAV and IBV, respectively.

Previous studies on IAV described a role for RNA structural motifs in the splicing of M and NS vRNAs. The corresponding regions of IBV and ICV genomes have been scanned with the RNAz program and regions of high thermodynamic stability and conserved RNA secondary structure overlapping the 5′ splice sites of IBV segment NS and ICV segments M and NS were detected. In contrast, the structural motifs detected close to the IAV 5′ splice sites of segments M and NS do not overlap with these sites (Dela-Moss et al., 2014).

Segment 8 (NS) of IAV is homologous to segments 8 and 7 of IBV and ICV, respectively, and in all cases NS2/NEP mRNA

is produced by splicing. In both IBV and ICV, the 5′ splice site is located into the stem of a conserved hairpin, possibly sequestering the splice site. Additionally, a conserved hairpin is also predicted specifically in segment 6 (M) of ICV, in which the M1 protein is produced by splicing, in contrast to IAV and IBV, in which M1 is translated from the unspliced vRNA (Dela-Moss et al., 2014). Moreover, in both IBV and ICV, the 3′ splice sites of the two shortest segments can fold either as a hairpin or as a pseudoknot, changing the accessibility of the splice site, as also described for IAV (Gulyaev et al., 2007; Moss et al., 2011, 2012a). Interestingly, naturally occurring mutations in regions bordering the splice sites preserve base pairing (Dela-Moss et al., 2014).

CONCLUSION

In this review, we focused on the relation between the RNA structure of influenza viruses and its function in the replication cycle. Most of our knowledge about influenza RNA structure and function deals with the vRNA/cRNA promoters and the splice sites of vRNAs. Concerning the promoter, the recent X-ray structures of the viral heterotrimeric polymerase alone and bound to the vRNA promoter (Pflug et al., 2014; Reich et al., 2014; Hengrung et al., 2015) shed a new light on previous functional studies, but further studies will be required to fully understand the viral replication and transcription mechanism. Given their key functional roles and their conservation among all segments of all IAV strains, vRNA and cRNA promoters might be good targets for the development of anti-IAV drugs. Several studies already investigated this possibility. A compound named DPQ [6,7-dimethoxy-2-(1-piperazinyl)-4-quinazolinamine] was found to bind the vRNA promoter and induces important changes in its 1D NMR imino resonance spectrum. In cell-based assays, DPQ reduces replication of IAV and IBV but it is not as potent as FDA-approved drugs (Lee et al., 2014). However, more potent derivatives of DPQ have been identified that pave the way for the design new drugs (Bottini et al., 2015).

Concerning the splice sites, several different structures have been proposed based on bioinformatics predictions. In several cases, chemical and enzymatic mapping experiments helped identify the actual structure. Interestingly, the IAV 3′ splice site of both segment M and NS mRNAs share the same structural feature, which has been shown to regulate splicing and viral replication. By contrast, the M and NS 5′ splice sites adopt different structures.

Until recently, our knowledge of the structure of the internal part of vRNAs within vRNPs was almost inexistent. However, rapid progress in this area has been made within the last months thanks to the coupling of several classical RNA biochemistry tools with next generation sequencing (Lee et al., 2017; Williams et al., 2018; Dadonaite et al., unpublished). These studies identified NP-rich and NP-poor regions in each vRNAs (Lee et al., 2017; Williams et al., 2018) and obtained genome wide structural information (Dadonaite et al., unpublished). Several intramolecular structures and intermolecular RNA–RNA

interactions have been proposed in these studies (Williams et al., 2018; Dadonaite et al., unpublished), which open the way for new functional and structural studies.

AUTHOR CONTRIBUTIONS

DF, MC, AP-S, RPS, J-CP, and RM conceived the review topic. DF, MC, and RM drafted the manuscript and DF generated the

figures. AP-S, RPS, and J-CP corrected and edited the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by Grant No. ANR-15-CE11-0014-01 'PSiFlu' from the Agence Nationale de la Recherche (ANR) to RM.

REFERENCES

- Anchisi, S., Guerra, J., Mottet-Osman, G., and Garcin, D. (2016). Mismatches in the influenza A Virus RNA panhandle prevent retinoic acid-inducible Gene I (RIG-I) sensing by impairing RNA/RIG-I complex formation. *J. Virol.* 90, 586–590. doi: 10.1128/JVI.01671-15
- Baudin, F., Bach, C., Cusack, S., and Ruigrok, R. W. (1994). Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO J.* 13, 3158–3165.
- Bottini, A., De, S. K., Wu, B., Tang, C., Varani, G., and Pellicchia, M. (2015). Targeting influenza a virus RNA promoter. *Chem. Biol. Drug Des.* 86, 663–673. doi: 10.1111/cbdd.12534
- Chaimayo, C., Hayashi, T., Underwood, A., Hodges, E., and Takimoto, T. (2017). Selective incorporation of vRNP into influenza A virions determined by its specific interaction with M1 protein. *Virology* 505, 23–32. doi: 10.1016/j.virol.2017.02.008
- Chen, J. L., Kennedy, S. D., and Turner, D. H. (2015). Structural features of a 3' splice site in influenza a. *Biochemistry* 54, 3269–3285. doi: 10.1021/acs.biochem.5b00012
- Crow, M., Deng, T., Addley, M., and Brownlee, G. G. (2004). Mutational analysis of the influenza virus cRNA promoter and identification of nucleotides critical for replication. *J. Virol.* 78, 6263–6270. doi: 10.1128/JVI.78.12.6263-6270.2004
- Dela-Moss, L. I., Moss, W. N., and Turner, D. H. (2014). Identification of conserved RNA secondary structures at influenza B and C splice sites reveals similarities and differences between influenza A, B, and C. *BMC Res. Notes* 7:22. doi: 10.1186/1756-0500-7-22
- Desselberger, U., Racaniello, V. R., Zazra, J. J., and Palese, P. (1980). The 3' and 5'-terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. *Gene* 8, 315–328. doi: 10.1016/0378-1119(80)90007-4
- Eisfeld, A. J., Neumann, G., and Kawaoka, Y. (2015). At the centre: influenza A virus ribonucleoproteins. *Nat. Rev. Microbiol.* 13, 28–41. doi: 10.1038/nrmicro3367
- Flick, R., Neumann, G., Hoffmann, E., Neumeier, E., and Hobom, G. (1996). Promoter elements in the influenza vRNA terminal structure. *RNA* 2, 1046–1057.
- Fodor, E., Pritlove, D. C., and Brownlee, G. G. (1994). The influenza virus panhandle is involved in the initiation of transcription. *J. Virol.* 68, 4092–4096.
- Fodor, E., Pritlove, D. C., and Brownlee, G. G. (1995). Characterization of the RNA-fork model of virion RNA in the initiation of transcription in influenza A virus. *J. Virol.* 69, 4012–4019.
- Fournier, E., Moules, V., Essere, B., Paillart, J.-C., Sirbat, J.-D., Cavalier, A., et al. (2012a). Interaction network linking the human H3N2 influenza A virus genomic RNA segments. *Vaccine* 30, 7359–7367. doi: 10.1016/j.vaccine.2012.09.079
- Fournier, E., Moules, V., Essere, B., Paillart, J.-C., Sirbat, J.-D., Isel, C., et al. (2012b). A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic Acids Res.* 40, 2197–2209. doi: 10.1093/nar/gkr985
- Gavazzi, C., Isel, C., Fournier, E., Moules, V., Cavalier, A., Thomas, D., et al. (2013a). An in vitro network of intermolecular interactions between viral RNA segments of an avian H5N2 influenza A virus: comparison with a human H3N2 virus. *Nucleic Acids Res.* 41, 1241–1254. doi: 10.1093/nar/gks1181
- Gavazzi, C., Yver, M., Isel, C., Smyth, R. P., Rosa-Calatrava, M., Lina, B., et al. (2013b). A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16604–16609. doi: 10.1073/pnas.1314419110
- Gerber, M., Isel, C., Moules, V., and Marquet, R. (2014). Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends Microbiol.* 22, 446–455. doi: 10.1016/j.tim.2014.04.001
- Gultyaev, A. P., Fouchier, R. A., and Olsthoorn, R. C. (2010). Influenza virus RNA structure: unique and common features. *Int. Rev. Immunol.* 29, 533–556. doi: 10.3109/08830185.2010.507828
- Gultyaev, A. P., Heus, H. A., and Olsthoorn, R. C. (2007). An RNA conformational shift in recent H5N1 influenza A viruses. *Bioinformatics* 23, 272–276. doi: 10.1093/bioinformatics/btl559
- Gultyaev, A. P., Spronken, M. I., Richard, M., Schrauwen, E. J. A., Olsthoorn, R. C. L., and Fouchier, R. A. M. (2016). Subtype-specific structural constraints in the evolution of influenza A virus hemagglutinin genes. *Sci. Rep.* 6:sre38892. doi: 10.1038/srep38892
- Gultyaev, A. P., Tsyganov-Bodounov, A., Spronken, M. I., van der Kooij, S., Fouchier, R. A. M., and Olsthoorn, R. C. L. (2014). RNA structural constraints in the evolution of the influenza A virus genome NP segment. *RNA Biol.* 11, 942–952. doi: 10.4161/rna.29730
- Hengrung, N., El Omari, K., Serna Martin, I., Vreede, F. T., Cusack, S., Rambo, R. P., et al. (2015). Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. *Nature* 527, 114–117. doi: 10.1038/nature15525
- Honda, A., Ueda, K., Nagata, K., and Ishihama, A. (1987). Identification of the RNA polymerase-binding site on genome RNA of influenza virus. *J. Biochem.* 102, 1241–1249. doi: 10.1093/oxfordjournals.jbchem.a122163
- Hsu, M. T., Parvin, J. D., Gupta, S., Krystal, M., and Palese, P. (1987). Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. *Proc. Natl. Acad. Sci. U.S.A.* 84, 8140–8144. doi: 10.1073/pnas.84.22.8140
- Ilyinskii, P. O., Schmidt, T., Lukashev, D., Meriin, A. B., Thoidis, G., Frishman, D., et al. (2009). Importance of mRNA secondary structural elements for the expression of influenza virus genes. *OMICS* 13, 421–430. doi: 10.1089/omi.2009.0036
- Jiang, T., Kennedy, S. D., Moss, W. N., Kierzek, E., and Turner, D. H. (2014). Secondary structure of a conserved domain in an intron of influenza A M1 mRNA. *Biochemistry* 53, 5236–5248. doi: 10.1021/bi500611j
- Jiang, T., Nogales, A., Baker, S. F., Martinez-Sobrido, L., and Turner, D. H. (2016). Mutations designed by ensemble defect to misfold conserved RNA structures of influenza A segments 7 and 8 affect splicing and attenuate viral replication in cell culture. *PLoS One* 11:e0156906. doi: 10.1371/journal.pone.0156906
- Kobayashi, Y., Dadonaite, B., van Doremalen, N., Suzuki, Y., Barclay, W. S., and Pybus, O. G. (2016). Computational and molecular analysis of conserved influenza A virus RNA secondary structures involved in infectious virion production. *RNA Biol.* 13, 883–894. doi: 10.1080/15476286.2016.1208331
- Leahy, M. B., Dobbyn, H. C., and Brownlee, G. G. (2001a). Hairpin loop structure in the 3' arm of the influenza A virus virion RNA promoter is required for endonuclease activity. *J. Virol.* 75, 7042–7049. doi: 10.1128/JVI.75.15.7042-7049.2001
- Leahy, M. B., Pritlove, D. C., Poon, L. L. M., and Brownlee, G. G. (2001b). Mutagenic analysis of the 5' Arm of the influenza A Virus virion RNA promoter defines the sequence requirements for endonuclease activity. *J. Virol.* 75, 134–142. doi: 10.1128/JVI.75.1.134-142.2001
- Lee, K. H., and Seong, B. L. (1998). The position 4 nucleotide at the 3' end of the influenza virus neuraminidase vRNA is involved in temporal regulation of transcription and replication of neuraminidase RNAs and affects the

- repertoire of influenza virus surface antigens. *J. Gen. Virol.* 79(Pt 8), 1923–1934. doi: 10.1099/0022-1317-79-8-1923
- Lee, M.-K., Bae, S.-H., Park, C.-J., Cheong, H.-K., Cheong, C., and Choi, B.-S. (2003). A single-nucleotide natural variation (U4 to C4) in an influenza A virus promoter exhibits a large structural change: implications for differential viral RNA synthesis by RNA-dependent RNA polymerase. *Nucleic Acids Res.* 31, 1216–1223. doi: 10.1093/nar/gkg214
- Lee, M.-K., Bottini, A., Kim, M., Bardaro, M. F., Zhang, Z., Pellicchia, M., et al. (2014). A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication. *Chem. Commun.* 50, 368–370. doi: 10.1039/c3cc46973e
- Lee, M.-K., Kim, H.-E., Park, E.-B., Lee, J., Kim, K.-H., Lim, K., et al. (2016). Structural features of influenza A virus panhandle RNA enabling the activation of RIG-I independently of 5'-triphosphate. *Nucleic Acids Res.* 44, 8407–8416. doi: 10.1093/nar/gkw525
- Lee, N., Le Sage, V., Nanni, A. V., Snyder, D. J., Cooper, V. S., and Lakdawala, S. S. (2017). Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucleic Acids Res.* 45, 8968–8977. doi: 10.1093/nar/gkx584
- Lenartowicz, E., Kesy, J., Ruskowska, A., Soszynska-Jozwiak, M., Michalak, P., Moss, W. N., et al. (2016). Self-folding of naked segment 8 genomic RNA of influenza A Virus. *PLoS One* 11:e0148281. doi: 10.1371/journal.pone.0148281
- Liu, G., Park, H.-S., Pyo, H.-M., Liu, Q., and Zhou, Y. (2015). Influenza A Virus panhandle structure is directly involved in RIG-I activation and interferon induction. *J. Virol.* 89, 6067–6079. doi: 10.1128/JVI.00232-15
- Luo, G. X., Luytjes, W., Enami, M., and Palese, P. (1991). The polyadenylation signal of influenza virus RNA involves a stretch of uridines followed by the RNA duplex of the panhandle structure. *J. Virol.* 65, 2861–2867.
- Marc, D., Barbachou, S., and Soubieux, D. (2013). The RNA-binding domain of influenza virus non-structural protein-1 cooperatively binds to virus-specific RNA sequences in a structure-dependent manner. *Nucleic Acids Res.* 41, 434–449. doi: 10.1093/nar/gks979
- Moss, W. N., Dela-Moss, L. I., Kierzek, E., Kierzek, R., Priore, S. F., and Turner, D. H. (2012a). The 3' splice site of influenza A segment 7 mRNA can exist in two conformations: a pseudoknot and a hairpin. *PLoS One* 7:e38323. doi: 10.1371/journal.pone.0038323
- Moss, W. N., Dela-Moss, L. I., Priore, S. F., and Turner, D. H. (2012b). The influenza A segment 7 mRNA 3' splice site pseudoknot/hairpin family. *RNA Biol.* 9, 1305–1310. doi: 10.4161/rna.22343
- Moss, W. N., Priore, S. F., and Turner, D. H. (2011). Identification of potential conserved RNA secondary structure throughout influenza A coding regions. *RNA* 17, 991–1011. doi: 10.1261/rna.2619511
- Noble, E., Mathews, D. H., Chen, J. L., Turner, D. H., Takimoto, T., and Kim, B. (2011). Biophysical analysis of influenza A virus RNA promoter at physiological temperatures. *J. Biol. Chem.* 286, 22965–22970. doi: 10.1074/jbc.M111.239509
- Park, C.-J., Bae, S.-H., Lee, M.-K., Varani, G., and Choi, B.-S. (2003). Solution structure of the influenza A virus cRNA promoter: implications for differential recognition of viral promoter structures by RNA-dependent RNA polymerase. *Nucleic Acids Res.* 31, 2824–2832. doi: 10.1093/nar/gkg387
- Pflug, A., Guilligay, D., Reich, S., and Cusack, S. (2014). Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* 516, 355–360. doi: 10.1038/nature14008
- Pflug, A., Lukarska, M., Resa-Infante, P., Reich, S., and Cusack, S. (2017). Structural insights into RNA synthesis by the influenza virus transcription-replication machine. *Virus Res.* 234, 103–117. doi: 10.1016/j.virusres.2017.01.013
- Priore, S. F., Kauffmann, A. D., Baman, J. R., and Turner, D. H. (2015). The Influenza A PB1-F2 & N40 start codons are contained within an RNA pseudoknot. *Biochemistry* 54, 3413–3415. doi: 10.1021/bi501564d
- Priore, S. F., Kierzek, E., Kierzek, R., Baman, J. R., Moss, W. N., Dela-Moss, L. I., et al. (2013a). Secondary structure of a conserved domain in the intron of influenza A NS1 mRNA. *PLoS One* 8:e70615. doi: 10.1371/journal.pone.0070615
- Priore, S. F., Moss, W. N., and Turner, D. H. (2012). Influenza A virus coding regions exhibit host-specific global ordered RNA structure. *PLoS One* 7:e35989. doi: 10.1371/journal.pone.0035989
- Priore, S. F., Moss, W. N., and Turner, D. H. (2013b). Influenza B virus has global ordered RNA structure in (+) and (–) strands but relatively less stable predicted RNA folding free energy than allowed by the encoded protein sequence. *BMC Res. Notes* 6:330. doi: 10.1186/1756-0500-6-330
- Rehwinkel, J., Tan, C. P., Goubau, D., Schulz, O., Pichlmair, A., Bier, K., et al. (2010). RIG-I detects viral genomic RNA during negative-strand RNA virus infection. *Cell* 140, 397–408. doi: 10.1016/j.cell.2010.01.020
- Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crépin, T., et al. (2014). Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* 516, 361–366. doi: 10.1038/nature14009
- Ruskowska, A., Lenartowicz, E., Moss, W. N., Kierzek, R., and Kierzek, E. (2016). Secondary structure model of the naked segment 7 influenza A virus genomic RNA. *Biochem. J.* 473, 4327–4348. doi: 10.1042/BCJ20160651
- Soszynska-Jozwiak, M., Michalak, P., Moss, W. N., Kierzek, R., and Kierzek, E. (2015). A conserved secondary structural element in the coding region of the influenza A virus nucleoprotein (NP) mRNA is important for the regulation of viral proliferation. *PLoS One* 10:e0141132. doi: 10.1371/journal.pone.0141132
- Tchatalbachev, S., Flick, R., and Hobom, G. (2001). The packaging signal of influenza viral RNA molecules. *RNA* 7, 979–989. doi: 10.1017/S1355838201002424
- Vasin, A. V., Petrova, A. V., Egorov, V. V., Plotnikova, M. A., Klotchenko, S. A., Karpenko, M. N., et al. (2016). The influenza A virus NS genome segment displays lineage-specific patterns in predicted RNA secondary structure. *BMC Res. Notes* 9:279. doi: 10.1186/s13104-016-2083-6
- Wang, J., Li, J., Zhao, L., Cao, M., and Deng, T. (2016). Dual roles of the hemagglutinin segment-specific noncoding nucleotides in the extended duplex region of the Influenza A Virus RNA promoter. *J. Virol.* 91:JVI.1931-16. doi: 10.1128/JVI.01931-16
- Williams, G. D., Townsend, D., Wylie, K. M., Kim, P. J., Amarasinghe, G. K., Kutluay, S. B., et al. (2018). Nucleotide resolution mapping of influenza A virus nucleoprotein-RNA interactions reveals RNA features required for replication. *Nat. Commun.* 9:465. doi: 10.1038/s41467-018-02886-w

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a shared affiliation, though no other collaboration, with the authors DF, MC, AP-S, RS, J-CP, RM.

Copyright © 2018 Ferhadian, Contrant, Printz-Schweigert, Smyth, Paillart and Marquet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

E. Les protéines chaperonnes de l'ARN

1. Généralités

Pour assurer leurs fonctions, les ARN doivent être repliés correctement. Etant donné leur grande flexibilité, le repliement des molécules d'ARN peut s'avérer difficile et l'ARN peut se retrouver piégé dans un état mal replié et inactif, dont il ne peut sortir spontanément à cause d'une trop grande énergie d'activation [31,95]. Pour pallier à ce problème, les cellules et les virus codent des protéines capables de remodeler la structure de l'ARN. Ces protéines sont divisées en deux groupes distincts, les hélicases et les chaperonnes. Les hélicases sont classées en superfamilles en fonction de leurs motifs hélicase conservés [112] et permettent des réarrangements structuraux de façon ATP-dépendante au contraire des protéines chaperonnes de l'ARN qui n'ont pas besoin d'ATP pour remodeler la structure de l'ARN [173]. De plus, les protéines caractérisées comme chaperonne de l'ARN ont d'autres rôles au sein de la cellule ou du cycle viral tels que la régulation de la transcription, l'export de l'ARN ou l'assemblage des RNP [204]. Les protéines chaperonnes de l'ARN possèdent généralement des domaines intrinsèquement désordonnés [43]. Plusieurs méthodes expérimentales ont été développées pour déterminer la capacité des protéines à remodeler la structure de l'ARN. Chaque méthode met en évidence les différentes propriétés de ces protéines, ces dernières n'étant pas toutes actives dans chaque approche (pour revue sur les méthodes expérimentales voir [205]). Les approches les plus utilisées font intervenir l'épissage en *cis* ou en *trans* des introns de groupe I ainsi que la capacité de coupure de l'ARN des ribozymes à « tête de marteau ».

Une classification des protéines chaperonnes de l'ARN a été réalisée selon leurs fonctions au sein de la cellule (pour revue complète sur les différentes classes de protéines chaperonnes voir [204] ainsi que le site internet <https://www.projects.mfpl.ac.at/rnachaperones/> qui recense les protéines possédant une activité chaperonne dans la littérature.

Les ribonucléoprotéines nucléaires hétérogènes (hnRNP) sont des protéines qui s'associent aux transcrits de l'ARN polymérase II et qui, par leur activité chaperonne, permettent les mécanismes d'épissage, d'export ou de traduction.

Les protéines à domaine « *cold shock* » (CSP) sont principalement retrouvées chez les bactéries et permettent de maintenir la structure de l'ARN à basse température,

condition dans laquelle les problèmes de repliement de l'ARN sont plus importants. Ainsi, les protéines chaperonnes CspA et CspE, identifiée chez *E. coli* jouent un rôle d'anti-terminateur de la transcription et améliorent l'efficacité de traduction [9].

Certaines protéines ribosomiques possèdent également une activité chaperonne. Les protéines S12 et S1 chez *E. coli* présentent, d'après les méthodes expérimentales présentées précédemment, la plus forte activité chaperonne. La totalité des 34 protéines de la grande sous-unité du ribosome d'*E. coli* a été analysée et près d'un tiers présentent une activité chaperonne [224].

2. Les protéines chaperonnes chez les virus à ARN

Les virus, et plus particulièrement les virus à ARN, possèdent au sein de leurs molécules d'ARN des structures variées, aussi bien dans les régions 5' et 3' non traduites que dans les régions codantes, et qui ont un rôle dans plusieurs étapes du cycle répliatif, comme la répliation, la traduction ou l'encapsidation. De nombreux virus codent donc des protéines possédant une activité chaperonne afin de garantir l'intégrité de ces structures (pour une revue voir [275]).

Le VIH-1 fait partie de ces virus, et code trois protéines capables de remodeler la structure de l'ARN, la plus importante étant la protéine de nucléocapside NCp7. L'activité chaperonne de cette protéine intervient lors de l'étape de rétro-transcription, en améliorant la fixation de l'ARN^{t_{Lys}3}, servant d'amorce, au niveau du PBS (*primer binding site*) ainsi qu'en améliorant l'efficacité du transfert du brin négatif lors de la transcription inverse (pour revue voir [135,136]). La protéine régulatrice Tat (*trans-activator of transcription*), par son activité chaperonne, est impliquée de façon similaire dans l'étape de transcription inverse, bien que Tat ne soit pas capable de réaliser des réarrangements structuraux spécifiques contrairement à NCp7 (pour revue voir [81]). La dernière protéine virale, Vif, dont l'activité chaperonne a été déterminée expérimentalement *in vitro*, pourrait jouer un rôle, en plus de la transcription inverse, dans la dimérisation de l'ARN génomique viral ainsi que dans sa maturation, de concert avec NCp7 [10,94].

Parmi les virus à ARN de polarité positive, les *Flaviviridae* et plus particulièrement le virus de l'Hépatite C (HCV) code une protéine chaperonne de l'ARN, appelée *core protein*, et qui agit sur la région 3' non traduite de l'ARN génomique viral. La fonction de ce réarrangement structural dans le cycle viral reste cependant à déterminer [225].

Les *Picornaviridae* et les *Coronaviridae* sont deux familles de virus à ARN de polarité positive qui codent également des protéines chaperonnes de l'ARN impliquées aussi bien dans le mécanisme de réplication virale pour la protéine N des *Coronaviridae* que dans la protection du génome, la traduction ou la réplication virale pour les protéines 3AB et 2CATPase des *Picornaviridae* [275].

Parmi les virus à ARN de polarité négative, seule la protéine de nucléocapside N du genre *Hantavirus* (famille des *Bunyaviridae*) est caractérisée comme protéine chaperonne de l'ARN. Elle serait impliquée dans l'initiation de la traduction et faciliterait la conformation en « poêle à frire » nécessaire à l'encapsidation du génome [162]. Les protéines chaperonne de l'ARN sont également retrouvées chez les virus à ARN double brin et plus particulièrement chez les virus de la famille des *Reoviridae*. Cependant, la propriété chaperonne de protéine NSP2 des *Rotavirus* est controversée, contrairement à la protéine VP5 des *Cypovirus* [275].

Le nombre de protéines codées par les virus à ARN possédant une activité chaperonne de l'ARN démontre l'importance du bon repliement de régions spécifiques des ARN viraux à toutes les étapes des cycles viraux.

F. Les techniques d'étude de la structure de l'ARN

Différentes techniques existent pour déterminer la structure 2D ou 3D d'un ARN. Les premières méthodes mises en place dans ce but utilisent des réactifs chimiques et des enzymes spécifiques de certaines bases ou de certains motifs structuraux. Plusieurs approches biophysiques ont également été développées pour résoudre la structure des ARN.

1. Les méthodes biophysiques

Pour étudier la structure de l'ARN, trois approches biophysiques sont principalement utilisées, la cristallographie, la résonance magnétique nucléaire (RMN) et la cryo-microscopie électronique (cryo-EM). Toutes ces techniques permettent de déterminer la structure 3D de l'ARN. La cristallographie de l'ARN s'est développée au milieu des années 90, passant de 31 structures disponibles en 1995 à près de 2500 en 2015, bien que toutes ne soient pas de l'ARN seul, mais aussi des complexes avec des protéines, comme le ribosome [264]. Cependant, cette méthode est

longue et compliquée à mettre en place, notamment parce qu'il faut tester de très nombreuses conditions de cristallisation (pour revue voir [98,210]).

L'utilisation de la RMN est limitée par la taille des molécules que l'on peut analyser, généralement autour de 75 nucléotides. Cependant, la structure de la région 5'-leader du VIH-1, formant un dimère de 230 kDa, a été élucidée par RMN, ouvrant son utilisation à de plus grandes molécules [138]. Le développement de la RMN à l'état solide permet d'analyser des ARN de plus grande taille sous forme de microcristaux, de solutions congelées, d'échantillons lyophilisés ou encore d'échantillons solubles. Cette technique a permis de d'étudier, par exemple, l'interaction entre la protéine Tat du VIH-1 avec la tige boucle TAR (pour revue voir [276]).

La cryo-EM quant à elle est une technique largement utilisée depuis plusieurs années, permettant d'obtenir, dans le meilleur des cas, des structures avec une résolution quasi-atomique. Cependant, cette technique a surtout été utilisée pour des protéines, et hormis les complexes tels que les ribosomes, l'ARN n'est que très peu été étudié par cryo-EM. Il existe moins de 10 structures de molécules d'ARN seules déposées à ce jour dans la banque de données de microscopie électronique (*electron microscopy data bank* - EMDB), et aucune d'elles ne présentent une résolution meilleure que 20Å. Le principal problème lié à cette technique est la limite de taille de l'échantillon. La plus petite protéine dont la structure a été élucidée à la résolution atomique a un poids moléculaire de 64 kDa [46,117]. L'analyse de la structure des ARN par cryo-EM doit donc être réalisée sur des échantillons de grande taille, alors que ce sont les grands ARN qui présentent la plus grande flexibilité, ce qui est problématique pour la résolution des structures [83]. L'étude de grands ARN par cryo-EM ne permet donc que de visualiser la structure globale de ces derniers [74]. Jusqu'à présent, les ARN de moins de 30 kDa étaient considérés trop petits pour être étudiés par cryo-EM, mais une étude a élucidé cette année la structure du duplex étendu formé par les structures DIS (*dimerization initiation site*) localisées dans la région 5' leader de l'ARN génomique du VIH-1 [286]. Ce duplex a un poids moléculaire de 30 kDa et la résolution de cette structure obtenue en cryo-EM est de 9 Å, ce qui ouvre de nouvelles perspectives pour le domaine de la détermination de la structure des ARN.

2. Les méthodes biochimiques

Même si les techniques précédemment évoquées se développent, la majorité des structures d'ARN publiées sont obtenues à la suite d'expériences de cartographies enzymatiques ou chimiques.

La cartographie enzymatique utilise des ribonucléases (RNase) ciblant les régions d'ARN en simple ou en double brin, permettant de déterminer l'état d'appariement de la molécule cible. Les nucléases les plus utilisées sont la RNase V1 (pour les ARNdb), la RNase S1 (pour les ARNsb), la RNase A (pour les nucléotides C et U dans les régions en simple brin) et la RNase T1 (pour les nucléotides G dans les régions en simple brin) [121]. Afin d'avoir les résultats les plus précis possibles, plusieurs conditions sont utilisées pour un même ARN, chacune d'entre elles utilisant une RNase différente. Jusqu'il y a peu, les fragments d'ARN obtenus après clivage étaient analysés sur gel, en utilisant des ARN marqués radioactivement à leurs extrémités (3' ou 5') avec en parallèle des pistes de séquençage permettant de déterminer le nucléotide précédant le site de clivage [113,121,179]. Les sites de clivage peuvent également être identifiés par extension d'une amorce radioactive et détectés sur gel. Avec l'essor du séquençage nouvelle génération, un nouveau protocole appelé PARS (*parallel analysis of RNA structure*) a été développé afin d'analyser les produits de clivage par séquençage haut débit, permettant d'obtenir une quantité d'informations beaucoup plus importante que sur gel, analysable plus facilement et plus précisément [116,257]. Cependant, et malgré les apports de ces dernières années, cette méthode reste limitée à l'étude de structures d'ARN *in vitro* étant donné que les RNase ne traversent pas les membranes cellulaires [124].

La cartographie chimique est basée sur la modification des nucléotides non appariés dans la structure de l'ARN à l'aide de sondes chimiques dans des conditions statistiques, avec une modification par molécule d'ARN. Deux grandes familles de sonde chimique sont utilisées, permettant la modification spécifique de certaines régions de l'ARN. Certaines sondes comme le DMS (diméthylsulfate) ou le CMCT (1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluenesulfonate) sont spécifiques de certains nucléotides (respectivement A/C et U/G) et la modification (**Figure 15**) est retrouvée au niveau de la face Watson-Crick des bases [124]. D'autres sondes sont universelles, dont les sondes SHAPE (*selective 2'-hydroxyl acylation*

analyzed by primer extension), qui acylent les nucléotides (**Figure 15**) au niveau du 2'OH porté par le ribose [271].

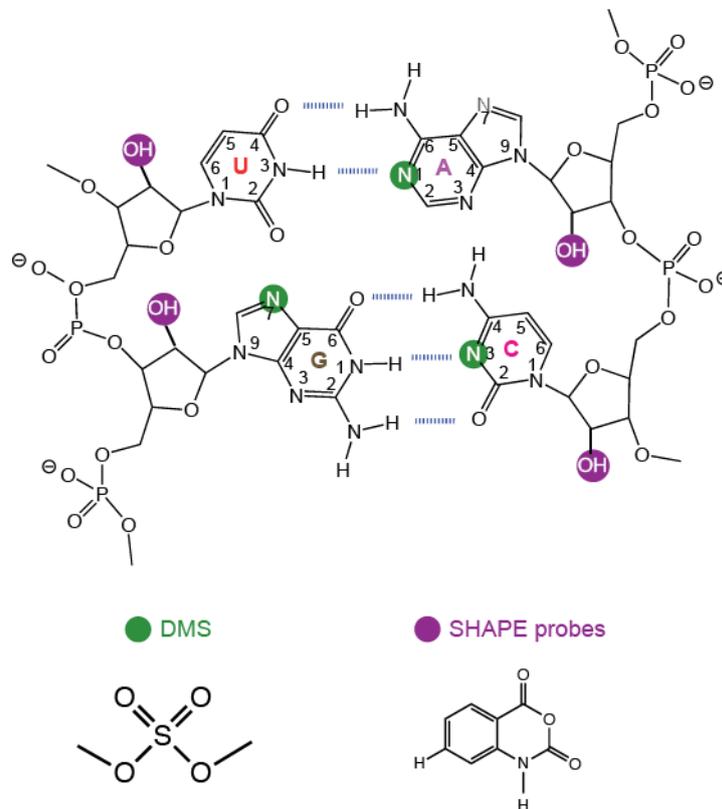


Figure 15 : Position des modifications produites par le DMS (en vert) et les sondes SHAPE (en violet). Les structures du DMS et du NMIA sont représentées.

Toutes ces modifications sont ensuite détectées par extension d'amorce, la rétrotranscriptase étant bloquée par la première modification rencontrée. Les premières expériences de cartographie chimique ont été réalisées à l'aide d'amorces marquées radioactivement et analysées sur gel de polyacrylamide [58]. Par la suite, les amorces radioactives ont été remplacées par des amorces fluorescentes et les ADN complémentaires (ADNc) sont analysés par électrophorèse capillaire [270]. Tout comme pour la cartographie enzymatique, la démocratisation du séquençage à haut débit a permis la mise au point de nouvelles méthodes combinant la cartographie chimique en solution et le séquençages au débit, aussi bien avec des sondes SHAPE [140,235] qu'avec le DMS [53,216,240]. Ces approches, appelées SHAPE-Seq, icSHAPE (**Figure 16A**) et DMS-Seq, font intervenir de nombreuses étapes pour la création des banques de séquençage à haut débit, basées sur la ligation d'adaptateurs. Ces multiples étapes augmentent le risque de perdre le matériel biologique et l'introduction de biais. De nouvelles méthodes, appelées MaP (*mutational profiling*),

ont été développées pour pallier à ces limitations et reposent non pas sur l'arrêt de la transcription inverse mais sur l'introduction d'une mutation dans l'ADNc en face du nucléotide modifié, en utilisant des conditions de transcription inverse spécifique (**Figure 16B**), notamment en présence d'ions manganèse [232,289]. L'analyse après séquençage permet de définir un taux de mutation pour chaque nucléotide, directement relié à sa flexibilité au sein de la structure. De plus, l'utilisation de conditions mutagène, plutôt que d'arrêts de l'enzyme, permet de s'affranchir de la modification en conditions statistiques. L'approche MaP permet donc d'analyser la totalité des nucléotides de chaque molécule d'ARN.

Les méthodes de cartographie chimique, contrairement à la cartographie enzymatique, font intervenir des sondes chimiques qui, pour certaines (DMS, NMIA, 1M7) peuvent traverser les membranes cellulaires et permettent donc de modifier directement les ARN *in cellula* ou *in viro* [233,235].

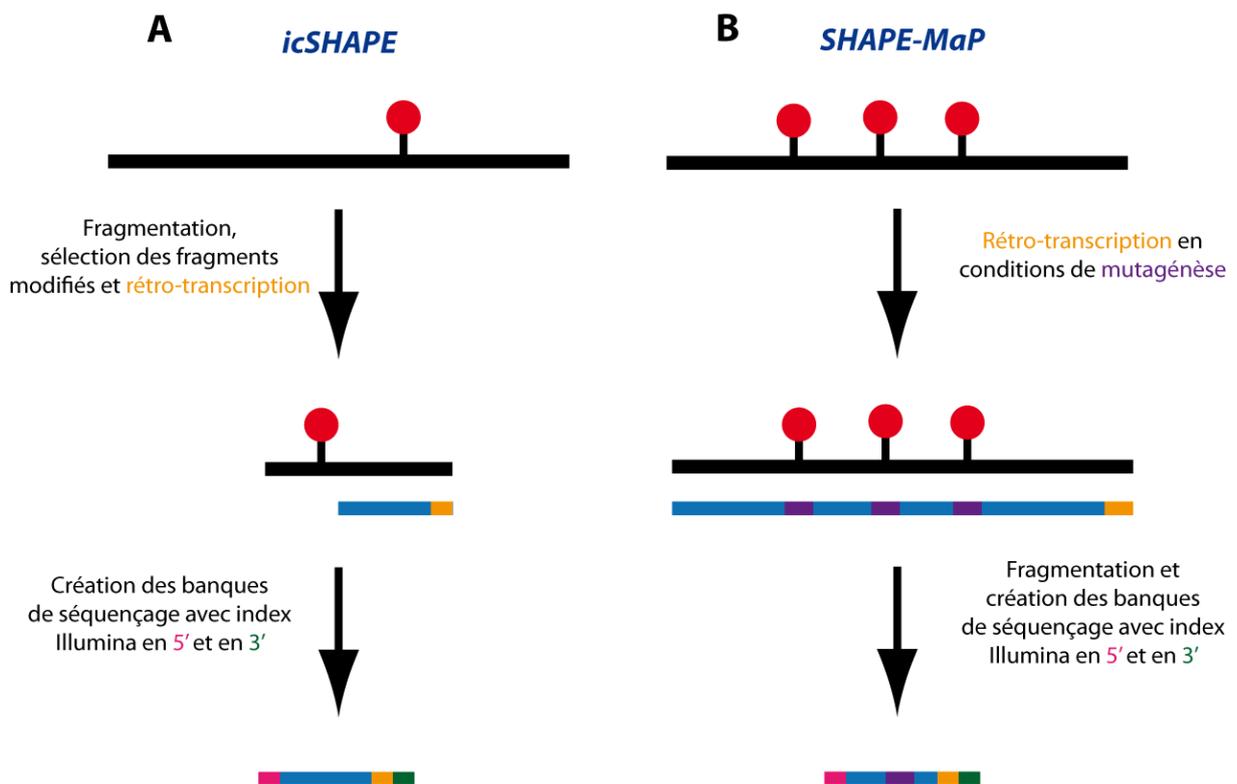


Figure 16 : Comparaison des principes des approches icSHAPE (A) et SHAPE-MaP (B).

PROBLÉMATIQUE
&
SUJET DE RECHERCHE

Problématique & sujet de recherche

Le génome des IVA est composé de huit segments d'ARN de polarité négative. Ces ARN sont présents au sein de la particule virale sous forme de RNPv, où un complexe polymérase hétérotrimérique est attaché aux extrémités 5' et 3' qui interagissent entre elles et le reste de l'ARN est associé à de multiples copies de la protéine virale NP.

La segmentation du génome chez les IVA présente des avantages évolutifs mais complique le processus d'empaquetage nécessaire à la production de nouvelles particules virales. Il est aujourd'hui admis, suite à de nombreuses études, que l'empaquetage est sélectif, ce qui implique une discrimination entre les ARNv et les autres ARN présents dans la cellule, ainsi qu'entre les différents ARNv.

Des études de microscopie électronique réalisées sur des particules virales mettent en évidence des contacts entre les RNPv, qui peuvent s'expliquer par des interactions spécifiques ARN-ARN ou ARN-protéine (virales ou cellulaires). Jusqu'à présent, aucune interaction spécifique entre les ARNv et des protéines jouant un rôle dans le mécanisme d'empaquetage des IVA, n'a été détectée. A l'inverse, notre équipe a mis en évidence l'existence de réseaux d'interactions entre les différents ARNv d'une même souche. L'ARNv étant contraint au sein d'une RNPv dans la particule virale, sa structure doit jouer un rôle primordial dans sa capacité à interagir avec les RNPv voisines. Très peu d'études à ce jour se sont intéressées à la structure des ARN des IVA, et la plupart d'entre elles se sont intéressées à des ARN transcrits *in vitro*, en absence de tout composant de la RNPv.

Le composant majeur de la RNPv étant la protéine virale NP, sa fixation à l'ARN doit être le déterminant majeur de la structure de l'ARN au sein du complexe ribonucléoprotéique. Je me suis donc intéressé, dans un premier temps, à la fixation *in vitro* de la NP à l'ARN en i) caractérisant les sites de fixation de la NP (taille, présence d'une séquence ou d'un motif structural consensus), ii) analysant l'impact de la fixation de la NP sur la structure de l'ARN et iii) identifiant une éventuelle activité chaperonne d'ARN de la protéine NP, ce qui n'a pas été clairement démontrée à ce jour.

Dans un second temps, j'ai cherché à identifier les régions de chaque ARNv impliquées dans les interactions entre les RNPv au sein des particules virales en m'intéressant à la structure de l'ARN au sein des complexes ribonucléoprotéiques. Le but de cette étude est i) de mieux comprendre le mécanisme d'empaquetage des RNPv

au sein des nouvelles particules virales, ii) d'identifier le réseau d'interaction entre les ARN_v pour la souche étudiée et iii) de caractériser l'impact de la structure de l'ARN_v au sein des RNP_v sur le mécanisme d'empaquetage.

Pour répondre à ces questions, j'ai utilisé la cartographie chimique en solution, et plus précisément l'approche SHAPE qui consiste en l'acylation du 2'OH du ribose des nucléotides flexibles (non appariés et non protégés par l'interaction avec des protéines) à l'aide d'une sonde chimique. Une étape de transcription inverse permet alors d'identifier les nucléotides non appariés. Les ADN_c sont ensuite analysés sur séquenceur capillaire dans le cas des expériences *in vitro*, grâce à l'utilisation d'amorces fluorescentes, ou par séquençage à haut débit dans le cas des expériences en culture cellulaire.

Les expériences *in vitro* ont dans un premier temps été réalisées sur la souche modèle de laboratoire de sous type H1N1, A/Puerto Rico/8/1934 (PR8) puis sur la souche A/WSN/1933 (WSN), de même sous-type, dont la protéine NP s'est avérée plus facile à surproduire. L'utilisation de deux souches différentes m'a tout de même permis d'identifier des similitudes entre les deux protéines NP étudiées, ce qui m'a permis de généraliser sur les propriétés de fixation de la protéine virale sur l'ARN.

Pour les expériences en culture cellulaire, j'ai collaboré avec l'équipe du Dr Martin Schwemmle (Institute of Virology, Freiburg, Allemagne) qui étudie un virus de sous-type H7N7, A/seal/Mass/1-SC35M/1980 (SC35M). La production des particules virales a été réalisée par cette équipe. La mise en place de la méthodologie icSHAPE (*in cell SHAPE*) s'étant révélée extrêmement difficile, j'ai opté pour la technique de SHAPE-MaP (*SHAPE-Mutational profiling*), comportant moins d'étapes expérimentales que j'ai pu adapter à mon système d'étude.

MATÉRIEL
&
MÉTHODES

Matériel & Méthodes

A. Transcription *in vitro*

Les séquences permettant l'expression des ARNv des souches PR8 et WSN sont clonés dans les plasmides pUC2000. Ce plasmide possède un promoteur T7 en orientation antisens, deux sites de restriction pour l'enzyme BsmBI qui permettent le clonage des séquences des différents segments des IVA et un site unique de restriction pour la linéarisation du plasmide (Bsh1236I pour ou Ecl136II en fonction du segment). Tous les plasmides utilisés au cours de ma thèse pour la transcription *in vitro* étaient disponibles à mon arrivée au laboratoire.

Avant l'étape de transcription 100 µg de plasmide sont linéarisés pendant une nuit à 37 °C avec 0,4 U d'enzyme par µg d'ADN dans le tampon 1X préconisé par le fournisseur de l'enzyme de restriction (Bsh1236I pour M et Ecl1236II pour NS). Une extraction au phénol-chloroforme (v/v) est réalisée et suivie d'une précipitation à l'éthanol.

Protocole de précipitation à l'éthanol :

Ajouter 0,1 vol d'acétate de sodium 3 M (pH 5), 3 vol d'éthanol 100 % et 1 µg de glycogène. Incuber 45 min sur bain carboglace-éthanol puis centrifuger 45 min à 21000 g à 4 °C. Laver les culots deux fois avec de l'éthanol 75 % à -20 °C (centrifuger 5 min à 21000 g à 4 °C entre chaque lavage) puis sécher 5 min à l'air libre.

Les culots sont repris dans 25 µl d'eau miliQ (Millipore). Une aliquote est prélevée et déposée sur gel d'agarose 0,8 % (p/v) TBE 1X, contenant 0,2 µg/ml de BET (Roth), afin de vérifier l'efficacité de la linéarisation.

La totalité de l'échantillon digéré est utilisée pour la transcription *in vitro*. La réaction est réalisée dans un volume final de 300 µl avec 4 mM de chaque NTP, 0,01 % (v/v) de Triton X-100, 400 U de RNasine RNase inhibitor (Promega), 1,7 mM de spermidine, 5 mM de DTT, 3 µl d'ARN polymérase T7 produite et purifiée au laboratoire, 30 µl de tampon 10X (400 mM Tris-HCl pH 8 à 37 °C ; 150 mM MgCl₂ ; 500 mM NaCl) et incubée 3 h à 37 °C. Après 1 h d'incubation, 3 µl d'ARN polymérase T7 sont ajoutés pour améliorer l'efficacité de transcription. Les échantillons sont ensuite traités pendant 2 h à 37 °C avec 50 U de DNase I recombinante RNase-free

(Roche) dans le tampon 1X du fournisseur (40 mM Tris-HCl ; 10 mM NaCl ; 6 mM MgCl₂ ; 1 mM CaCl₂ ; pH 7,9). Après une extraction au phénol-chloroforme (v/v) l'ARN est précipité puis repris dans 500 µl d'eau MilliQ.

Pour la purification des ARN, une colonne de tamis moléculaire TSK-Gel® G4000SW (Tosoh Bioscience) est installée sur un système ÄKTApurifier (GE Healthcare) et équilibrée à un débit de 1 ml/min dans un tampon 0,2 M acétate de sodium et 1% (v/v) de méthanol. Les échantillons transcrits *in vitro* sont ensuite passés sur la colonne à un débit de 1 ml/min et les fractions (1 ml) contenant des acides nucléiques (identifiées par absorbance à 254 nm) sont précipitées par ajout de 3 vol d'éthanol. Les échantillons sont ensuite repris dans 100 à 300 µl d'eau MilliQ et la qualité des ARN est contrôlée sur gel (7 x 10 cm) de polyacrylamide 8%, urée 8M en conditions dénaturantes avec une migration à 200V pendant 2 h.

B. Production et purification de la protéine NP

Les séquences permettant l'expression des protéines NP des souches PR8 et WSN sont clonées dans les vecteurs d'expressions bactériens pET28a et pET22b respectivement, sous la dépendance d'un promoteur T7. Ces plasmides codent pour la résistance aux antibiotiques kanamycine et ampicilline respectivement. La construction pour l'expression de la protéine NP de WSN (**Figure 17B**) nous a été fournie par le Dr Thibaut Crépin (Grenoble). Le plasmide contenant la séquence codant la protéine NP de PR8 a été produit au laboratoire en utilisant les sites de restriction NcoI (extrémité 5') et XhoI (extrémité 3') présents dans la cassette de clonage du vecteur pET28a et plaçant le tag 6-Histidine (6-His) à l'extrémité C-terminale de la protéine (**Figure 17A**).

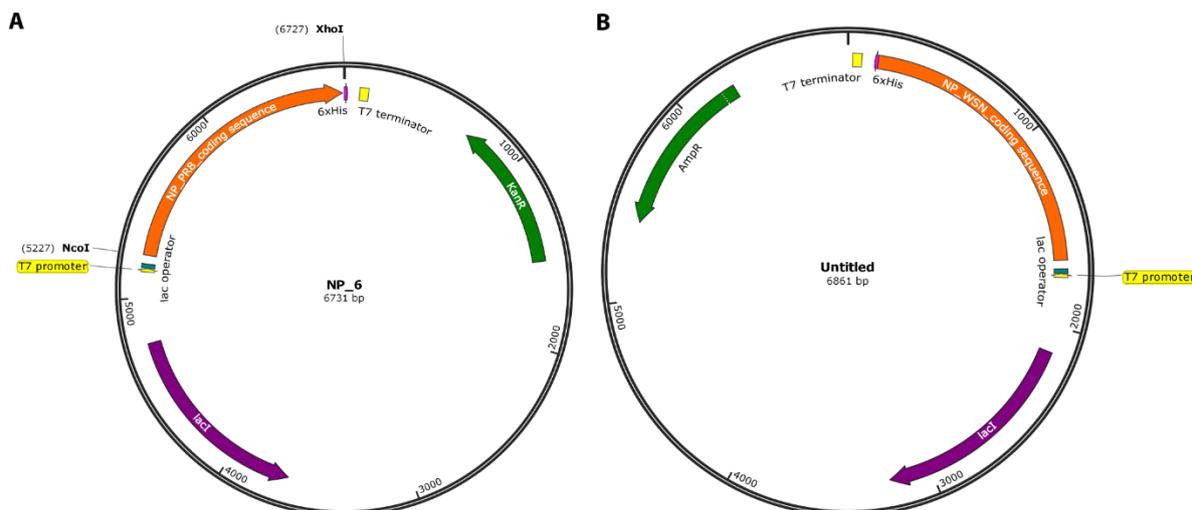


Figure 17 : Cartes des plasmides (A) pET28a et (B) pET22b permettant l'expression dans *E. coli* des protéines NP des souches PR8 et WSN respectivement.

1. Induction de la protéine

1) Premier protocole d'induction de la protéine NP (souche PR8) :

Pour l'induction de l'expression de la protéine NP, des bactéries compétentes *E. coli* BL21 (DE3), exprimant constitutivement l'ARN polymérase T7, sont utilisées à partir d'un stock glycérol.

Protocole de création d'un stock glycérol :

50 ng de plasmide sont ajoutés à 50 μ l de bactéries compétentes et incubés 30 min sur la glace. Un choc thermique de 45 s à 42°C puis 2 min sur la glace est réalisé avant ajout de 200 μ l de LB (Roth) sans antibiotique et incubation 1 h à 37 °C. Les bactéries sont ensuite étalées sur boîte LB agar + antibiotique et incubées la nuit à 37 °C. Le lendemain, 1 ml de LB sans antibiotique est ajouté et la gélose est grattée pour récupérer les colonies présentes sur la boîte. Après ajout d'un volume de glycérol 100%, le stock glycérol qui est conservé à -80 °C.

La veille de l'induction, 100 ml de LB + kanamycine (30 mg/ml) sontensemencées avec le stock glycérol et incubés la nuit à 37 °C sous agitation. La DO₆₀₀ de cette pré-culture est mesurée et deux flasques contenant 2 l de LB + kanamycine (30 mg/ml) sontensemencées à une DO₆₀₀ théorique de 0,05. Les cultures sont incubées à 37 °C sous agitation jusqu'à ce que la DO₆₀₀ soit comprise entre 0,7 et 0,9, puis induites avec 0,5 mM d'IPTG (Euromedex). L'induction est réalisée à 18 °C, sous agitation, durant

toute la nuit. Les cultures sont centrifugées le lendemain pendant 15 min à 4000 g et les culots bactériens sont utilisés pour la suite du protocole.

2) Protocole final d'induction de la protéine NP (souche WSN) :

Pour l'induction de l'expression de la protéine NP, des bactéries compétentes *E. coli* BL21 (DE3) pLysS, exprimant constitutivement l'ARN polymérase T7, sont utilisées à partir d'un stock glycérol.

La veille de l'induction 100 ml de LB + kanamycine (30 mg/ml) + chloramphénicol (34 mg/ml) sontensemencés avec le stock glycérol et incubés sur la nuit à 37 °C sous agitation. La DO₆₀₀ de cette pré-culture est mesurée et deux flasques contenant 2 l de LB + kanamycine (30 mg/ml) + chloramphénicol (34 mg/ml) sontensemencées à une DO₆₀₀ théorique de 0,05. Les cultures sont incubées à 37 °C sous agitation jusqu'à ce que la DO₆₀₀ soit comprise entre 0,7 et 0,9 puis plongées dans un bain eau-glace et agitées jusqu'à ce que la température du milieu atteigne 18°C. Les flasques sont ensuite incubées à 18°C sous agitation pendant 30 min avant d'induire la production de la protéine avec 1 mM d'IPTG (Euromedex). L'induction est réalisée à 18°C, sous agitation, durant toute la nuit. Les cultures sont centrifugées le lendemain pendant 15 min à 4000 g et les culots bactériens sont utilisés pour la suite du protocole.

2. Purification de la protéine

1) Premier protocole de purification de la protéine NP (souche PR8) :

Les culots bactériens sont repris dans 60 ml de tampon de lyse (50 mM Tris-HCl pH 7,5 à 20 °C ; 300 mM NaCl ; 5 mM β-mercaptoethanol) et lysés par « french press » en présence de 25 µg de lysozyme. Le lysat est centrifugé une première fois à 11000 g pendant 15 min (4 °C) puis ultra-centrifugé à 160000 g pendant 1 h (rotor SW-32Ti, Beckman Coulter). Le surnageant est ensuite passé à un débit de 1 ml/min sur une colonne « His-Trap™ HP » (GE Healthcare) d'un volume d'1 ml (0,7 x 2,5 cm), branchée à un système ÄKTAprime plus (GE Healthcare) et équilibrée avec le tampon de lyse. La colonne est ensuite lavée deux fois avec le tampon de lyse contenant 20 mM d'imidazole (Merck) avant d'éluer les protéines avec un tampon fortement concentré en imidazole (50 mM Tris-HCl pH 7,5 à 20 °C ; 300 mM NaCl ; 5 mM β-mercaptoethanol ; 800 mM imidazole). Les fractions (1 ml) contenant des protéines (identifiées par absorbance à 280 nm) sont ensuite passées à un débit de 2 ml/min sur

colonne « HiTrap™ Heparin HP » (GE Healthcare) d'un volume d'1 ml (0,7 x 2,5 cm), branchée à un système ÄKTApurifier (GE Healthcare) et équilibrée avec un tampon composé de 50 mM Tris-HCl pH 7,5 à 20 °C ; 300 mM NaCl et 5 mM β-mercaptoethanol. L'élution est réalisée par un gradient en sel, en passant du tampon A (50 mM Tris-HCl pH 7,5 à 20 °C ; 300 mM NaCl et 5 mM β-mercaptoethanol) au tampon B (50 mM Tris-HCl pH 7,5 à 20 °C ; 2 M NaCl et 5 mM β-mercaptoethanol) sur un volume de 40 ml. Les fractions (1 ml) contenant des protéines (identifiées par absorbance à 280 nm) sont collectées et déposées sur gel SDS-PAGE 12,5 %. Après révélation au bleu de Coomassie les fractions contenant la protéine d'intérêt (bande visible à 55 kDa) sont regroupées et dialysées contre un tampon trimère (50 mM Tris-HCl pH 7,5 à 20 °C ; 150 mM NaCl et 5 mM β-mercaptoethanol) sur la nuit, sous agitation à 4°C.

2) Protocole final de purification de la protéine NP (souche WSN) :

Les culots bactériens sont repris dans 80 ml de tampon de lyse (50 mM Tris-HCl pH 7,5 à 20 °C, ; 300 mM NaCl ; 5 mM β-mercaptoethanol ; 1 M NDSB) et lysés par sonication puis le lysat est centrifugé à 11000 g pendant 1 h (4 °C). Le surnageant est ensuite passé à un débit de 2 ml/min sur colonne « His-Trap™ FF crude » (GE Healthcare) d'un volume de 5 ml (1,6 x 2,5 cm) branchée à un système ÄKTAprime plus (GE Healthcare) et équilibrée avec un tampon composé de 20 mM Tris-HCl pH 7,5 à 20 °C ; 300 mM NaCl ; 5 mM β-mercaptoethanol ; 50 mM imidazole. La colonne est ensuite lavée avec un tampon à haut sel (20 mM Tris-HCl pH 7,5 à 20 °C ; 1 M NaCl ; 5 mM β-mercaptoethanol ; 50 mM imidazole) puis un tampon à bas sel (20 mM Tris-HCl pH 7,5 à 20 °C ; 100 mM NaCl ; 5 mM β-mercaptoethanol ; 50 mM imidazole) avant d'éluer les protéines avec un tampon fortement concentré en imidazole (50 mM Tris-HCl pH 7,5 à 20 °C ; 100 mM NaCl ; 5 mM β-mercaptoethanol ; 800 mM imidazole). Les fractions (1 ml) contenant des protéines (identifiées par absorbance à 280 nm) sont ensuite passées à un débit de 2 ml/min sur colonne « HiTrap™ Heparin HP » (GE Healthcare) d'un volume de 5 ml (1,6 x 2,5 cm), branchée à un système ÄKTApurifier (GE Healthcare) et équilibrée avec un tampon composé de 50 mM Tris-HCl pH 7,5 à 20 °C ; 100 mM NaCl et 5 mM β-mercaptoethanol. L'élution est réalisée par un gradient en sel, en passant du tampon A (50 mM Tris-HCl pH 7,5 à 20 °C ; 100 mM NaCl et 5 mM β-mercaptoethanol) au tampon B (50 mM Tris-HCl pH 7,5 à 20 °C ; 1,5 M NaCl et 5 mM β-mercaptoethanol) sur un volume de 30 ml. Les fractions (1 ml)

contenant des protéines sont collectées et déposées sur gel SDS-PAGE 12,5 %. Après révélation au bleu de Coomassie les fractions contenant la protéine d'intérêt (bande visible à 55 kDa) sont regroupées et dialysées contre un tampon monomère (50 mM Tris-HCl pH 7,5 à 20 °C ; 50 mM NaCl et 5 mM β -mercaptoéthanol) ou un tampon trimère (50 mM Tris-HCl pH 7,5 à 20 °C ; 150 mM NaCl et 5 mM β -mercaptoéthanol) sur la nuit, sous agitation à 4°C (**Figure 18**).

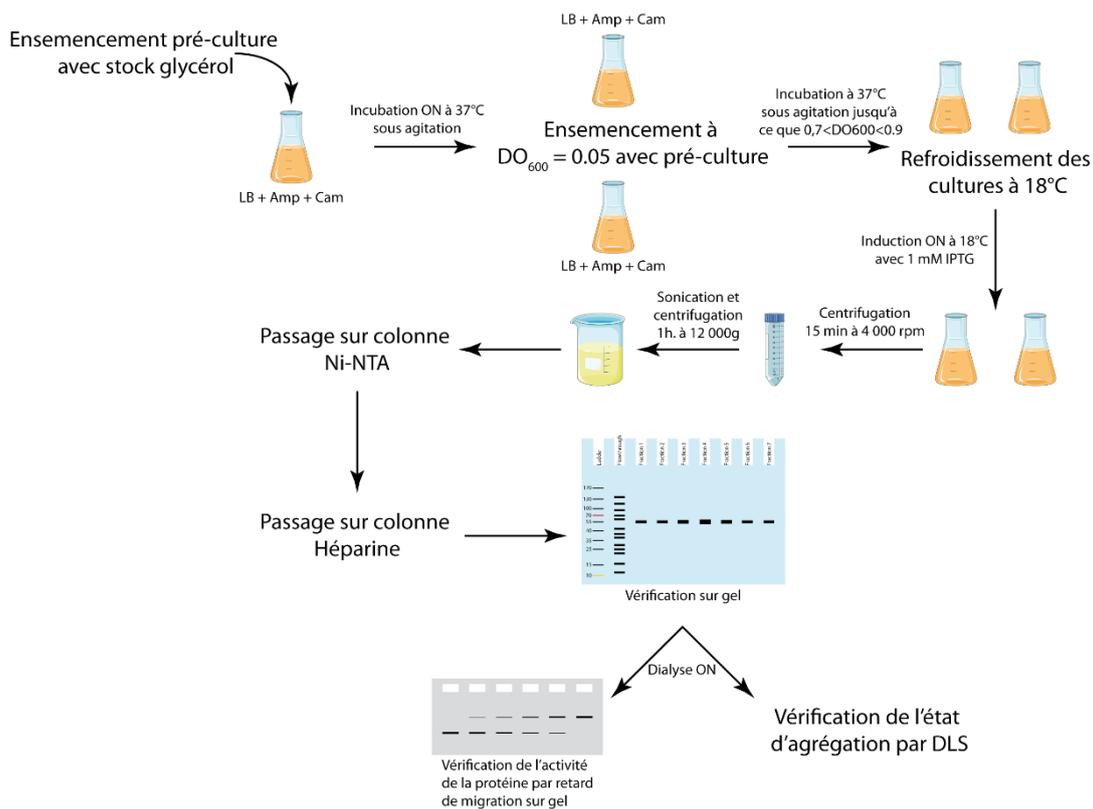


Figure 18 : Stratégie expérimentale d'induction et de purification de la protéine NP des IVA.

3) Suite du protocole après dialyse :

Les échantillons sont centrifugés 10 min à 21000 g (4 °C) et la moitié du volume est prélevée et mise dans un nouveau tube afin d'éliminer les agrégats. Un test de Bradford est effectué pour déterminer la concentration en protéine des échantillons et l'état d'oligomérisation de la protéine NP est vérifié en DLS (*dynamic light scattering*) sur un Zetasizer Nano S (Malvern Panalytical), en utilisant 50 μ l d'échantillon à 20 °C. Dans les cas où une trop forte agrégation est observée, les échantillons sont ultra-centrifugés 1 h à 100000 g (rotor S45A, Hitachi) et la moitié du volume est prélevée et à nouveau analysée en DLS.

C. Retard de migration sur gel

L'activité de fixation de la protéine à l'ARN est testée par retard de migration sur gel d'agarose en conditions natives. Pour cela, 1 pmole d'ARN NS est placée 2 min à 90 °C puis 2 min sur la glace avant d'être incubée 15 min à 20 °C dans 50 mM Tris pH 7,5 à 20 °C ; 50 mM NaCl si la NP monomérique est utilisée ou dans 50 mM Tris pH 7,5 à 20 °C ; 150 mM NaCl si la NP trimérique est utilisée, dans un volume final de 15 µl. Pendant ce temps, la protéine NP est incubée 20 min à 37 °C avant d'être ajoutée à des concentrations croissantes à l'ARN, le ratio NP/nucléotide allant de 1/100 à 1/10. L'ensemble est incubé 30 min à 37 °C puis déposé sur gel agarose 0,8 % dans du TBM (TB 0,5X + 1 mM MgCl₂) et la migration est réalisée pendant 2 h 30 min à 150 V et à 4 °C.

D. SHAPE *in vitro*

1. Etape de modification :

La protéine NP utilisée pour les expériences de SHAPE *in vitro* est sous forme trimérique ou monomérique afin de déterminer les différences de fixation dans ces deux conditions. Les tampons de repliement de l'ARN utilisés sont différents en fonction de l'utilisation d'une protéine monomérique (50 mM Tris-HCl pH 7,5 à 20 °C ; 50 mM NaCl) ou trimérique (50 mM Tris-HCl pH 7,5 à 20 °C ; 150 mM NaCl). Quatre conditions sont utilisées pour les expériences de SHAPE *in vitro* :

- **Neg :** *Témoin négatif, l'ARN non complexé à la NP et non modifié.* Pour les ARN NS et M, 3 et 4 pmoles d'ARN sont utilisées, respectivement, et placées 2 min à 90 °C puis 2 min sur la glace avant d'être incubées 15 min à 20 °C dans le tampon de repliement. Un volume de DMSO (diméthylsulfoxyde) anhydre équivalent à celui de NMIA (N-méthylisatoic anhydride) utilisé dans les échantillons modifiés est ajouté et l'échantillon est incubé 50 min à 20 °C. Une précipitation est réalisée en utilisant du NaCl 4 M à la place de l'acétate de sodium 3 M.
- **NoNP :** *ARN nu, l'ARN non complexé à la NP mais modifié.* Pour les ARN NS et M, 3 et 4 pmoles d'ARN sont utilisées, respectivement, et placées 2 min à 90 °C puis 2 min sur la glace avant d'être incubées 15 min à 20 °C dans le tampon

de repliement. L'ARN est ensuite modifié avec une concentration finale de 20 mM de NMIA durant 50 min à 20 °C.

- **PK :** *ARN complexé à la NP, puis déprotéinisé et modifié.* Pour les ARN NS et M, 3 et 4 pmoles d'ARN sont utilisées, respectivement, et placées 2 min à 90 °C puis 2 min sur la glace avant d'être incubées 15 min à 20 °C dans le tampon de repliement pendant que la NP (monomérique ou trimérique) est incubée 20 min à 37 °C. Les complexes ARN NP sont formés 30 min à 37 °C par ajout de la NP à l'ARN à un ratio de 1 protéine pour 20 nucléotides. Le complexe est ensuite traité avec 4 U de Protéinase K pendant 1 h à 37°C. Une extraction phénol-chloroforme (v/v) est réalisée suivie d'une extraction chloroforme (v/v), puis d'une précipitation en utilisant du NaCl 4 M au lieu d'acétate de sodium 3 M. Les culots sont repris dans le tampon de repliement et l'ARN est ensuite modifié avec une concentration finale de 20 mM de NMIA durant 50 min à 20 °C.
- **NP :** *ARN complexé à la NP, puis modifié et déprotéinisé.* Pour les ARN NS et M, 3 et 4 pmoles d'ARN sont utilisées, respectivement, et placées 2 min à 90 °C puis 2 min sur la glace avant d'être incubées 15 min à 20 °C dans le tampon de repliement pendant que la NP (monomérique ou trimérique) est incubée 20 min à 37 °C. Les complexes ARN NP sont formés 30 min à 37°C par ajout de la NP à l'ARN à un ratio de 1 protéine pour 20 nucléotides. Le complexe est ensuite modifié avec une concentration finale de 20 mM de NMIA puis précipité en utilisant du NaCl 4 M. Les culots sont repris dans le tampon de repliement et le complexe est traité avec 4 U de Protéinase K pendant 1 h à 37 °C. Une extraction phénol-chloroforme (v/v) est réalisée, suivie d'une extraction chloroforme (v/v) puis d'une précipitation en utilisant du NaCl 4 M.

Après précipitation, pour chaque échantillon, les culots sont repris dans 27 ou 36 µl d'eau MilliQ (pour les ARN NS et M, respectivement).

2. Etape de transcription inverse :

Afin de couvrir la totalité des ARN NS et M, plusieurs amorces fluorescentes sont utilisées. Pour chaque amorce, 9 µl d'échantillon modifié sont utilisés pour l'étape de transcription inverse, auxquels sont ajoutés 10 pmoles d'amorce portant en 5' le fluorophore Vic (absorbance maximale à 538 nm et émission maximale à 554 nm), puis incubés 2 min à 90 °C et 2 min sur la glace. La réaction est ensuite réalisée dans un volume final de 20 µl contenant 2 mM de chaque dNTP, 2 U de la rétrotranscriptase

d'AMV (*Avian Myeloblastosis Virus*) et 2 µl de tampon 10X (250 mM Tris-HCl pH 8,3 ; 500 mM KCl ; 2 mM DTT ; 50 mM MgCl₂) du fournisseur (Life Sciences) et incubée 20 min à 42 °C puis 10 min à 50 °C et 30 min à 60 °C. Une extraction phénol-chloroforme (v/v) est réalisée, suivie d'une extraction chloroforme (v/v).

En parallèle de l'extension des échantillons modifiés, une piste de séquençage est réalisée en utilisant de l'ARN non modifié en présence de ddGTP. Pour cela, 1 pmole d'ARN non modifié est incubée 2 min à 90 °C et 2 min sur la glace avec 10 pmoles d'amorce portant en 5' le fluorophore Ned (absorbance maximale à 546 nm et émission maximale à 575 nm). La réaction est ensuite réalisée dans un volume final de 20 µl contenant 0,5 mM de dGTP et 2 mM de chaque dNTP restant ; 5 mM de ddGTP ; 2U de la rétrotranscriptase d'AMV et 2 µl de tampon 10X (250 mM Tris-HCl pH 8,3 ; 500 mM KCl ; 2 mM DTT ; 50 mM MgCl₂) du fournisseur (Life Sciences) et incubée 20 min à 42 °C puis 10 min à 50 °C et 30 min à 60 °C. Une extraction phénol-chloroforme (v/v) est réalisée, suivie d'une extraction chloroforme (v/v).

Les deux échantillons (piste séquençage et échantillon modifié) sont réunis dans le même tube et mis à précipiter et les culots sont repris dans 10 µl de *highly deionized* formamide (Applied Biosystem), permettant de garder les ADNc en conditions dénaturantes. Les échantillons sont ensuite incubés 5 min à 90 °C puis 5 min sur la glace avant d'être centrifugés 15 min à 21000 g (4°C) puis chargés sur séquenceur capillaire 3130xl Genetic Analyzer (Applied Biosystems).

E. Marquage radioactif d'amorces

Pour la mise au point de certaines étapes de SHAPE *in viro* des amorces radioactives sont utilisées afin de visualiser sur gel les produits générés au cours de ces étapes.

Dans un volume final de 20 µl, 10 pmoles d'amorce sont marquées radioactivement avec 100 µCi de γ -³²P-ATP, 15 U d'enzyme T4 Polynucleotide Kinase (PNK) dans le tampon A (dans le cas où les amorces ne portent pas de 5'-P ; 50 mM Tris-HCl pH 7,6 à 25 °C ; 10 mM MgCl₂ ; 5 mM DTT ; 0,1 mM spermidine) ou B (dans le cas où les amorces portent un 5'-P ; 50 mM imidazole-HCl pH 6,4 à 25 °C ; 18 mM MgCl₂ ; 5 mM DTT ; 0,1 mM spermidine ; 0,1 mM ADP). La réaction est incubée 1 h à 37 °C et les amorces sont purifiées sur deux colonnes successives Micro Bio-Spin® (Bio-Rad).

Avant purification les colonnes sont centrifugées 2 min à 1000 g pour éliminer le tampon de stockage, puis les échantillons sont chargés et centrifugés 4 min à 1000 g. L'efficacité de marquage est mesurée par comptage en mode Tcherenkov.

F. SHAPE *in viro*

1. Production et purification des particules virales

Neuf flasques T175 de cellules MDCK-II à confluence sont infectées avec le virus A/SC35M à une MOI (multiplicity of infection) de 0,001 PFU (plaque forming unit) par cellule. Après 24 h, un total de 200 ml de milieu de culture est collecté et centrifugé 30 min à 2000 g à 4°C. Le surnageant est récupéré et séparé en six aliquotes de 33 ml qui sont chargées délicatement sur 5 ml de coussin de sucrose (30% dans le tampon NTC contenant 100 mM NaCl ; 20 mM Tris-HCl pH 7,4 et 5 mM CaCl₂) dans six tubes en polypropylène (Beckman Coulter, ref #344058). Les particules virales sont purifiées par ultracentrifugation pendant 90 min à 112000 g à 4 °C avec un rotor SX32 (Beckman Coulter). Les culots sont repris dans un total de 2,5 ml de tampon NTC et séparés en deux aliquotes chargées sur deux tubes en polypropylène (Beckman Coulter, ref #344059) contenant un gradient OptiPrep (Sigma, D1556) préparé par superposition des couches suivantes : 40 %, 1.5 ml ; 35 %, 1.5 ml ; 30 %, 1.5 ml ; 20 %, 3 ml ; 15 %, 1 ml ; 10 %, 1 ml dans le tampon NTC. Les particules virales sont séparées par ultracentrifugation pendant 150 min à 209000 g à 4 °C avec un rotor SW41Ti (Beckman Coulter). La bande contenant les particules virales, visible entre les concentrations en OptiPrep de 20 % et 30 %, est collectée par le côté du tube grâce à une seringue et les échantillons sont regroupés.

2. Modification de l'ARN

La moitié des particules virales purifiées sont stockées dans le tampon de modification composé de 50 mM HEPES pH 8 ; 200 mM NaCl ; 0,1 mM EDTA et 10 % sérum foetal bovin. Le reste de l'échantillon est traité 15 min à 30 °C avec le tampon de désassemblage composé de 1 % Triton X-100 (p/v) ; 1 mg/ml lysolécithine ; 5 mM MgCl₂ ; 100 mM KCl ; 1,5 mM DTT ; 5 % glycérol et 10 mM Tris-HCl pH 8. Après incubation, du sérum foetal de bœuf est ajouté à une concentration finale de 10 %.

Les particules virales intactes et désassemblées sont ensuite réparties en trois tubes correspondant à trois conditions de modification : un contrôle non modifié, et deux échantillons modifiés avec des concentrations en réactif chimique différentes. Pour le protocole d'icSHAPE, les échantillons sont modifiés 75 min à 37 °C avec des concentrations finales de 3,9 et 7,8 mM de Ai-N₃ (N-azidobutyl-anhydride isatoïque), tandis que pour le SHAPE-MaP ce sont des concentrations finales de 10 et 30 mM de 1M7 (synthétisé au laboratoire par le Dr Valérie Vivet-Boud) qui sont utilisées pendant 2 min à 37 °C. L'ARN est ensuite extrait à l'aide du kit Direct-zol™ RNA Miniprep (Zymo Research) puis précipité avec 0,1 volume d'acétate de sodium 3 M, 3 volumes d'EtOH 100 % et 1 µg de glycogène et conservé à -20°C jusqu'à utilisation.

3. Protocole d'icSHAPE :

Ce protocole est adapté du protocole publié en 2015 par Spitale *et coll.* [235]. Dans ce protocole, différents adaptateurs sont ligués en 3' et en 5' des échantillons. Les adaptateurs 3' possèdent un barcode unique permettant la différenciation des échantillons traités. Les adaptateurs 5' ont tous la même séquence mais ceux qui sont utilisés pour les échantillons non modifiés possèdent un groupement amine à l'extrémité 5'. Ce dernier permet l'ajout d'une biotine par réaction chimique, tandis que pour les échantillons modifiés la biotine sera ajoutée par click-chemistry sur la modification apportée par le réactif Ai-N₃.

Les échantillons sont centrifugés 45 min à 21000 g (4 °C), les culots sont lavés deux fois avec de l'éthanol 75 % refroidi à -20 °C (5 min de centrifugation à 21000 g, 4 °C, entre chaque lavage) puis séchés 5 min à l'air avant d'être repris dans 8 µl d'eau MilliQ et incubés 1 min à 90 °C puis 2 min sur la glace. La fragmentation des ARN est réalisée avec 1 µl de réactif de fragmentation 5X (Ambion life technologies) pendant 1,5 min à 80 °C, puis 1 µl de solution stop 5X (Ambion life technologies) est ajouté pour arrêter la réaction (les fragments obtenus présentent des extrémités 5'OH et 3'P). Une purification des échantillons est réalisée avec le kit RNeasy Mini (Qiagen) en ajoutant 35 µl de tampon RLT et 100 µl d'EtOH 100 % avant de suivre le protocole du fournisseur. L'élution est réalisée dans 2 x 50 µl et le volume est réduit à 6 µl par lyophilisation. La déphosphorylation de l'extrémité 3' des échantillons est réalisée dans un volume final de 10 µl avec 40 U de RNasin, 10 U de T4 PNK (ThermoFisher Scientific) et 2 µl de tampon PNK 5X (350 mM Tris-HCl pH 6,8 ; 50 mM MgCl₂ ; 25 mM DTT) et incubée 1 h à 37 °C avant d'ajouter l'adaptateur 3' par ligation.

Protocole de ligation :

Aux 10 µl d'échantillon sont ajoutés 50 pmoles d'adaptateur, 1 mM d'ATP, 1 mM de DTT ; 2,5 µl de tampon 10X (500 mM Tris-HCl pH 7,5 à 25 °C ; 100 mM MgCl₂ ; 10 mM DTT – NEB), 8 µl de PEG8000 50 % (NEB) et 15 U de T₄ RNA ligase I (NEB). La réaction est incubée à 16 °C durant nuit.

Après ajout de 25 µl d'eau MilliQ, 185 µl de tampon RLT et 400 µl d'EtOH 100 %, une purification des échantillons ligés est réalisée à l'aide du kit RNeasy Mini (Qiagen) en suivant le protocole du fournisseur. L'élution est réalisée dans 2 x 50 µl et le volume est réduit jusqu'à 7 µl par lyophilisation. Une étape de phosphorylation des extrémités 5' des fragments est ensuite réalisée dans un volume de 10 µl avec 1 mM d'ATP, 10 U de T₄ PNK et 1 µl de tampon A PNK 10X (500 mM Tris-HCl pH 7,6 à 25 °C ; 100 mM MgCl₂ ; 50 mM DTT ; 1 mM spermidine). L'incubation est réalisée 1 h à 37 °C puis 20 min à 65 °C avant de réaliser la ligation d'un adaptateur en 5' sur ces échantillons. Ces derniers sont purifiés à l'aide du kit RNeasy Mini (Qiagen) en suivant le protocole du fournisseur après ajout de 25 µl d'eau MilliQ, 185 µl de tampon RLT et 400 µl d'EtOH 100 %. L'élution est réalisée dans 2 x 50 µl et le volume est réduit jusqu'à 35 µl par lyophilisation. La liaison de la biotine aux différents échantillons est réalisée selon deux protocoles différents :

- Pour les échantillons non modifiés : Ajout de 1 µl de EZ-Link™ Sulfo-NHS-SS-Biotin (ThermoFisher Scientific) à 1 mM aux 35 µl d'échantillon et incubation 30 min à 37 °C.
- Pour les échantillons modifiés : Ajout de 1 µl de DBCO-S-S PEG3 Biotin (Broadpharm) à 1 mM aux 35 µl d'échantillon et incubation 30 min à 37 °C.

Les échantillons sont ensuite purifiés sur deux colonnes consécutives Micro Bio-Spin®. Avant purification les colonnes sont centrifugées 2 min à 1000 g pour éliminer le tampon de stockage, puis les échantillons sont chargés et centrifugés 4 min à 1000 g. Le volume est réduit à 10 µl par lyophilisation. La sélection des fragments portant une biotine est réalisée sur billes streptavidine (MyOne C1 - Invitrogen). Pour chaque échantillon 10 µl de billes sont utilisées et lavées deux fois dans 200 µl de tampon A (0,1 M NaOH ; 0,05 M NaCl) puis deux fois dans 200 µl de tampon B (0,1 M NaCl) et enfin trois fois dans le tampon 1X (5 mM Tris-HCl pH 7,5 ; 0,5 mM EDTA ; 1 M NaCl). A chaque étape de lavage les billes sont incubées avec le tampon 2 min à température

ambiante puis posée 2 min sur un portoir magnétique avant d'éliminer le liquide. Les billes sont resuspendues dans 10 µl de tampon 2X (10 mM Tris-HCl pH 7,5 ; 1 mM EDTA ; 2 M NaCl) et les 10 µl d'échantillon sont ajoutés et incubés 30 min à 20 °C sous agitation pour éviter une sédimentation des billes trop importante. Les billes sont ensuite lavées 4 fois avec 200 µl de tampon 1X et l'éluion est réalisée 30 min à 37 °C avec 200 µl de DTT 50 mM fraîchement préparée. Les échantillons sont ensuite mis à précipiter et les culots sont repris dans 11 µl d'eau milliQ.

Les échantillons sont ensuite incubés 2 min à 90 °C puis 2 min sur glace avec 10 pmoles d'amorce de transcription inverse. L'étape d'extension est réalisée dans un volume final de 20 µl en présence de 5 mM DTT ; 0,5 mM de chaque dNTP ; 40 U de RNasin ; 200 U de SuperScript IV (Invitrogen) et 4 µl de tampon SSIV 5X (Invitrogen) et incubée 20 min à 42 °C, 30 min à 50 °C et 10 min à 60 °C. Les échantillons sont ensuite déposés sur gel polyacrylamide 6 % en conditions dénaturantes en parallèle d'un marqueur de taille des ARN (Riboruler Low Range RNA Ladder – ThermoFisher) et de l'amorce de transcription inverse seule. La migration est réalisée à 200 V jusqu'à ce que le bleu de bromophénol ait migré au milieu du gel et la révélation est réalisée par ombrage aux UV. Les parties du gel contenant les fragments d'une taille supérieures de 100 à 200 bases à celle de l'amorce de transcription inverse sont découpées et déposées dans un tube de 0,5 ml dont le fond est troué, lui-même disposé dans un tube 1,5 ml. Après 1 min de centrifugation à 12000 g, 300 µl de « crush soak buffer » (500 mM acétate d'ammonium ; 1 mM EDTA ; 0,1 % SDS) sont ajoutés et l'incubation est réalisée durant la nuit à 50°C. Le surnageant est ensuite récupéré et précipité et les culots sont repris dans 12 µl d'eau MilliQ.

La circularisation des échantillons est réalisée dans un volume final de 20 µl avec 2,5 mM MnCl₂ ; 1 M Bétaine ; 2 µl de tampon CircLigase II 10X (0,33 M Tris-acétate pH 7,5 ; 0,66 M acétate de potassium et 5 mM DTT – Epicentre) et 100 U de CircLigase II (Epicentre) pendant 120 min à 60 °C puis 10 min à 80 °C. Une quantification du matériel est réalisée par PCR quantitative (qPCR), dans un volume final de 15 µl avec 1 ou 5 µl d'échantillon, 3 µl de tampon HF 5X (ThermoFisher Scientific), 200 µM de chaque dNTP, 3 pmoles de chaque amorce, Evagreen 0,5 X et 0,3 U de polymérase Phusion (ThermoFisher Scientific). Après une étape de dénaturation de 30 s à 98 °C, 40 cycles d'amplification sont réalisés (15 s à 98 °C, 15 s à 55 °C et 30 s à 72 °C). Une

vérification de 5 µl de produit de qPCR est réalisée sur gel agarose 0,8 % TBE 1X avec une migration de 20 min à 135 V.

4. Protocole de SHAPE-MaP :

Ce protocole est adapté du protocole publié en 2015 par Smola *et coll.* [232]. Les échantillons sont centrifugés 45 min à 21000 g (4 °C), les culots sont lavés deux fois avec de l'éthanol 75 % à -20 °C (5 min de centrifugation à 21000 g, 4 °C, entre chaque lavage) puis séchés 5 min à l'air avant d'être repris dans 10 µl d'eau miliQ. A chaque échantillon sont ajoutés 1 µl de « random primer mix » (NEB) et l'incubation est réalisée 5 min à 65 °C puis 2 min sur la glace. Les échantillons sont ensuite incubés 2 min à 25 °C en présence du tampon MaP (50 mM Tris-HCl pH 8 ; 75 mM KCl ; 6 mM MnCl₂ ; 10 mM DTT et 0,5 mM de chaque NTPs) puis sont ajoutés 200 U de SuperScript II et la réaction d'élongation est réalisée à 25 °C pendant 10 min puis 3 h à 42 °C et enfin 15 min à 70 °C. Les échantillons sont purifiés sur colonnes Micro Bio-Spin® (Bio-Rad). Avant purification les colonnes sont centrifugées 2 min à 1000 g pour éliminer le tampon de stockage, puis les échantillons sont chargés et centrifugés 4 min à 1000 g.

Le volume est ajusté à 68 µl et le second brin ADN est synthétisé 2,5 h à 16 °C dans un volume final de 80 µl après ajout de 8 µl de NEBNext second strand synthesis reaction buffer 10X et 4 µl de NEBNext second strand synthesis enzyme mix. L'ADNdb est alors purifié à l'aide du kit « Gel and PCR Clean Up » selon le protocole du fournisseur (Macherey-Nagel). L'élution est réalisée dans 10 µl d'eau miliQ et 10 ng d'ADNdb sont fragmentés dans un volume final de 20 µl avec 10 µl de « Tagment DNA buffer » et 5 µl de « *Amplicon Tagment mix* » pendant 5 min à 55 °C. Les échantillons sont ensuite refroidis à 10 °C et une fois cette température atteinte 5 µl de « Neutralize Tagment buffer » sont ajoutés et les échantillons sont incubés 5 min à 20 °C.

L'ajout des index Illumina est réalisé par PCR, dans un volume final de 50 µl en utilisant les 25 µl d'échantillon précédemment fragmenté, 15 µl de « Nextera PCR master mix » et 5 µl de chaque index. Les échantillons sont ensuite incubés 3 min à 72 °C, 30 s à 95 °C puis suivent 14 cycles d'amplification (10 s à 95 °C, 30 s à 55 °C et 30 s à 72 °C) et sont finalement incubés 5 min à 72 °C. Les échantillons sont ensuite purifiés sur 75 µl de billes « NEBNext Sample Purification Beads ». Après 5 min d'incubation à température ambiante, les tubes sont placés 5 min sur portoir

magnétique et le surnageant est éliminé. Deux lavages de 30 s chacun avec 200 μ l d'EtOH 80 % à température ambiante, fraîchement préparé, sont réalisés et après le second lavage les billes sont séchées 5 min à l'air. L'élution est réalisée avec l'ajout de 17 μ l de tampon TE 0,1X et une incubation de 2 min à température ambiante après avoir mélangé énergiquement. Les tubes sont alors placés 5 min sur portoir magnétique et le surnageant est collecté et stocké dans un nouveau tube. La quantité de matériel est évaluée à l'aide d'une qPCR sur 1 μ l de chaque échantillon.

RÉSULTATS

Résultats

A. Production et purification de la protéine NP

1. Souche PR8

Les expériences de SHAPE *in vitro* devant initialement être réalisées sur les ARNv de la souche PR8, j'ai tenté de surproduire et purifier la protéine NP de cette même souche. Avant de purifier la protéine NP j'ai réalisé des tests d'induction à l'IPTG (**Figure 19**) en incubant les cultures bactériennes à différentes températures (18 °C et 37 °C) et pendant différents temps (**Figure 19A et B**). Ces premiers tests ont été réalisés avec des bactéries compétentes *E. coli* BL21 (DE3) et n'ont montré aucune induction de la production de la protéine NP dans ces conditions.

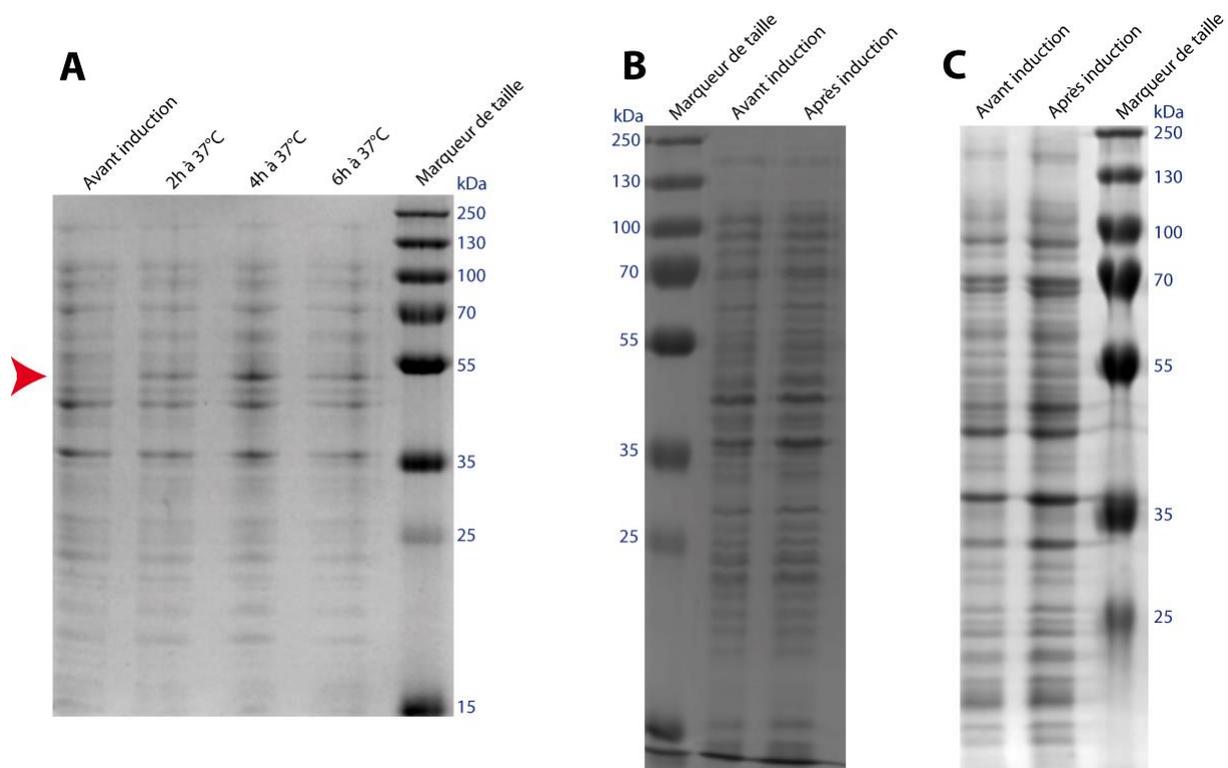


Figure 19 : Induction de la protéine NP PR8 sous différentes conditions. (**A et B**) La souche *E. coli* BL21 (DE3) est transformée avec le plasmide pET28a et l'induction est réalisée à 37 °C pendant 2, 4 et 6 h (**A**) ou à 18°C sur la nuit (**B**). (**C**) La souche *E. coli* BL21 (Rosetta2) pLysS est transformée avec le plasmide pET28a et l'induction est réalisée à 18 °C sur la nuit.

À 37 °C, une bande apparaît après induction dans toutes les conditions (**Figure 19A**, flèche rouge), mais la taille ne correspond pas à la protéine NP (56 kDa) et l'intensité est faible contrairement à ce qui est attendu lors de l'induction de la production d'une protéine dans *E. coli*. À 18 °C aucune différence n'est observée avant et après induction (**Figure 19B**). La séquence codante de la protéine NP clonée dans le plasmide pET28a n'étant pas optimisée pour la production en bactérie, il est possible que l'absence d'induction soit liée à un blocage de la traduction en raison de la présence de codons rares chez la bactérie. Pour vérifier cela, j'ai utilisé des bactéries *E. coli* BL21 (Rosetta2) pLysS, qui expriment les ARNt permettant de décoder les codons rares. Après 4 h d'induction à 37 °C avec 0,5 mM d'IPTG, aucune induction n'est détectée sur gel (**Figure 19C**).

Cependant, le fait que l'induction ne soit pas visible sur gel n'indique pas nécessairement une absence de production de la protéine par les bactéries. J'ai donc réalisé la purification de la protéine NP après une induction avec 0,5 mM d'IPTG à 18 °C sur la nuit. Ces conditions ont été choisies, car récemment publiées dans la littérature pour la purification de la protéine NP de la souche WSN [125]. La première étape de la purification (**Figure 20**) est réalisée par chromatographie d'affinité de type IMAC (*immobilized metal affinity chromatography*), consistant en l'immobilisation d'un atome métallique, ici le nickel, qui permet la rétention de certains polypeptides, ici le tag constitué de six résidus histidines. Des aliquotes sont prélevées à chaque étape afin de contrôler l'évolution de la purification sur gel SDS-PAGE. On observe sur le spectre UV_{280nm} obtenu au cours du passage sur la colonne (**Figure 20A**) et sur le gel (**Figure 20B**) qu'une quantité importante de protéine n'est pas retenue et se retrouve dans la fraction « *flowthrough* ». Deux lavages avec deux concentrations salines différentes (1 M NaCl puis 150 mM NaCl) permettent d'éliminer les protéines fixées de façon non spécifiques à la colonne, tandis que le dernier pic sur le spectre UV_{280nm} correspond à l'élution spécifique des protéines possédant une étiquette His₆ de la colonne (**Figure 20A**). On observe sur gel des bandes d'une taille attendue de 56 kDa dans les fractions d'élution, bien que de nombreux contaminants soient présents (**Figure 20B**).

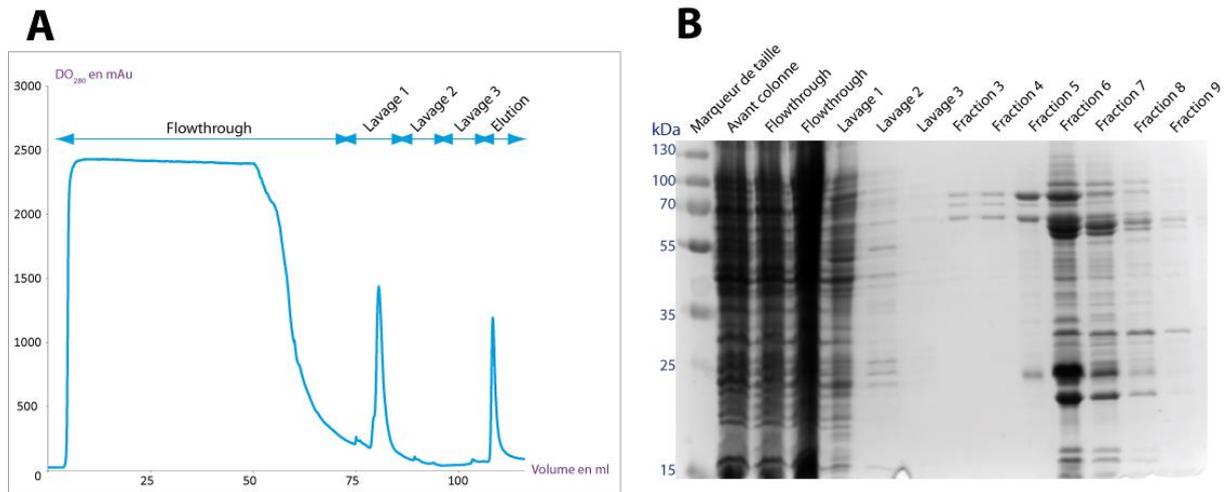


Figure 20 : Première étape de purification de la protéine NP PR8 sur colonne Ni-NTA. **(A)** Spectre UV_{280nm} en fonction du volume passé sur la colonne obtenu au cours de la purification par le système AKTA. Les différentes étapes sont indiquées et l'éluion comprend les fractions 1 à 15. **(B)** Vérification de la purification sur gel SDS-PAGE. A chaque étape une aliquote a été prélevée et analysée.

J'ai ensuite réalisé une seconde étape de purification afin d'éliminer le plus de contaminants possible (**Figure 21**). Cette étape est réalisée sur une colonne Héparine, qui retient les protéines capables de fixer les acides nucléiques. Le spectre UV_{280nm} montre trois pics majeurs au cours du gradient de sel permettant l'éluion (**Figure 21A**). Toutes les fractions correspondant à ces pics sont analysées sur gel SDS-PAGE (**Figure 21B et C**). On observe que les deux premiers pics correspondent à des contaminants (**Figure 21B**), alors que le troisième contient plusieurs protéines d'une taille proche de 56 kDa (**Figure 21C**).

Pour confirmer que la protéine purifiée est bien la NP, ainsi que pour identifier les contaminants présents dans les fractions d'intérêt, une analyse par spectrométrie de masse (MALDI-TOF) est réalisée, directement sur les bandes dans les pistes d'éluion, par la plateforme de protéomique de l'institut sur Autoflex III (Bruker). Les résultats indiquent que les trois bandes migrant autour de 56 kDa correspondent toutes à la protéine NP de la souche PR8, ce qui laisse supposer qu'une dégradation de la protéine a lieu. Cette dégradation a vraisemblablement eu lieu avant la lyse des bactéries étant donné que des inhibiteurs de protéase sont utilisés au cours de la purification. De plus, c'est vraisemblablement la partie N-terminale qui est dégradée, puisque ces protéines ont été retenues sur la colonne Ni-NTA, et que l'étiquette His₆ est localisée à l'extrémité C-terminale de la protéine.

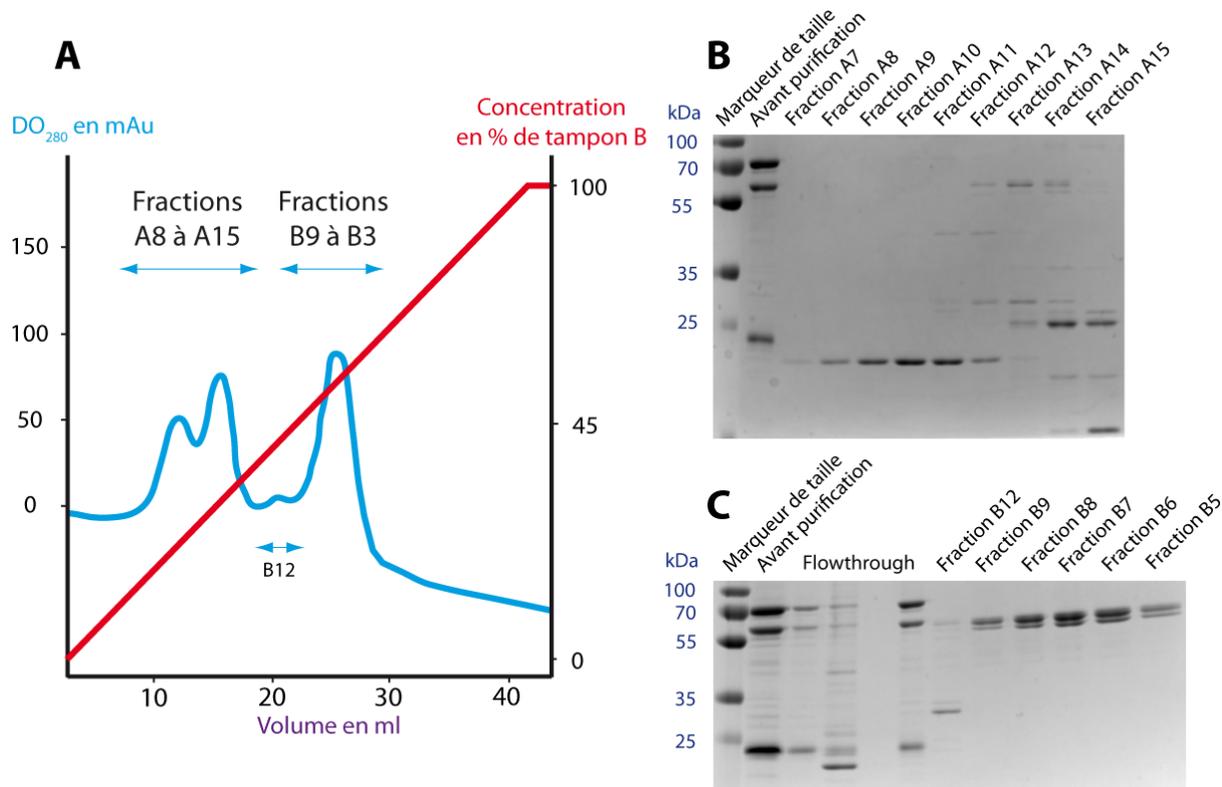


Figure 21 : Seconde étape de purification de la protéine NP PR8 sur colonne Héparine. **(A)** Spectre UV_{280nm} (en bleu) et gradient de tampon B (en rouge) en fonction du volume passé sur la colonne obtenus au cours de la purification sur système AKTA. **(B et C)** Analyse des fractions sur gel SDS-PAGE.

Les fractions contenant les trois bandes correspondant à la NP PR8 sont regroupées et dialysées contre un tampon favorisant la trimérisation de la protéine. L'état d'agrégation après l'étape de dialyse est vérifié par DLS (**Figure 22**). L'allure du corrélogramme, représentant la superposition de dix mesures, ne montre pas de signe d'agrégation de la protéine (**Figure 22A**). Dans le cas contraire, le corrélogramme met plus de temps à revenir à la ligne de base. La distribution en intensité (**Figure 22C**) est bimodale, avec un premier pic monodisperse présentant un rayon hydrodynamique d'environ 4,5 nm et un second pic polydisperse présentant un rayon hydrodynamique d'environ 200 nm. La distribution en nombre (**Figure 22B**) est unimodale et monodisperse avec un rayon hydrodynamique de 3,25 nm. Ces résultats indiquent que le pic le plus important obtenu dans la distribution en intensité ne représente que peu de molécules qui diffusent fortement. D'après le calculateur intégré au logiciel Zetasizer, les molécules analysées ont une taille d'environ 53 kDa soit une protéine monomérique. Ces résultats sont inattendus, puisqu'on s'attendait à observer une protéine majoritairement trimérique dans ces conditions, mais la NP sous forme

monomérique pouvant également fixer l'ARN [242] ce résultat n'empêche pas la poursuite de l'étude.

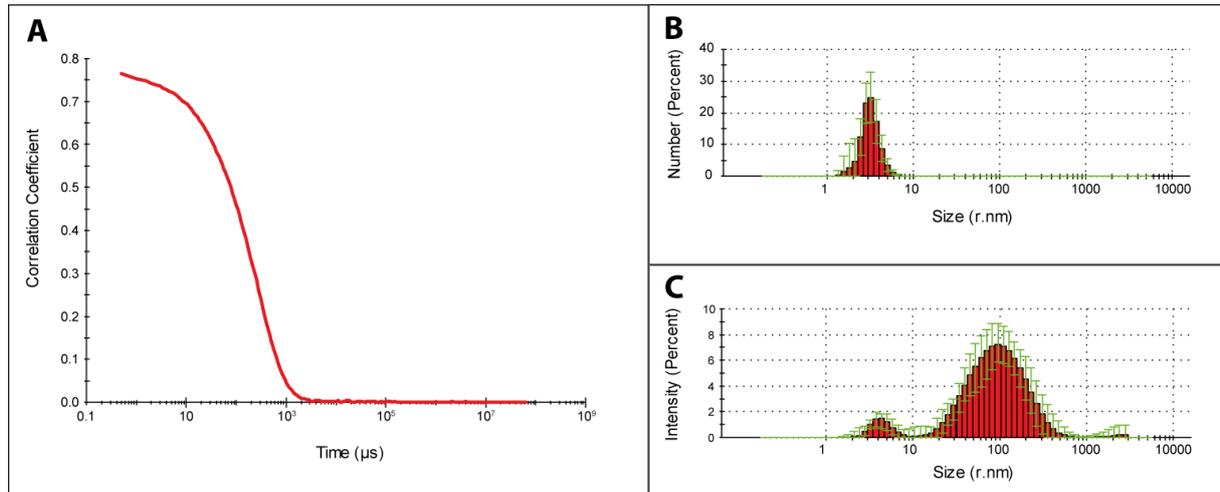


Figure 22 : Analyse par DLS de la protéine NP PR8 purifiée et dialysée contre 50 mM Tris-HCl pH7.5, 300 mM NaCl et 5 mM β -mercaptoethanol. (A) Fonction d'autocorrélation. (B) Distribution en nombre et (C) en intensité. Représentation de la moyenne de dix mesures consécutives.

J'ai ensuite testé la fixation de la protéine NP PR8 à l'ARNv NS de la même souche à différents ratios allant d'une protéine pour cent nucléotides à une protéine pour dix nucléotides par retard de migration sur gel électrophorétique (**Figure 23**). On observe sur le gel que la piste contenant l'ARN seul contient deux bandes qui correspondent à un ARN monomérique (flèche rouge) ou dimérique (flèche verte). A partir du ratio 1/50 une nouvelle bande est visible et correspond à l'ARNv complexé à la NP (flèche orange). L'intensité de cette bande augmente avec la quantité de protéine ajoutée jusqu'au ratio 1/10 où elle disparaît. Jusqu'à ce ratio, une seule bande est observée, indiquant que la fixation de la NP à l'ARN est coopérative : il n'y a pas d'intermédiaire entre l'ARNv nu et le complexe ribonucléoprotéique qui semble donc avoir une stœchiométrie bien définie. De plus, à partir du ratio 1/40, on observe la disparition de la bande correspondant au dimère d'ARN et aucune nouvelle bande pouvant correspondre à un dimère de complexe n'apparaît, suggérant que les complexes ARN-NP identiques n'interagissent pas entre eux. Lors de l'empaquetage, les IVA doivent incorporer une copie de chaque RNPv dans les particules virales néoformées, il faut donc éviter que deux RNPv identiques interagissent entre elles pour ne pas empaqueter deux segments identiques. On observe également que la totalité de l'ARN est complexé avec la NP au ratio 1/10, où la fixation de la NP devient hétérogène avec la présence d'un retard de

migration diffus. Cela indique qu'à partir d'une certaine quantité de protéine NP, la fixation n'est plus régulée et devient aspécifique.

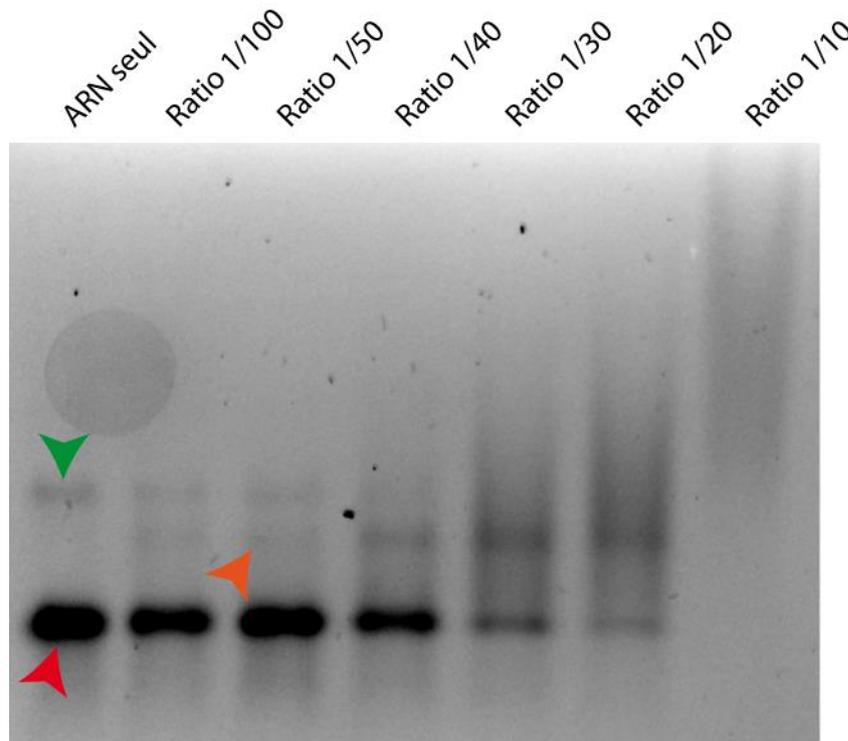


Figure 23 : Test de l'activité de la protéine NP PR8 par retard de migration sur gel. Chaque piste contient 50 nM d'ARN NS de la souche PR8. Les ratios indiquent le nombre de nucléotides pour lequel une protéine est ajoutée (ratio 1/50 : une protéine ajoutée pour 50 nucléotides). La révélation est réalisée par transillumination aux UV. Les flèches indiquent les trois états de l'ARN retrouvés.

À la suite de ces résultats, j'ai utilisé la protéine NP PR8 pour la cartographie chimique en solution par la méthode SHAPE (voir §I.F.2.), mais la faible quantité purifiée ainsi que la présence de trois bandes distinctes après purification reste un problème.

2. Souche WSN

Afin de contourner les problèmes liés à la production de la protéine NP PR8, j'ai décidé de changer la souche d'étude pour les expériences de SHAPE *in vitro*. Le choix de cette souche est motivé par plusieurs articles dans la littérature qui décrivent l'utilisation de la NP WSN. Par exemple, la première structure cristallographique publiée dans Nature en 2006 est celle de la protéine NP de la souche WSN [278]. Avant de purifier la protéine, j'ai réalisé un test d'induction (**Figure 24**) en utilisant les bactéries compétentes *E.coli* BL21 (DE3) ou BL21 (Rosetta2) pLysS. Dans les deux

conditions, l'induction a été réalisée à 18 °C durant la nuit avec une concentration de 1 mM d'IPTG. Contrairement à l'induction de la protéine NP PR8, pour la protéine NP WSN, la température du milieu de culture est refroidie à 18 °C avant ajout de l'IPTG. On observe sur les gels que l'induction dans les bactéries BL21 (DE3) n'est pas visible (**Figure 24A**), tandis qu'une bande de forte intensité est visible à la taille attendue de 56 kDa après induction des bactéries BL21 (Rosetta2) pLysS (**Figure 24B**).

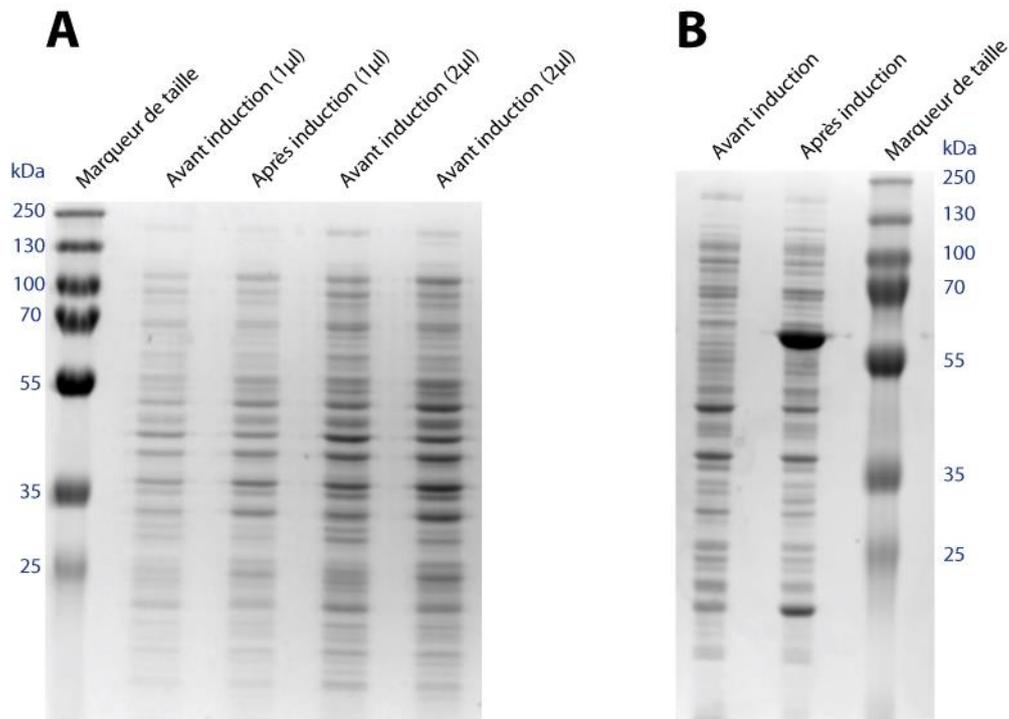


Figure 24 : Induction de la protéine NP WSN dans différentes conditions. (A) La souche *E. coli* BL21 (DE3) est transformée avec le plasmide pET22b et l'induction est réalisée à 18°C durant la nuit. Deux volumes différents d'échantillon sont déposés afin de mieux visualiser une faible induction. (B) La souche *E. coli* BL21 (Rosetta2) pLysS est transformée avec le plasmide pET22b et l'induction est réalisée à 18°C durant la nuit.

Après induction, les bactéries sont lysées et le surnageant est purifié par chromatographie d'affinité (**Figure 25**). Le spectre UV_{280nm} obtenu au cours de cette étape de purification est similaire à celui obtenu lors de la purification de la NP PR8 (**Figures 25A** et **20A**). Sur le gel de contrôle de la purification (**Figure 25B**), on observe qu'une forte quantité de protéine NP WSN n'est pas retenue sur la colonne et se retrouve dans la fraction *flowthrough*. Le repliement de ces protéines pourrait enfouir le tag His₆ au sein de sa structure le rendant inaccessible aux ions nickel de la colonne. L'interaction entre la NP et des protéines bactériennes pourrait également expliquer la présence de la protéine dans cette fraction *flowthrough*. De même, si la capacité maximale de fixation de la colonne est atteinte, le surplus de protéine ne sera

pas retenu sur la colonne et sera retrouvé dans cette fraction. Après élution, on remarque sur gel que très peu de contaminants sont présents contrairement à ce qui a été observé précédemment lors de la purification de la NP PR8 (**Figure 25B**). Si comme proposé précédemment la capacité maximale de fixation de la colonne est atteinte avec la protéine NP, alors les contaminants ne peuvent pas être retenus. Cette hypothèse semble néanmoins peu probable étant donné la quantité de protéines retrouvée dans les fractions de lavage, il est donc possible que les contaminants soient trop dilués lors de l'élution et donc non visibles sur le gel.

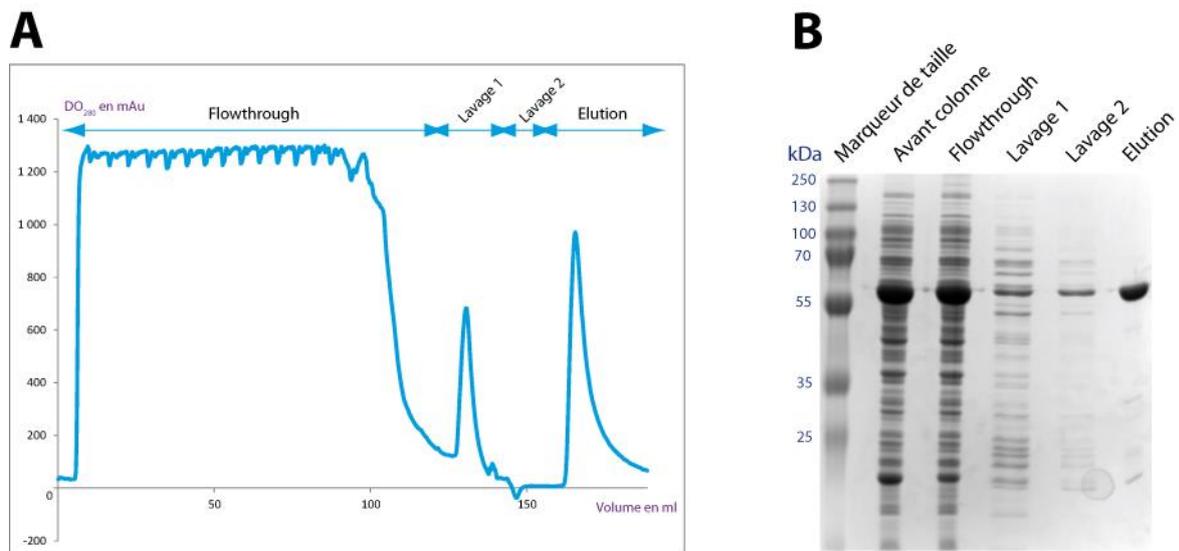


Figure 25 : Première étape de purification de la protéine NP WSN sur colonne Ni-NTA. **(A)** Spectre UV_{280nm} en fonction du volume passé sur la colonne. Les différentes étapes de la chromatographie sont indiquées en haut de la figure. **(B)** Vérification de la purification sur gel SDS-PAGE. A chaque étape une aliquote a été prélevée et analysée.

Le spectre UV_{280nm} de la seconde purification (**Figure 26A**) est à nouveau similaire à celui obtenu lors de la purification de la NP PR8 (**Figure 21A**), bien qu'ici l'intensité du pic majeur d'élution soit beaucoup plus importante, atteignant 2300 mAu pour le plus important contre 100 mAu pour la NP PR8. Un premier pic, de plus faible intensité, est observé en début de gradient et correspond à des contaminants. Sur le gel, on observe que le second pic contient la protéine d'intérêt à la taille attendue (**Figure 26B**). Contrairement à la protéine NP PR8, une seule bande est visible et beaucoup plus intense. On observe sur le gel que des contaminants sont présents malgré cette deuxième étape de purification, bien que l'intensité des bandes soit très faible comparée à celle de la protéine NP. La composition des fractions d'élution a été analysée par la plateforme de protéomique de l'institut sur Q-ExactivTM

(ThermoFisher Scientific). Les résultats de cette analyse montrent que très peu de contaminants sont présents, dans de très faibles proportions, et la protéine NP de la souche WSN est bien identifiée.

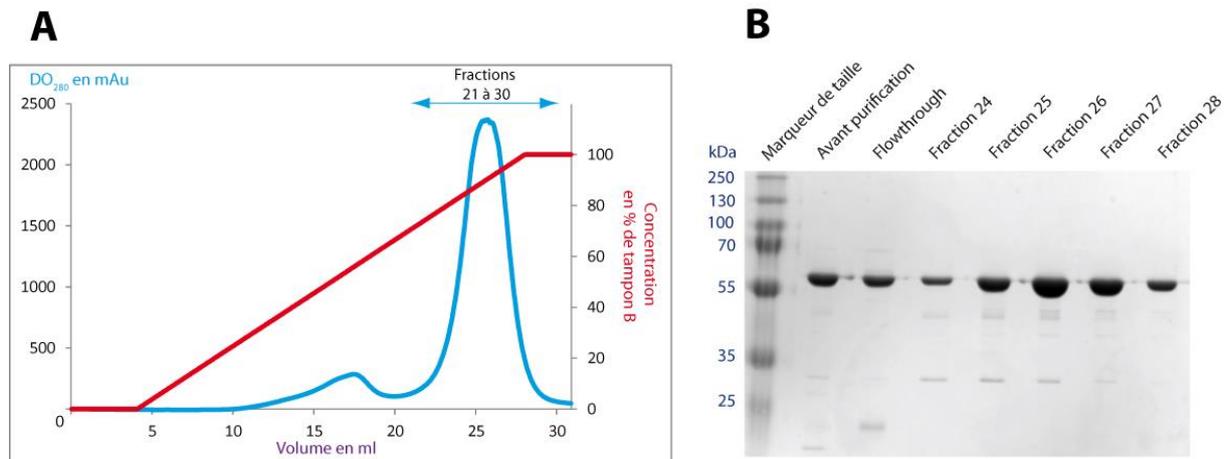


Figure 26 : Seconde étape de purification de la protéine NP WSN sur colonne Héparine. **(A)** Spectre UV₂₈₀ (en bleu) et gradient de tampon B (en rouge) en fonction du volume passé sur la colonne. **(B)** Vérification de la purification sur gel SDS-PAGE par analyse des fractions indiquées sur le spectre UV.

Les fractions les plus riches en protéine NP sont rassemblées puis dialysées contre deux tampons différents, facilitant la formation de monomères ou de trimères. Cet état d'oligomérisation est vérifié par DLS (**Figure 27**). Les profils DLS des deux échantillons sont similaires : l'allure des corrélogrammes, représentant la superposition de dix mesures, ne montre aucun signe d'agrégation de la protéine (**Figure 27A et C**). La distribution en intensité de la protéine monomérique (**Figure 27B**) est unimodale et monodisperse avec un rayon hydrodynamique de 3,5 nm, tandis que cette même distribution pour la protéine trimérique (**Figure 27D**) est également unimodale mais polydisperse, avec un rayon hydrodynamique d'environ 4,75 nm. Le logiciel Zetasizer indique que la taille des particules est d'environ 63 kDa et 130 kDa pour les protéines monomériques et trimériques, respectivement, ce qui confirme qu'elles sont dans l'état d'oligomérisation attendu.

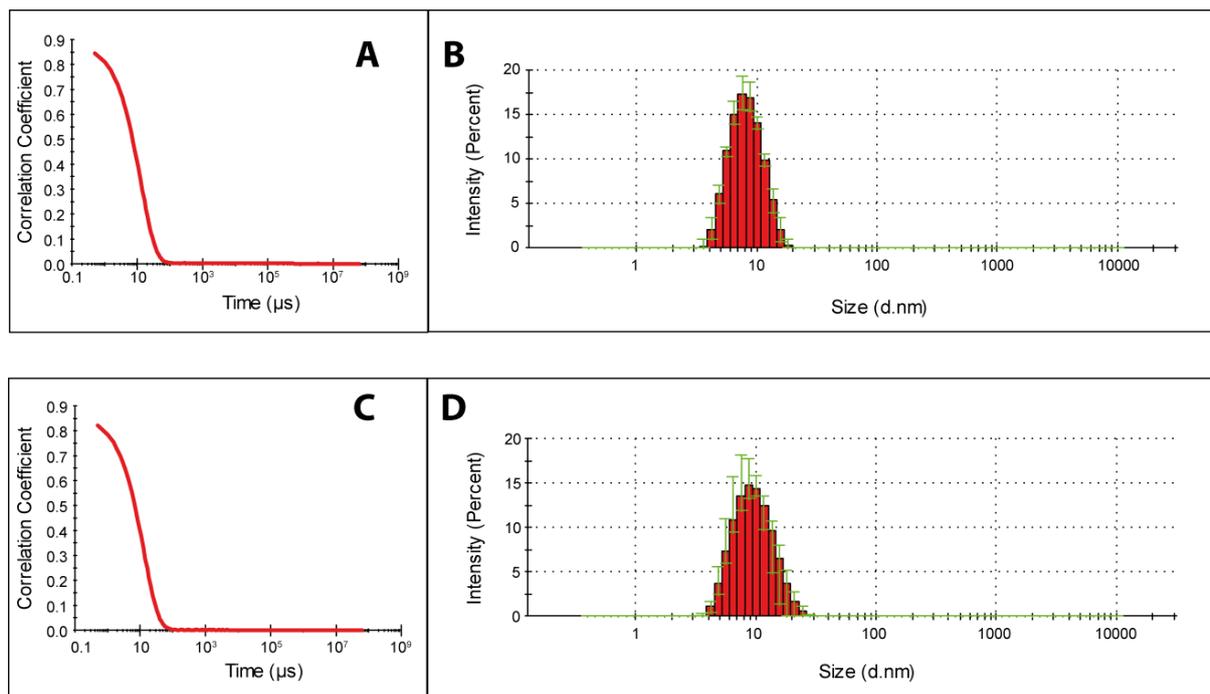


Figure 27 : Analyse de la protéine NP WSN, monomérique (**A** et **B**) et trimérique (**C** et **D**), purifiée et dialysée, par DLS. (**A** et **C**) Fonction d'autocorrélation (**B** et **D**) Distribution en intensité. Représentation de la moyenne de dix mesures consécutives.

J'ai ensuite vérifié l'activité de ces deux protéines par retard de migration sur gel électrophorétique en utilisant l'ARNv NS WSN transcrit *in vitro* (**Figure 28**). Différentes quantités de NP ont été utilisées lors de ces expériences afin d'obtenir des ratios allant d'une protéine pour cent nucléotides à une protéine pour vingt nucléotides. On observe que les protéines monomériques et trimériques sont capables de fixer l'ARN mais plusieurs différences sont à noter. Premièrement, on observe un retard de migration à un ratio plus faible pour la NP trimérique, dès le ratio 1/90 (**Figure 28B**), alors n'est observé qu'à partir du ratio 1/80 pour la protéine NP monomérique (**Figure 28A**). Deuxièmement, on observe qu'à partir du ratio 1/70 la totalité de l'ARN est retardé en présence de NP monomérique (**Figure 28A**) alors qu'une bande, de faible intensité, représentant l'ARN libre est toujours visible au ratio 1/60 en présence de NP trimérique (**Figure 28B**). En outre, à partir du ratio 1/50 la migration est identique pour les deux protéines, avec un retard de plus en plus important avec l'augmentation de la quantité de protéine ajoutée. Contrairement à ce qui a été observé avec la protéine NP PR8 le retard de migration est plus diffus et la position de la bande dépend de la quantité de protéine ajoutée. Cependant, les protéines sont fonctionnelles et seront utilisées pour les expériences de SHAPE *in vitro*. Ces résultats de retard de migration sur gel seront à confirmer ultérieurement.

En effet, des résultats différents ont été obtenus pour cette même expérience avec une protéine NP de moins bonne qualité, présentant des agrégats et un état d'oligomérisation moins bien défini.

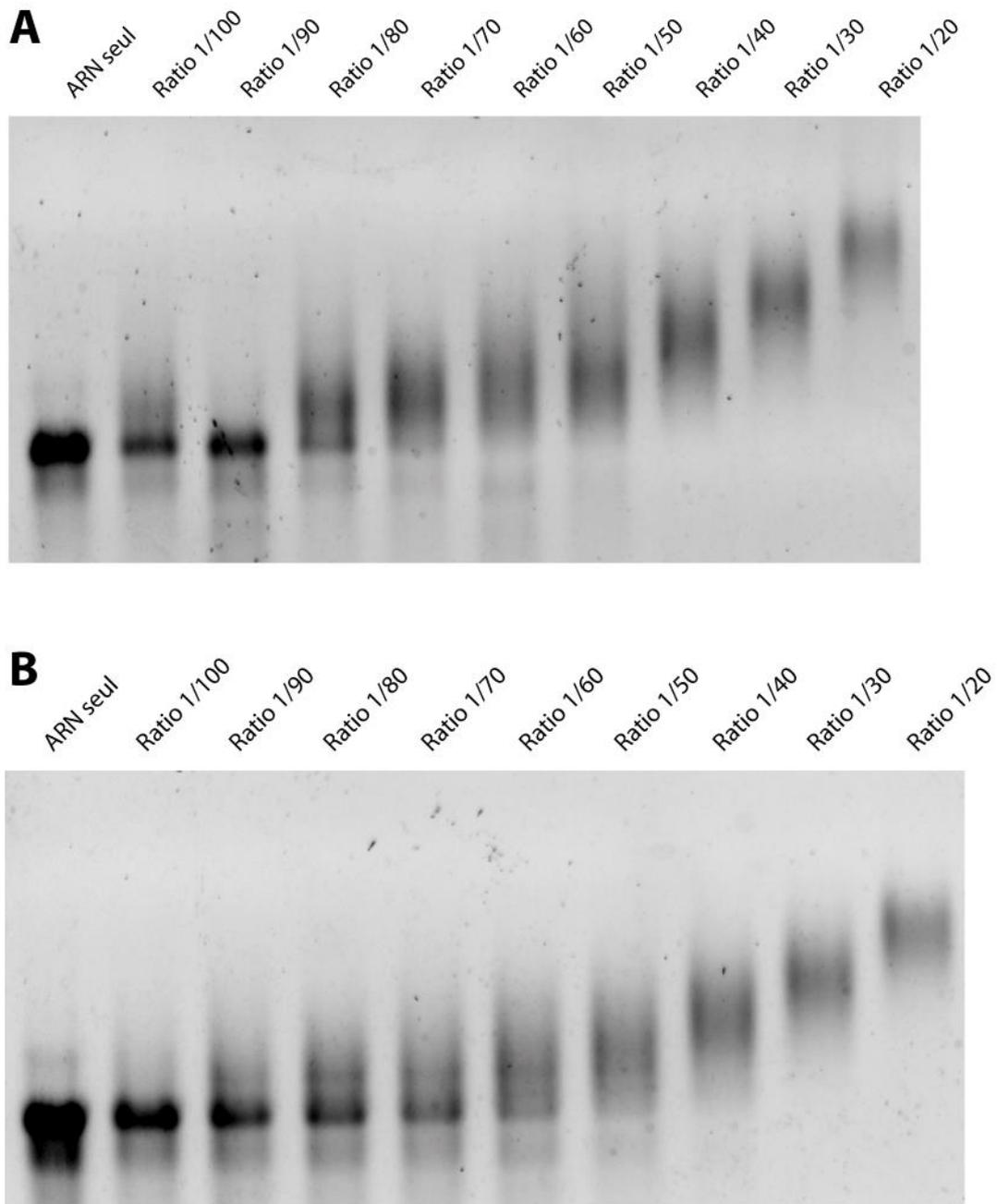


Figure 28 : Vérification de l'activité de la protéine NP WSN monomérique (A) et trimérique (B) par retard de migration sur gel. Les ratios indiquent le nombre de nucléotides pour lequel une protéine est ajoutée (ratio 1/50 : une protéine ajoutée pour 50 nucléotides).

B. Mise au point du protocole de SHAPE *in vitro*

Dans le but de caractériser la fixation de la NP à l'ARN et de mettre en évidence une éventuelle activité chaperonne, j'ai utilisé la méthode SHAPE sur des ARN nus (condition NoNP), des ARN complexés à la protéine (condition NP) et des ARN complexés à la protéine puis déprotéinisés (condition PK). Un problème s'est posé pour ces deux dernières conditions quant à la récupération de l'ARN fixé à la protéine NP. Afin de dissocier la NP, j'ai traité les complexes ribonucléoprotéiques (formés avec la NP monomérique) avec la protéinase K en présence ou non de SDS. Après traitement, les échantillons sont déposés sur gel pour contrôler l'efficacité de la protéinase K. On observe (**Figure 29**) que la protéinase K seule ne suffit pas à récupérer un ARN libre : il y a toujours un retard de migration bien qu'il soit moins marqué que la condition non traitée (complexe NS/NP). L'ajout de 1% ou 2% de SDS permet cependant d'éliminer les protéines NP résiduelles fixées à l'ARN qui migre à la même hauteur que l'ARN seul. Le traitement protéinase K supplémenté de SDS semble donc adéquat pour récupérer l'ARN après avoir formé le complexe.

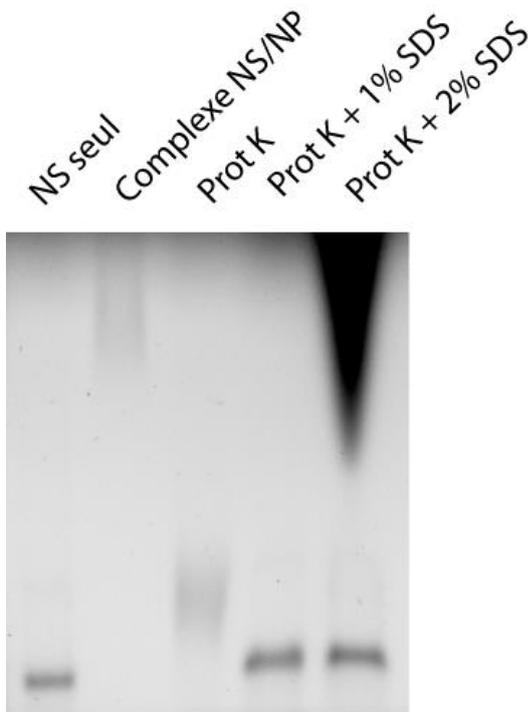


Figure 29 : Récupération de l'ARN complexé à la NP. Après formation, les complexes ARN/NP sont traités avec la protéinase K en présence ou en absence de SDS puis déposés sur gel d'agarose 0,8%.

Malgré ces résultats encourageants, les expériences de SHAPE réalisées avec la protéine NP ne donnent pas de résultats dans les conditions PK et NP : aucun signal fluorescent après rétrotranscription n'est visible lors de la séparation des ADNc par électrophorèse capillaire. J'ai donc voulu déterminer si le problème était lié à l'étape

de rétrotranscription ou à l'étape de précipitation qui suit le traitement à la protéinase K et l'extraction phénol-chloroforme. Pour cela, j'ai réalisé les mêmes contrôles que précédemment en ajoutant les étapes d'extraction phénolique et de précipitation à l'éthanol. On observe sur le gel (**Figure 30**) que deux conditions empêchent la récupération de l'ARN (pistes 11 et 20). Ces conditions correspondent à un ARN, complexé (piste 20) ou non (piste 11) à la NP, traité avec la protéinase K supplémenté de 1% SDS, et précipité à l'éthanol après une extraction phénol-chloroforme. Le même traitement en absence de SDS (piste 8 et 17) permet la récupération de l'ARN, tout comme le traitement SDS seul (piste 5). Avant l'étape de précipitation à l'éthanol, toutes les conditions en présence de protéinase K et de SDS permettent de récupérer l'ARN (pistes 9, 10, 18 et 19).

ARN NS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Protéine NP	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+
Protéinase K	-	-	-	-	-	+	+	+	+	+	+	-	-	-	+	+	+	+	+
SDS 1%	-	-	-	+	+	-	-	-	+	+	+	-	-	-	-	-	+	+	+
Phénol/Chlo	-	+	+	+	+	-	+	+	-	+	+	-	+	+	-	+	+	-	+
Précipitation	-	-	+	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-	+

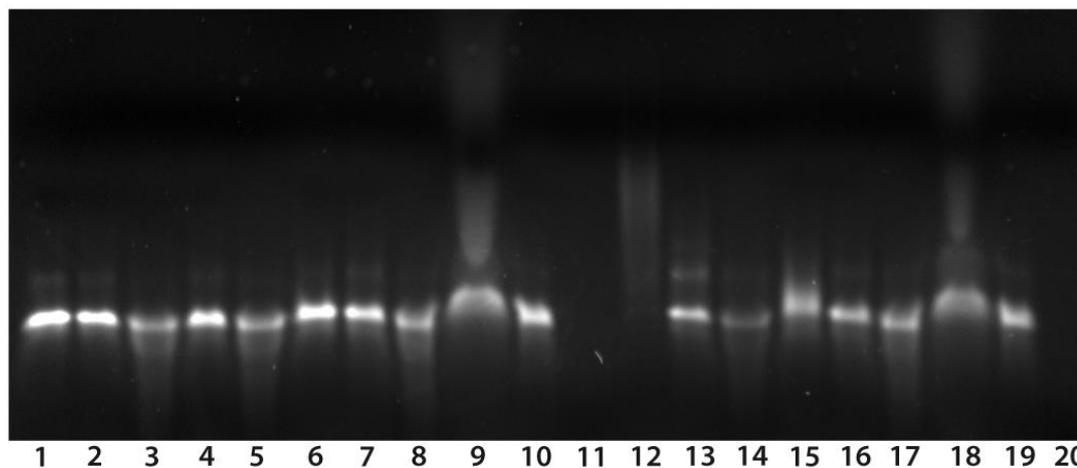


Figure 30 : Récupération de l'ARN complexé à la protéine NP. Après formation (pistes 12 à 20), les complexes subissent différents traitements (protéinase K, SDS, extraction phénolique, précipitation à l'éthanol) et sont déposés sur gel d'agarose 0,8%. Les ARN non complexés à la NP sont également traités dans les mêmes conditions (pistes 2 à 11).

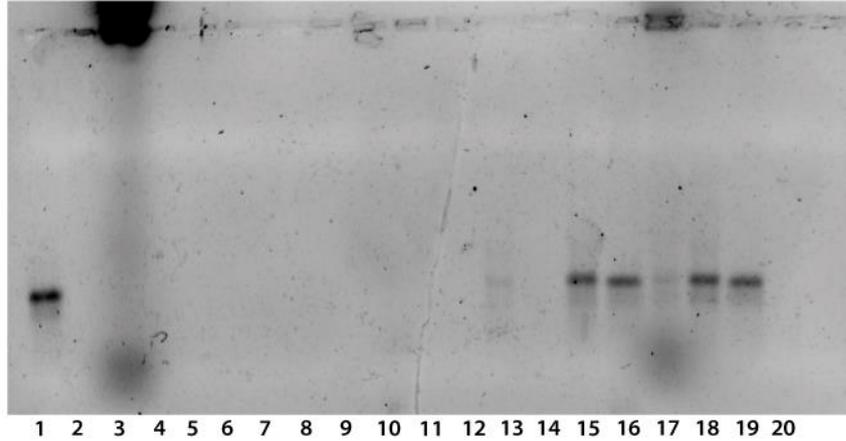
Ces résultats permettent de conclure à une incompatibilité entre la protéinase K et le SDS au moment de l'étape de précipitation à l'éthanol, indépendamment de la présence de NP. Le traitement protéinase K seul, suivi d'une extraction phénolique et d'une précipitation à l'éthanol, étant suffisant pour récupérer l'ARN (piste 17), c'est la

condition qui a été choisie pour les expériences de SHAPE *in vitro* pour les conditions NP et PK.

Les conditions déterminées par ce nouveau test ont permis de résoudre les problèmes rencontrés pour la condition PK, puisque dans ces conditions un signal fluorescent est observé après rétrotranscription et séparation des ADNc par électrophorèse capillaire. Cependant, pour la condition NP les problèmes n'ont pas été résolus par ces essais. En effet, après rétrotranscription à l'aide d'amorces fluorescentes, aucun signal n'est visible pour cette condition, qui fait intervenir la modification chimique à une étape différente par rapport à la condition PK. En effet, pour la condition PK, le complexe ribonucléoprotéique est formé, déprotéinisé puis modifié tandis que pour la condition NP, c'est le complexe qui est modifié avant d'être déprotéinisé. Etant donné cette différence de traitement et l'absence de signal après transcription inverse dans la condition NP, il est possible que le NMIA (utilisé pour la modification dans les expériences de SHAPE) interfère avec la NP et inhibe le traitement par la protéinase K, entraînant l'ARN dans la phase organique après extraction phénol-chloroforme. Pour confirmer cela, j'ai réalisé un nouveau test, en utilisant cette fois-ci les protéines monomériques (**Figure 31A**) et trimériques (**Figure 31B**). Pour ce test, l'ARN est complexé (pistes 2 à 15) ou non (pistes 16 à 20) à la protéine NP, et traité (pistes 3 à 11 et 16 à 20) ou non (pistes 12 à 15) avec le réactif chimique NMIA. Les mêmes traitements que précédemment (**Figure 30**) sont réalisés. Un traitement à 90°C pendant 2 min a été ajouté (pistes 8 à 11) dans le but de dénaturer la protéine NP et de récupérer l'ARN plus facilement. On observe sur le gel qu'aucune condition ne permet de récupérer l'ARN après traitement du complexe avec le NMIA (pistes 3 à 11). En absence de NP, l'ARN est récupéré (pistes 16 à 18) à l'exception des conditions traitées avec la protéinase K (pistes 19 et 20). Ces résultats semblent démontrer une incompatibilité entre le NMIA et la protéine NP, conformément à l'hypothèse émise.

ARN NS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Protéine NP	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	
NMIA	-	-	+	+	+	+	+	+	+	+	-	-	-	-	+	+	+	+	+	
Protéinase K	-	-	-	-	-	+	+	-	-	+	+	-	+	+	-	-	-	-	+	+
90°C	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	
Phénol/Chlo	-	-	-	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	
Précipitation	-	-	-	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	

A



ARN NS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Protéine NP	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	
NMIA	-	-	+	+	+	+	+	+	+	+	-	-	-	-	+	+	+	+	+	
Protéinase K	-	-	-	-	-	+	+	-	-	+	+	-	-	+	+	-	-	-	+	+
90°C	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	
Phénol/Chlo	-	-	-	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	
Précipitation	-	-	-	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	

B

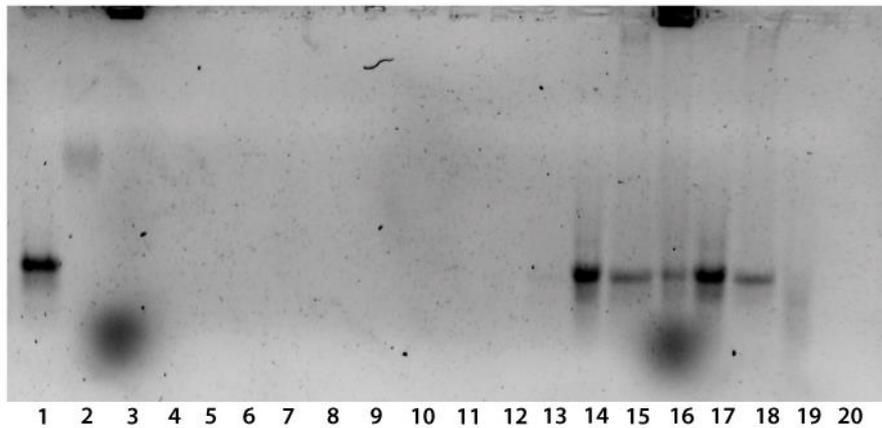


Figure 31: Récupération de l'ARN complexé à la protéine NP monomérique (A) ou trimérique (B). Après formation, les complexes subissent différents traitements (protéinase K, incubation à 90°C pendant 2 min, extraction phénolique, précipitation à l'éthanol) et sont déposés sur gel d'agarose 0,8%.

Afin de s'affranchir de cette incompatibilité, j'ai réalisé un dernier test dans lequel le complexe ribonucléoprotéique (formé avec la protéine monomérique) est précipité après traitement au NMIA dans le but de d'éliminer le surplus de réactif. Le traitement protéinase K est réalisé sur le complexe précipité, suivi d'une extraction phénolique et d'une nouvelle précipitation. On observe sur le gel (Figure 32) que la première précipitation après traitement au NMIA (piste 3) permet déjà de récupérer un ARN qui

migre comme un ARN nu. Après la deuxième précipitation qui suit le traitement protéinase K et l'extraction phénolique (piste 4) l'ARN est aussi récupéré et cette condition sera utilisée dans le protocole de SHAPE *in vitro* pour la condition NP.

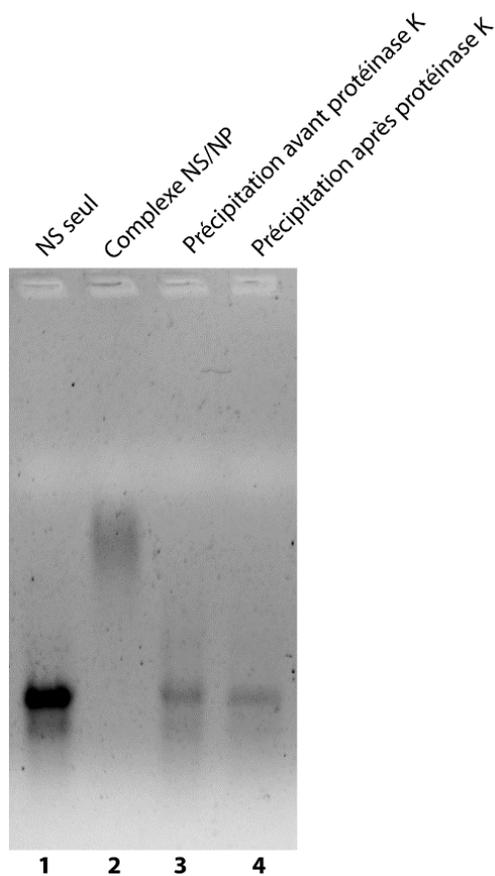


Figure 32 : Récupération de l'ARN complexé à la protéine NP monomérique. Après formation (piste 2), les complexes sont précipités à l'éthanol (piste 3) puis traités avec la protéinase K avant une seconde précipitation à l'éthanol (piste 4) et dépôt sur gel d'agarose 0,8%.

C. Cartographie chimique *in viro*

1. Mise au point de la modification

La partie *in viro* de mon travail de thèse a été réalisée en collaboration avec l'équipe de Martin Schwemmler à Freiburg (Allemagne) et plus particulièrement avec Hardin Bolte, doctorant, qui a produit et purifié le virus SC35M sauvage. Nous avons dans un premier temps évalué la capacité de nos réactifs à modifier l'ARNv à l'intérieur des particules virales. Pour cela nous avons utilisé des particules virales purifiées sur coussin de sucrose qui ont été conservées intactes ou traitées avec un tampon de désassemblage qui permet de déstabiliser les interactions inter-RNPv. Cette stratégie nous permettra, en comparant les deux conditions, d'identifier les régions impliquées dans ces interactions. La modification de l'ARNv a ensuite été réalisée avec deux

réactifs chimiques, le NMIA et le DMS. J'ai ensuite réalisé une étape de transcription inverse avec une amorce radioactive spécifique des segments PB1 ou NA avant de déposer les échantillons sur gel polyacrylamide 8% en conditions dénaturantes (**Figure 33**). On observe sur le gel des régions qui deviennent réactives lorsque la modification est réalisée sur les particules désassemblées, validant ainsi notre approche expérimentale. Le traitement des particules virales par le tampon de désassemblage permet de révéler des régions impliquées dans des interactions ARN-ARN au sein des particules virales intactes (**Figure 33**, flèches bleues). On remarque également que des régions sont réactives en présence de DMS mais non réactives en présence de NMIA (**Figure 33**, flèches orange). Ces résultats suggèrent que les deux approches de cartographie chimique choisies sont complémentaires et leur combinaison sera nécessaire pour déterminer précisément les régions impliquées dans les interactions intermoléculaires entre RNPv lors de l'empaquetage sélectif. Pour le segment PB1, on observe qu'il y a plus de signal lorsque la modification a été réalisée sur des particules désassemblées. De même, le signal est plus important dans la condition DMS comparée à la condition NMIA. Dans le cas du segment NA on remarque une tendance inverse, avec un signal plus important pour la modification dans les particules intactes. Cependant, pour ce segment également la condition DMS montre le meilleur signal. Indépendamment du segment étudié, on observe plus d'arrêts de la transcription inverse pour les particules virales désassemblées dans les conditions non modifiées (pistes « témoin négatif »). Cette première expérience permet de montrer que la modification de l'ARN au sein des particules virales est réalisable, avec une meilleure efficacité en utilisant du DMS. Afin d'obtenir une population homogène de particules virales, permettant d'éliminer celles qui sont endommagées et qui peuvent introduire un bruit de fond important lors de la comparaison entre particules virales intactes et désassemblées, il pourrait s'avérer utile d'ajouter une deuxième étape de purification des particules virales.

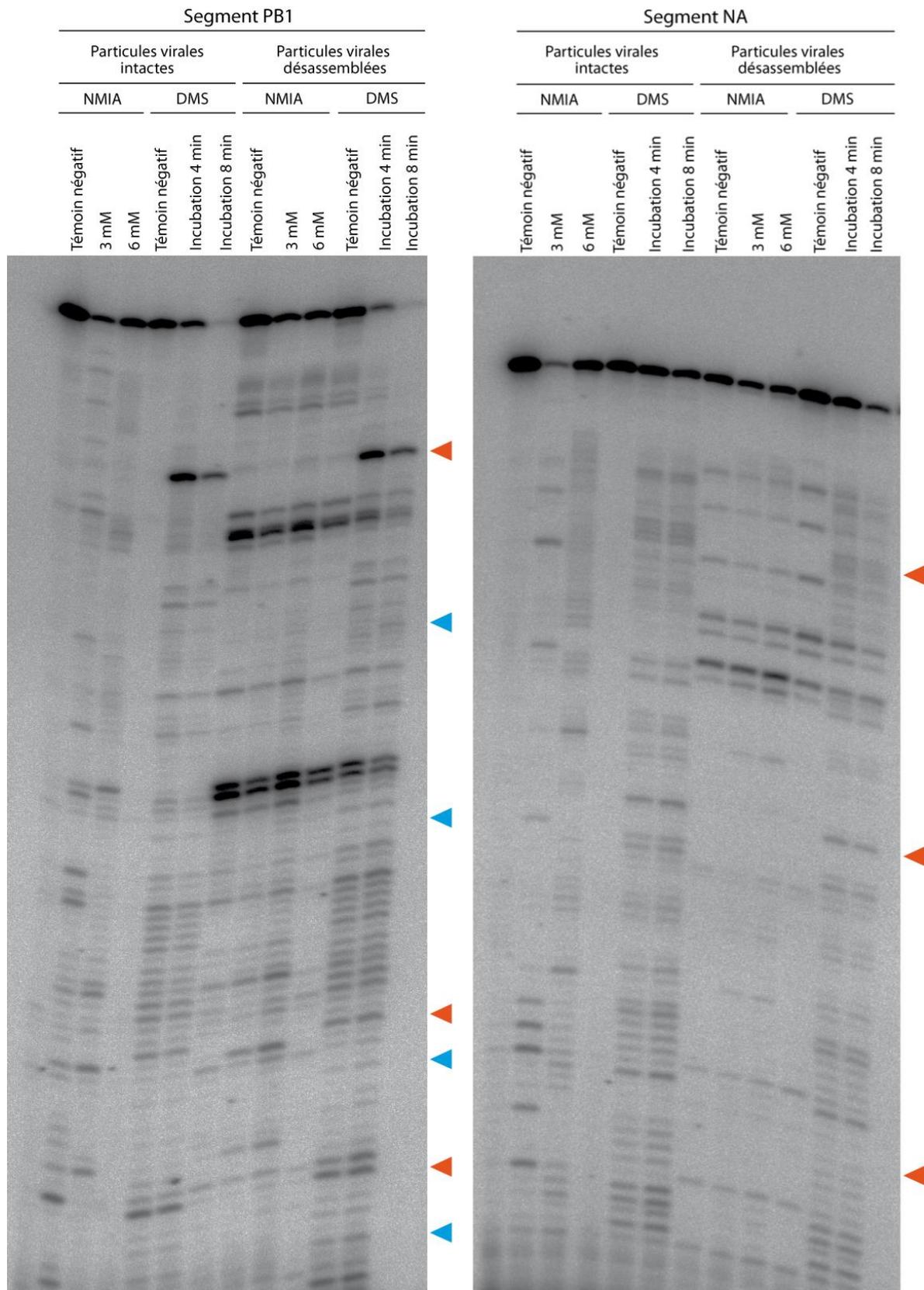


Figure 33 : Détection des modifications de l'ARN *in vitro* par transcription inverse grâce à des amorces radioactives s'hybridant sur les segments 2 (PB1) et 6 (NA). Les flèches bleues désignent les différences entre les conditions intactes et désassemblées tandis que les flèches orange désignent les différences entre les conditions modifiées au NMIA et DMS.

Nous avons choisi de réaliser une étape de purification sur gradient OptiPrep à la suite du coussin de sucrose. Cette étape permettant de séparer les particules virales en fonction de leur densité, la sédimentation des particules intactes et des particules endommagées sera différente. Pour ce nouveau test, nous avons utilisé le DMS et le réactif SHAPE Ai-N₃, qui sera utilisé pour le protocole d'icSHAPE. L'étape de transcription inverse a été réalisée à l'aide d'un primer radioactif s'hybridant au segment NA. On observe sur le gel (**Figure 34**) qu'il y a moins de modifications détectées dans les particules virales purifiées en deux étapes. La modification par le réactif Ai-N₃ des particules virales intactes, ayant subi deux purifications successives, semble très faible, mais cela provient plus certainement d'un problème lors du dépôt des échantillons sur le gel, étant donné que l'intensité de la bande correspondant à l'ADNc de pleine longueur est très faible. On observe que certains nucléotides sont réactifs en présence de DMS mais pas en présence d'Ai-N₃, le réactif SHAPE (**Figure 34**, flèches noires). La modification au DMS est réalisée au niveau des appariements Watson-Crick, tandis que le réactif SHAPE modifie les nucléotides au niveau du 2'OH du ribose. Les nucléotides présentant une sensibilité au DMS et pas à l'Ai-N₃ ne sont donc pas impliqués dans une interaction ARN-ARN mais plus vraisemblablement dans une interaction ARN-protéine. Ces résultats indiquent donc qu'une stratégie de cartographie chimique simple, avec un seul réactif, peut entraîner des biais dans l'interprétation des réactivités du fait de l'impossibilité de discriminer les interactions ARN-ARN des interactions ARN-NP. On observe également sur le gel que certaines régions, réactives dans la condition simplement purifiée (coussin de sucrose uniquement) le sont bien moins après ajout d'une étape de purification sur gradient d'OptiPrep (**Figure 34**, flèches bleues). Ces résultats suggèrent que des particules virales endommagées sont présentes après la purification sur coussin de sucrose, et qu'elles sont éliminées après purification sur gradient OptiPrep.

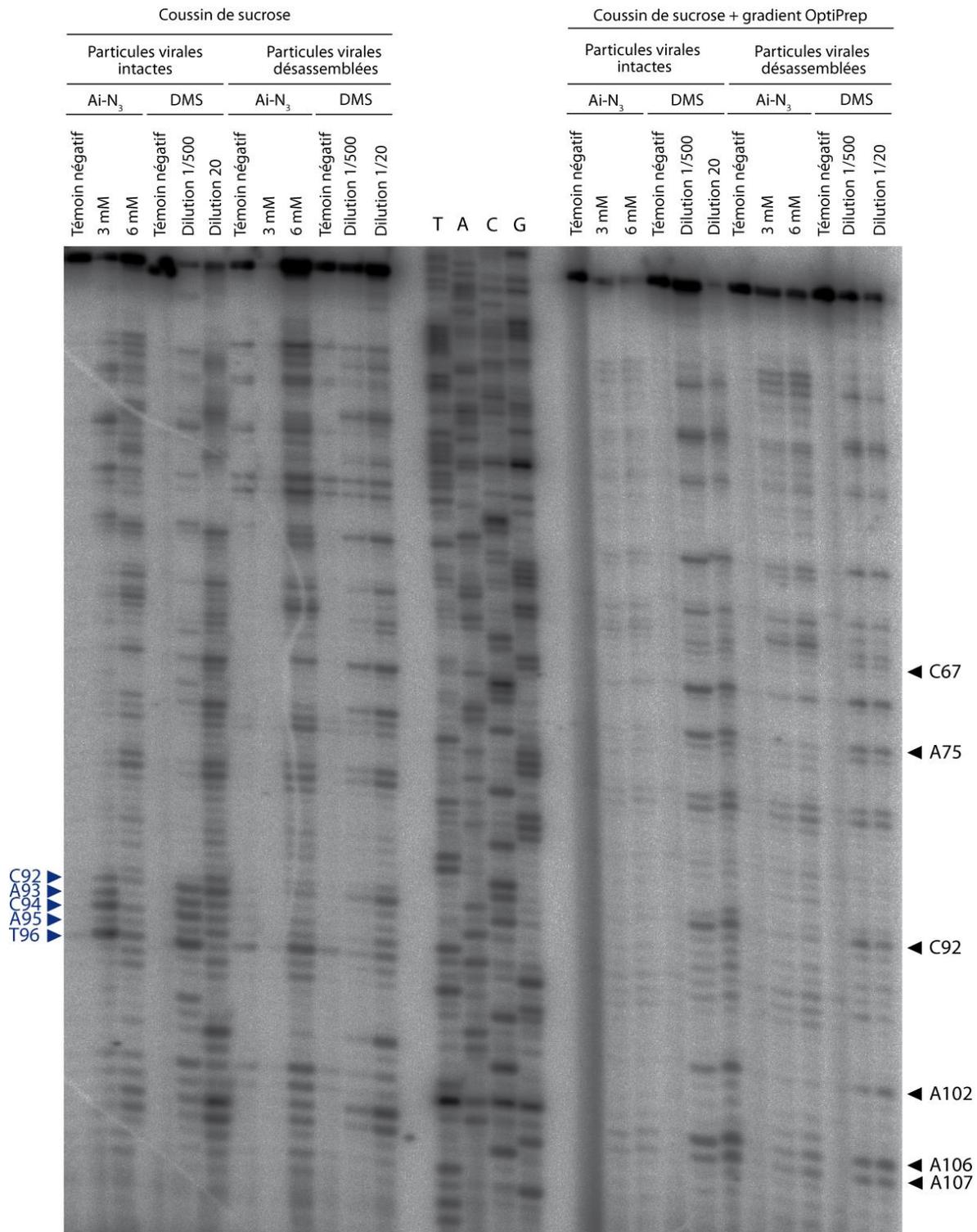


Figure 34 : Détection des modifications de l'ARN *in vitro* par transcription inverse grâce à l'utilisation d'amorces radioactives s'hybridant sur le segment 6 (NA). Les pistes A, T, G et C représentent les pistes de séquençage. Les flèches bleues désignent les différences entre les échantillons simplement et doublement purifiés tandis que les flèches noires désignent les différences entre les conditions modifiées au Ai-N₃ et DMS.

2. Mise en place du protocole icSHAPE

Notre stratégie pour le protocole d'icSHAPE diffère légèrement de celui publié par Spitale *et coll.* (**Figure 35**). En effet, après l'étape de 'click-chemistry' qui permet l'ajout d'une biotine sur la modification induite par la réactif chimique et la sélection sur billes de streptavidine, l'élution est réalisée par réduction du pont disulfure présent entre la biotine et le nucléotide modifié, contrairement au protocole original qui réalise l'élution par traitement RNase après rétrotranscription [235]. De plus, nous utilisons une étape de ligation supplémentaire après fragmentation, qui permet l'ajout d'un oligonucléotide en 5' de chaque fragment afin de discriminer les produits de rétrotranscription tronqués (pause de l'enzyme ou modification chimique) des produits de pleine longueur.

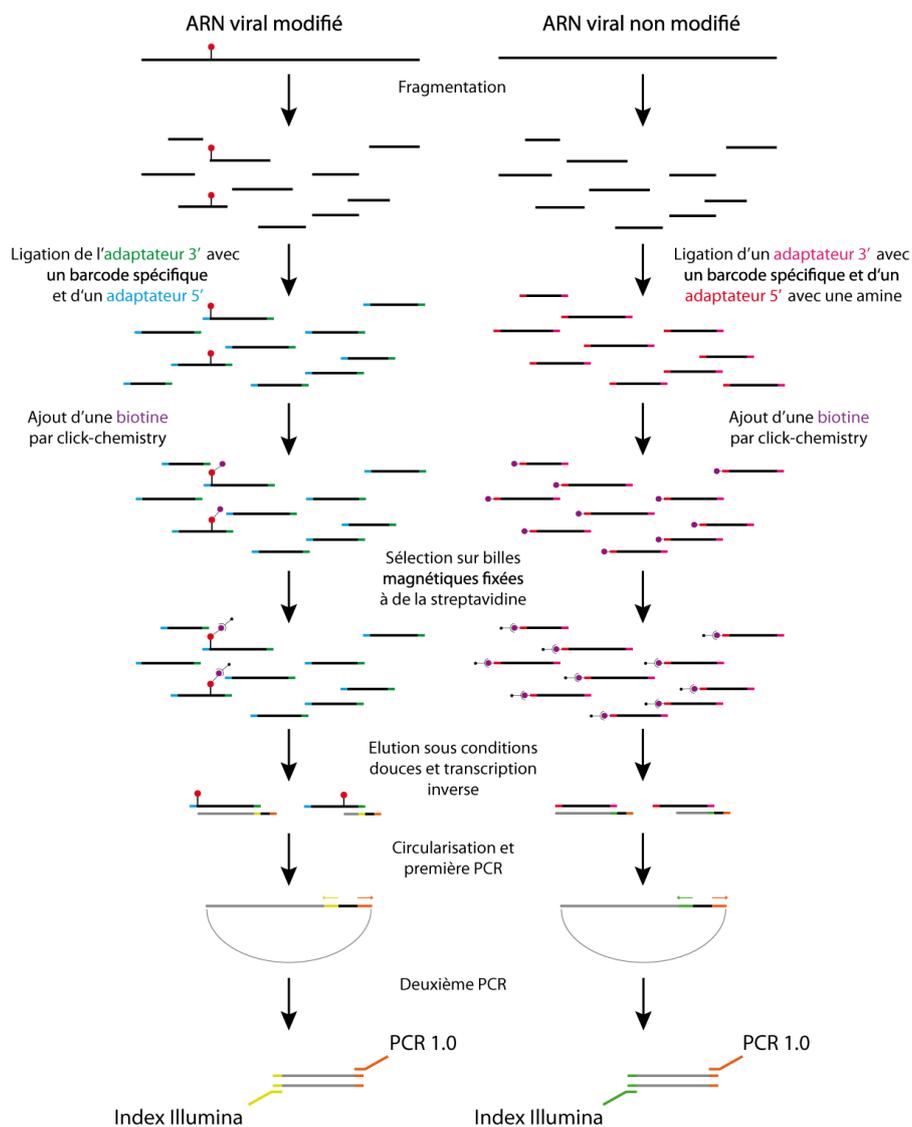


Figure 35 : Stratégie expérimentale du protocole d'icSHAPE développé au cours de ma thèse.

Avant d'appliquer le protocole icSHAPE sur les ARNv purifiés à partir des particules virales intactes et désassemblées modifiées, j'ai tenté de mettre au point certaines étapes de ce protocole sur un mélange équimolaire des huit ARNv de la souche WSN transcrits *in vitro* non modifiés. Les quantités utilisées ont été calculées pour refléter la quantité de matériel obtenue après extraction de l'ARN viral à la suite des étapes de purification des particules virales et de modification, réalisées sur 200 ml de surnageant provenant de l'infection de six flasques T175 à confluence. J'ai utilisé des oligonucléotides radioactifs afin de visualiser l'efficacité de chaque étape du protocole sur gel dénaturant acrylamide-urée (**Figure 36**). Dans les pistes 1 et 2, l'adaptateur 3' est radioactif et le protocole est stoppé avant ou après la purification sur colonne. Dans les pistes 3 et 4, l'adaptateur 3' est non marqué, alors que l'adaptateur 5' est radioactif et le protocole est stoppé avant ou après la purification sur colonne. Enfin, dans la piste 5, les deux adaptateurs 3' et 5' sont non marqués, alors que l'amorce de transcription inverse est radioactive. A toutes les étapes, on observe sur le gel un *smear*, dû à l'étape de fragmentation aléatoire réalisé en début de protocole. Les bandes très intenses des pistes 1, 2 et 3 correspondent aux oligonucléotides radioactifs utilisés. L'adaptateur 3' non ligué n'est donc pas éliminé efficacement lors de la purification sur colonne (piste 2) alors que l'adaptateur 5' libre l'est (piste 4). La présence d'adaptateur 3' libre lors de l'étape de ligation de l'adaptateur 5' peut induire la ligation des deux adaptateurs entre eux. L'étape de rétrotranscription est efficace, comme l'indique la présence d'un *smear* sur la piste 5, confirmant la présence d'ARNv jusqu'à cette étape.

Les résultats obtenus avec ce premier test étant encourageant, j'ai appliqué le protocole d'icSHAPE sur les particules virales intactes ou désassemblées modifiées par l'Ai-N3. La quantification par qPCR réalisée après l'étape de circularisation en fin de protocole a donné un résultat négatif et la vérification sur gel a confirmé l'absence de matériel biologique. L'étape de circularisation n'ayant pas été testée préalablement, il est possible qu'elle soit à l'origine de ce résultat. Alternativement, un problème lors de l'étape de ligation de l'adaptateur 3' (auquel s'hybride l'amorce de transcription inverse) ou lors de la transcription inverse peut également expliquer ce résultat négatif. Etant donné la faible quantité de matériel de départ (10^{11} copies d'ARN par segment), il est difficile de suivre à chaque étape la présence d'acide nucléique. Ce protocole comportant de nombreuses purifications sur colonne et de précipitations à l'éthanol, la perte de matériel peut intervenir à de nombreux endroits.

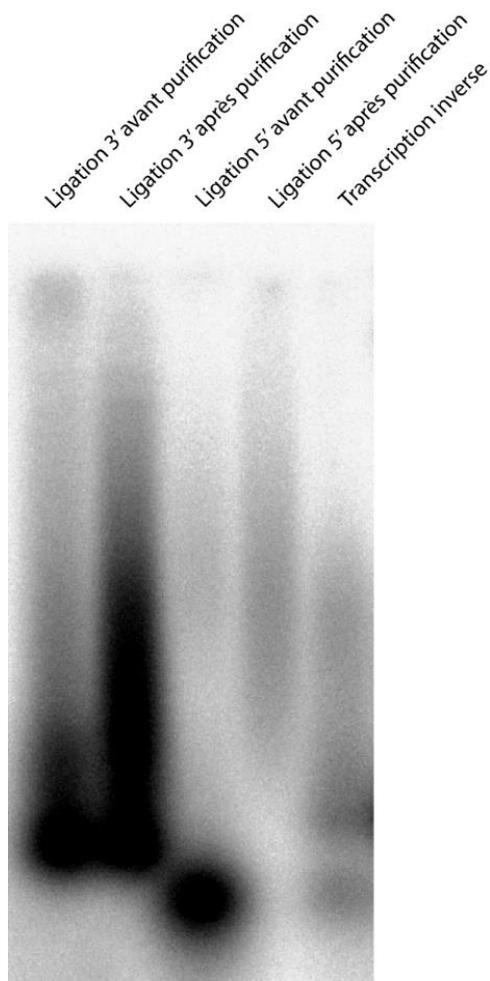


Figure 36 : Test des étapes du protocole icSHAPE. Les adaptateurs 3' (pistes 1 et 2), 5' (pistes 3 et 4) et l'amorce RT (piste 5) sont marqués radioactivement et détectés sur gel.

3. Mise en place du protocole de SHAPE-MaP

Afin d'avoir une alternative au protocole icSHAPE en cas de difficultés persistantes, nous avons choisi d'utiliser l'approche SHAPE-MaP [232], qui utilise des conditions de transcription inverse spécifique favorisant l'introduction de mutations en face des modifications de l'ARNv, plutôt qu'un arrêt de la polymérase, comme dans les techniques traditionnelles de cartographie en solution de l'ARN (**Figure 16**). Ce protocole comporte moins d'étapes pour la création de banques de séquençage et a récemment été appliqué au virus Influenza [45]. Comme pour le protocole icSHAPE, j'ai d'abord testé le protocole de SHAPE-MaP sur un mélange équimolaire des huit ARNv de la souche WSN transcrits *in vitro*. Les échantillons sont dans un premier temps rétro-transcrits dans les conditions MaP à l'aide d'un mélange d'amorces aléatoires, puis le second brin est synthétisé. L'ADN double brin ainsi obtenu est alors fragmenté et étiqueté à l'aide de transposomes. Ces tags permettent d'ajouter par PCR les index Illumina en 5' et 3' nécessaires au séquençage haut débit. A la fin du protocole,

après l'étape de PCR, un dépôt sur gel a été réalisé afin de visualiser la présence de matériel. Deux bandes sont visibles sur le gel (**Figure 37**) et chacune de ces bandes a été découpée. L'ADN contenu a été purifié et cloné dans un vecteur pJET (ThermoFisher Scientific), 4 clones ont été séquencés et enfin comparé aux bases de données du NCBI.

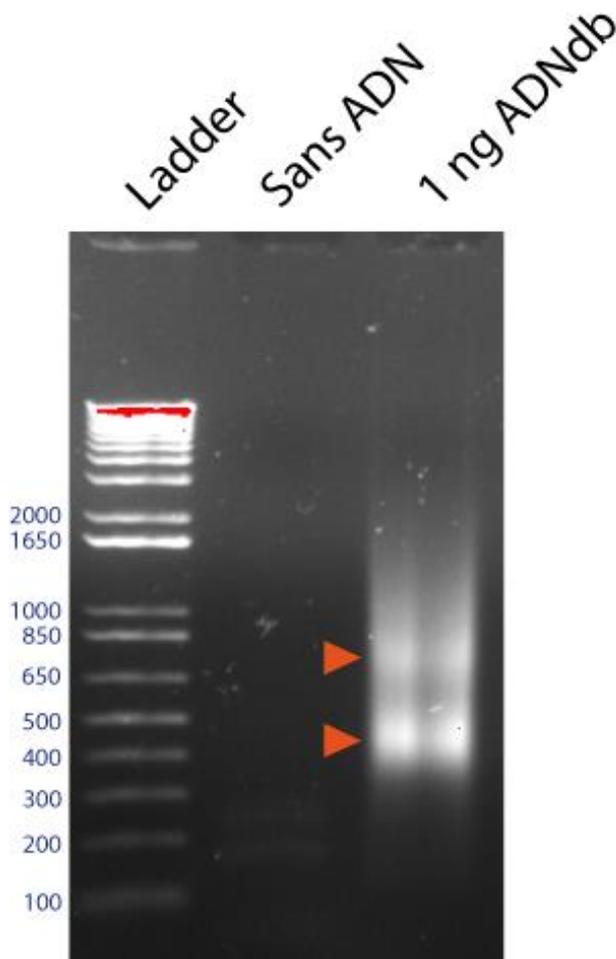


Figure 37 : Contrôle de la présence d'ADN après la dernière étape de PCR du protocole SHAPE-MaP. La PCR a été réalisée en absence d'ADN ou avec 1 ng d'ADN double brin (ADNdb) obtenu après transcription inverse et synthèse du second brin. Le dépôt a été réalisé sur gel d'agarose 0,8 % et la révélation est réalisée par transillumination aux UV.

On observe (**Figure 38**) que les séquences clonées correspondent aux ARNv de la souche WSN (deux clones correspondants aux segments 4 et 6 sont présentés ici). Ces résultats montrent que le protocole de SHAPE-MaP permet de créer des banques pour le séquençage haut débit, et j'ai confirmé que ces banques correspondent au matériel de départ et non à des multimères d'index et adaptateurs. Cependant, je ne retrouve pas dans les séquences obtenues après clonage les index 5' et 3' ajoutés par PCR, les séquences de la grippe étant directement encadrées par la séquence du vecteur pJET. Néanmoins, la PCR semble avoir fonctionné et les séquences de la souche WSN ont bien été clonées après la réalisation du protocole entier. La prochaine étape sera d'appliquer ce protocole aux échantillons biologiques *in viro*.

TTCATTGTTTTTTCATCATAGTGAATAAAATCAACTGCTTTAACACTTGTGCC
 TGAACACCATATCCATCCGGCGTAATACGACTCACTATAGGGAGAGCGGCCGCC
 AGATCTTCCGGATGGCTCGAGTTTTTCAGCAAGATGGAATGATAGATGGATGG
 TATGGTTATCATCATCAGAATGAACAGGGATCAGGCTATGCAGCGGATCAAAA
 AAGCACACAAAATGCCATTAACGGGATTACAAACAAGGTGAACTCTGTTATCG
 AGAAAATGAACACTCAATTCACAGCTGTGGGTAAGAATTCAACAACCTTAGAA
 AAAAGGATGGAAAATTTAAATAAAAAAGTTGATGATGGGTTTCTGGACATTGG
 GACATATAATGCAGAATTGTTAGTTCTACTGGAAAATGAAAGGACTTTGGATT
 TCCAATCTTCTAGAAGATCTCCTACAATATTCTCAGCTGCCATGGAAAATCGA
 TGTCTTCTTTTATTCTCTCAAGATTTTCAGGCTGTATATTAACACTTATATTA
 AGAACTATGCTAACCACCTCATCAGGAACCGTTGTAGGTGGCGTGGGTTTCTT
 GGCAATCGACTCTCATGAAAACACGAGCTAAATATTCAATATGTTCTCTTGA
 CCAACTTTATTCTGCATTTTTTTTGAACGAGGTTAGAGCAAGCTTCAGGAAAC
 TGAGACAGGAATTTTATTAATAAAATTTAAATTTTGAAGAAAGTTCAGGTTAAT

Influenza A virus (A/WSN/1933(H1N1)) viral cRNA, segment 4, complete sequence
 Sequence ID: [LC333185.1](#) Length: 1775 Number of Matches: 1

Range 1: 1106 to 1393 [GenBank](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
532 bits(288)	2e-147	288/288(100%)	0/288(0%)	Plus/Plus

TCCATTCATTGTTTTTTCATCATAGTGAATAAAATCAACTGCTTTA
 AACTTGTGCCTGAACACCATATCCATCCGGCGTAATACGACTCACT
 ATAGGGAGAGCGGCCCGCCAGATCTTCCGGATGGCTCGAGTTTTTCAG
 CAAGATTTTTTTTTTTTTTTTTTTTTTTTGGAGGGTATTAGTAGAAAAC
 AAGGAGTTTTTTGAACAAACTACTTGTCAATGGTGAACGGCAACTCA
 GCACCGTCTGGCCAAGACCAATCTACAGTATCACTATTCACACCACA
 AAAAGAAAATGATGCTCCCACTAGTCCAGATTGCGTCTCCTCAGGTA
 GCCCCGTGATTAATCAACCCAGAAGCAAGGCCCTCATACAGTCTAGC
 CCTGTTAGCTCAGGATGTTGAACGAAACTTCCGCTGTACCCTGACCG
 ATTAGTTATTGCCACAACATCTTGTCTCATAGAGAACCTACTATCAG
 TCTCTGTCCATCCATTAGGATCCCAAATCATCTCAAACATCTTTCTA
 GAAGATCTCCTACAATATTCTCAGCTGCCATGGAAAATCGATGTTCT
 TCTTTTATTCTCTCAAGATTTTCAGGCTGTATATTAACACTTATATTA
 AAGAACTATGCTAACCACCTCATCAGGAACCGTTGTAGGTGGCGTGG
 GTTTTCTTGGCAATCGACTCTCATGAAAACACGAGCTAAATATTCA
 ATATGTTCTCTTGACCAACTTTATTCTGCATTTTTTTTGAACGAGG

Influenza A virus (A/WSN/1933(H1N1)) viral cRNA, segment 6, complete sequence
 Sequence ID: [LC333187.1](#) Length: 1409 Number of Matches: 1

Range 1: 1081 to 1409 [GenBank](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
608 bits(329)	3e-170	329/329(100%)	0/329(0%)	Plus/Minus

Figure 38: Résultat du séquençage des séquences clonées dans le vecteur pJet après protocole SHAPE-MaP. Une aliquote des banques de séquençage a été clonée dans ce vecteur et quatre plasmides ont été séquencés et comparés aux bases de données du NCBI. La séquence de deux clones est présentée ici à titre d'exemple. En rouge sont représentées les séquences correspondant aux segments de la souche WSN. En vert est représentée la séquence du vecteur pJet. En noir est représentée une séquence dont l'origine est inconnue.

D. Résultats SHAPE *in vitro*

1. Comparaison des structures entre les souches PR8 et WSN

Pour toutes les conditions utilisées au cours de ma thèse, la structure 2D de l'ARN a été calculée et représentée à l'aide des logiciels RNAstructure (version 6.0) [209] et vaRNA [48]. Le premier permet de prédire la structure de l'ARN en combinant les données thermodynamiques et les réactivités SHAPE de chaque nucléotide. Le second permet d'améliorer la représentation graphique et d'associer à chaque nucléotide une couleur reflétant sa réactivité en suivant une échelle linéaire (**Figure 39**) allant de la réactivité minimale (0) à la réactivité maximale (1,5). Les nucléotides dont la réactivité excède 1,5 seront représentés avec la couleur correspondant à la réactivité maximale. Les nucléotides non analysés sont marqués en gris.



Figure 39 : Echelle de couleur correspondant à la valeur de réactivité déterminée par SHAPE.

Les premières structures représentées (**Annexes 1 à 4**) correspondent aux ARN M et NS des souches PR8 (**Annexes 1 et 2**) et WSN (**Annexe 3 et 4**) dans la condition NoNP (**Figure 40**), qui correspond à l'ARN replié dans tampon, en absence de protéine. Les ARN M et NS des souches PR8 et WSN présentent chacun une identité de séquence de 96%, mais leurs structures déterminées au cours de ma thèse sont différentes. Pour l'ARN M (**Annexes 1 et 3** pour les souches PR8 et WSN, respectivement), certaines régions présentent la même structure comme les tiges-boucles formées par les nucléotides 34-61 et 539-558 (encadrées en bleu) mais ce ne sont que de courtes régions. En analysant la région 565-953 (encadrée en rouge) on observe, pour la souche PR8, une boucle interne de laquelle partent six structures secondaires. Cette même région pour la souche WSN est divisée en deux parties connectées entre elles par région double brin, avec une première boucle interne de laquelle partent trois tiges-boucles (région 567-670) et une seconde boucle interne de laquelle partent quatre structures secondaires (région 700-931). Parmi toutes ces tiges boucles, une seule est identique entre les deux structures et se trouve entre les nucléotides 575 et 588. Des différences sont également observées pour l'ARN NS (**Annexes 2 et 4**) pour les souches PR8 et

WSN, respectivement) bien que de grandes régions présentent la même structure. En effet, les régions 412-535 et 693-819 (encadrées en bleu) forment la même structure entre ces deux souches. Cependant, d'autres régions présentent des structures radicalement différentes : par exemple les nucléotides 215-260 qui forment une tige boucle (encadrée en rouge) dans la souche WSN, alors qu'ils font partie d'une hélice double brin pour la souche PR8.

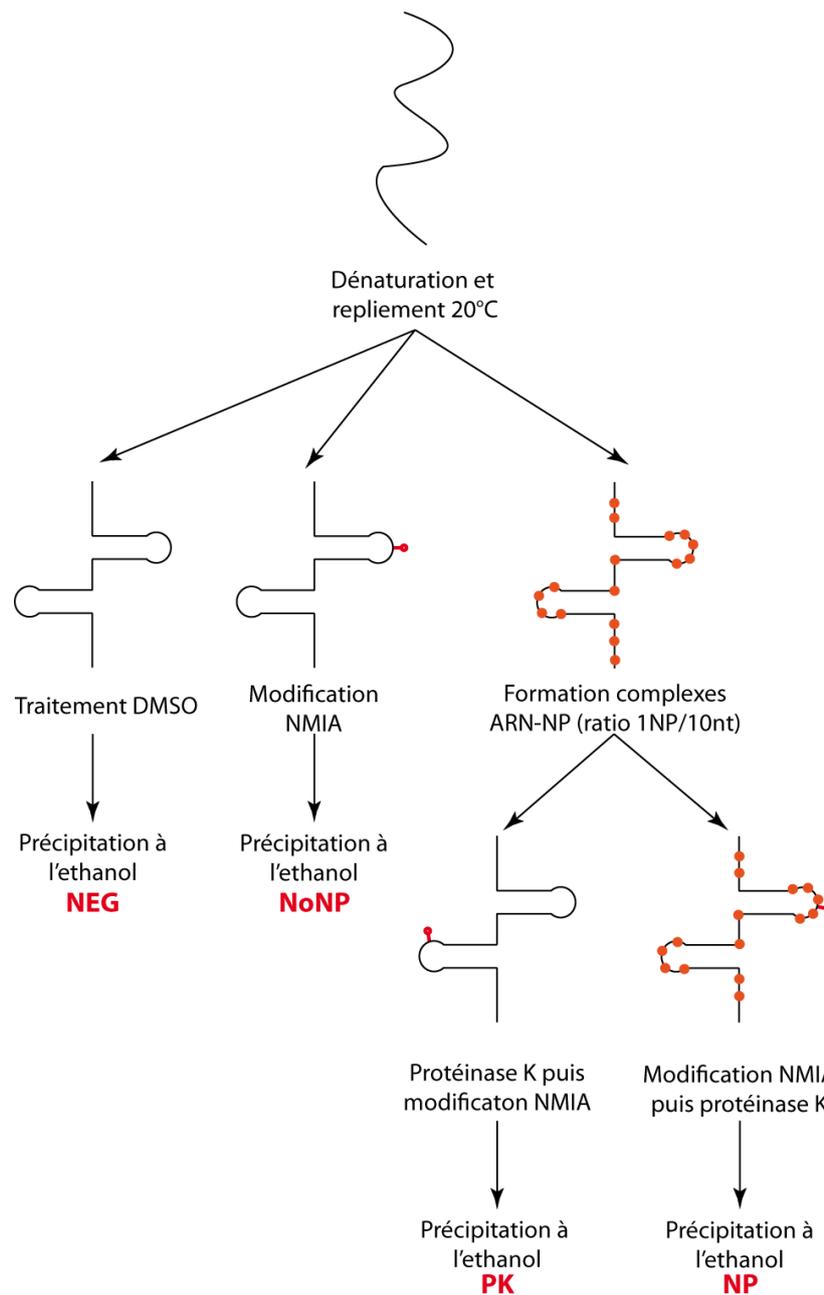


Figure 40 : Conditions utilisées pour la cartographie chimique *in vitro*. L'ARN est replié en absence (**NEG** et **NoNP**) ou en présence (**PK** et **NP**) de la protéine NP sous forme monomérique ou trimérique. La modification ou non (**NEG**) au NMIA est ensuite réalisée directement (**NoNP** et **NP**) ou après déprotéinisation (**PK**).

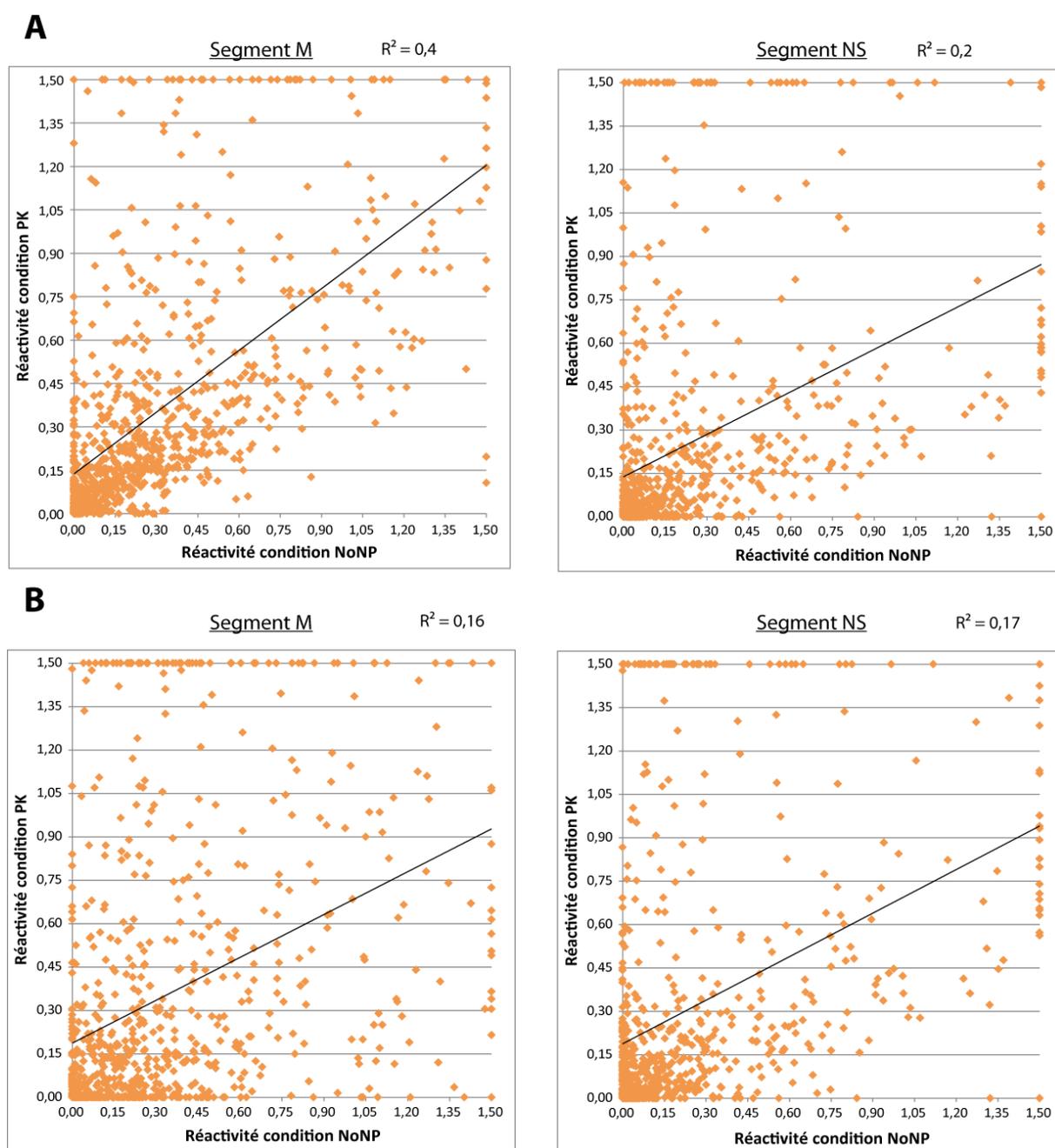


Figure 41 : Représentation de la réactivité de chaque nucléotide pour les ARN M et NS dans les conditions NoNP et PK. Chaque point représente un nucléotide en fonction de sa réactivité après repliement en absence et en présence (A) de NP monomérique ou (B) de NP trimérique. La courbe de tendance linéaire ainsi que son coefficient de corrélation (R^2) sont représentés.

2. Activité chaperon de la protéine NP

Afin de déterminer si la NP possède une activité chaperone de l'ARN, j'ai comparé, la réactivité des ARN NS et M de la souche WSN dans la condition PK, correspondant à l'ARN replié en présence de NP puis déprotéinisé avant d'être modifié chimiquement, à la condition NoNP (**Figure 40**). Cette étude a été menée avec la protéine NP sous forme monomérique ou trimérique. Dans le cas de la protéine monomérique, 703 nucléotides du segment NS et 883 nucléotides du segment M ont été analysés, soit respectivement 79% et 86% des segments NS et M. Le même nombre de nucléotides du segment NS et 826 nucléotides (80%) du segment M ont été analysés avec la protéine trimérique.

Avant de m'intéresser aux structures prédites pour les conditions PK, j'ai dans un premier temps montré que la protéine NP a la capacité de remodeler la structure de l'ARN. Pour cela, j'ai comparé (**Figure 41**), pour chaque nucléotide, les réactivités en condition NoNP (axe des abscisses) et en condition PK (axe des ordonnées). Les nucléotides dont la réactivité augmente dans la condition PK seront retrouvés en haut à gauche du graphique tandis que les nucléotides dont la réactivité diminue dans cette même condition seront retrouvés en bas à droite. Le coefficient directeur de toutes les courbes de tendance est inférieur à 1, ce qui indique une diminution globale des réactivités. Cependant, ces courbes de tendance coupent l'axe des ordonnées au-dessus de 0, ce qui signifie que les nucléotides peu réactifs dans la condition NoNP deviennent légèrement plus réactifs dans la condition PK (d'environ 0,15). Le coefficient de corrélation (R^2) de toutes les courbes de tendance est globalement assez faible (à l'exception du segment M replié avec la NP monomérique), ce qui implique que les réactivités sont très différentes entre les conditions NoNP et PK, reflétant un remodelage important de la structure par la protéine NP, la réactivité des nucléotides étant directement liée à leur flexibilité au sein de la structure.

J'ai ensuite comparé les structures pour chacun des segments en condition chaperonnée ou non en fonction de l'état d'oligomérisation de la protéine NP. Je me suis d'abord intéressé à la stabilité de ces structures dans les différentes conditions à l'aide des données d'énergie libre obtenues par le logiciel RNAstructure. J'ai choisi, pour chaque condition, la structure la plus stable en tenant compte des contraintes SHAPE, puis la stabilité a été recalculée sans les termes de pseudo-énergie associés aux contraintes SHAPE et biaisant les données d'énergie libre (**Figure 42**). Les structures

formées en présence de la protéine NP trimérique sont moins stables (-254,4 et -229,2 kJ/mol pour les ARN M et NS respectivement) qu'en absence de protéine (-263,6 et -232,2 kJ/mol), et celles formées en présence de NP monomérique (-265,5 et -239,1 kJ/mol) sont marginalement plus stables que celles formées en absence de protéine. La NP monomérique, mais pas la NP trimérique, pourrait avoir une faible activité chaperonne d'ARN. Cependant, le rôle d'une protéine chaperonne de l'ARN est de permettre le repliement de la molécule dans la conformation active, sans pour autant obligatoirement augmenter la stabilité thermodynamique [95]. Le repliement en présence de la protéine monomérique ou trimérique induisant un changement global de la réactivité des nucléotides (**Figure 41**), les deux états d'oligomérisation de la NP possèdent une activité chaperonne. Dans le cas des IVA, la protéine NP a un rôle de protection du génome contre la dégradation, permet l'import du génome dans le noyau et peut, par sa fixation, exposer des régions de l'ARNv à l'extérieur afin de promouvoir des interactions ARN-ARN. Cependant, d'après mes résultats, la protéine NP n'augmente pas la stabilité des structures des ARNv de manière importante.

	NoNP	PK monomérique	PK trimérique
Segment M	-263,6	-265,5	-254,4
Segment NS	-232,2	-239,1	-229,2

Figure 42 : Stabilité des structures dans les conditions NoNP et PK en présence de protéine monomérique ou trimérique pour les segments M et NS exprimée en kJ/mol.

Pour compléter ces résultats j'ai comparé la distribution de la taille des hélices entre les structures obtenues avant et après effet chaperon de la protéine NP sous forme monomérique ou trimérique (**Figure 43**). Pour cela, j'ai utilisé les structures de chaque condition dont les énergies libres ne diffèrent pas de plus de 10% et j'ai analysé la distribution moyenne de la taille des hélices au sein de ces structures. Pour le segment M (**Figure 43A**), on observe que les structures remodelées par la protéine NP présentent une quantité d'hélices longues plus importantes que les structures non remodelées. Les hélices composées d'au moins sept paires de base correspondent respectivement à 20% et 21% des hélices pour les structures remodelées avec la protéine NP monomérique et trimérique, contre 16% pour les structures non remodelées. Les mêmes résultats sont obtenus avec l'ARN NS (**Figure 43B**), avec une proportion d'hélices composées d'au moins sept paires de base de 16% et 15% pour les structures remodelées avec la protéine NP monomérique et trimérique respectivement,

contre 12% pour les structures non remodelées. Ces résultats montrent que la protéine NP, indépendamment de son état d'oligomérisation, induit la formation de plus d'hélices longues, soit une stabilisation des structures qui n'est pas visible par l'analyse des paramètres thermodynamiques seuls.

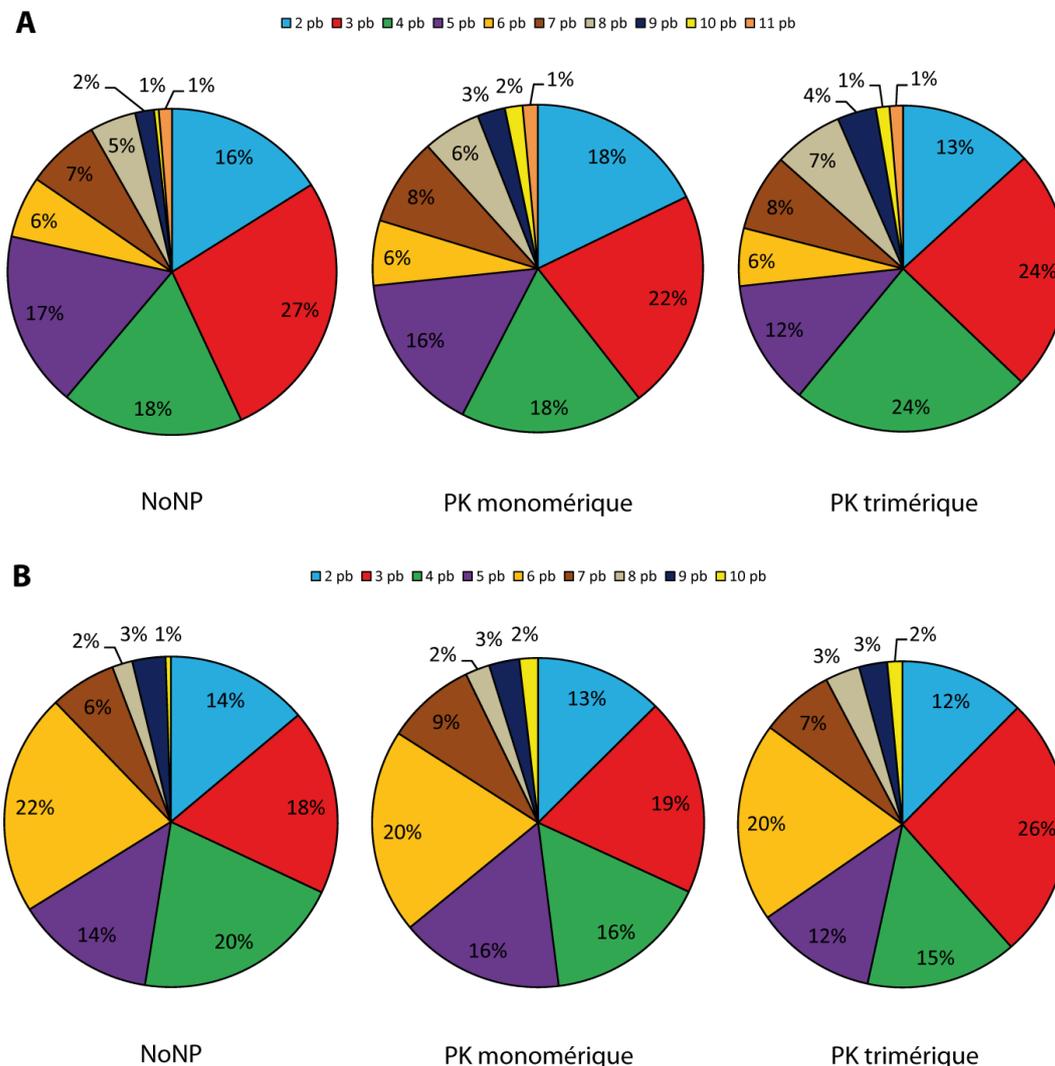


Figure 43 : Répartition de la taille des hélices dans les structures des ARN (A) M et (B) NS dans les conditions de repliement en absence (NoNP) ou en présence (PK) de protéine NP sous forme monomérique ou trimérique.

J'ai ensuite comparé les différentes structures obtenues lors des expériences de SHAPE *in vitro* et cette analyse permet de conclure à l'impact de la fixation de la NP sur le repliement des ARN. Les structures complètes des ARN NS et M, chapéronnés avec les protéines NP monomériques et trimériques sont présentés en Annexe (Annexes 5 à 8) ; certaines régions de ces structures sont analysées plus en détails ci-dessous.

Pour l'ARN M, dans toutes les conditions, on observe une structure très ramifiée, composée de nombreuses tiges-boucles et boucles internes. En comparant les trois structures ([Annexes 3, 5 et 6](#)), on observe que certaines structures sont identiques dans les différentes conditions. La région correspondant aux nucléotides 567 à 670 par exemple, est similaire entre les différentes conditions bien que plusieurs différences soient observées ([Figure 44](#)). En effet, cette région est formée d'une boucle interne d'où partent trois tige-boucles. On observe, pour cette région, que la structure de l'ARN chaperonné par la protéine monomérique ([Figure 44B](#)) est identique à la structure de l'ARN libre ([Figure 44A](#)), indiquant que cette protéine n'a pas d'effet chaperon sur cette structure.

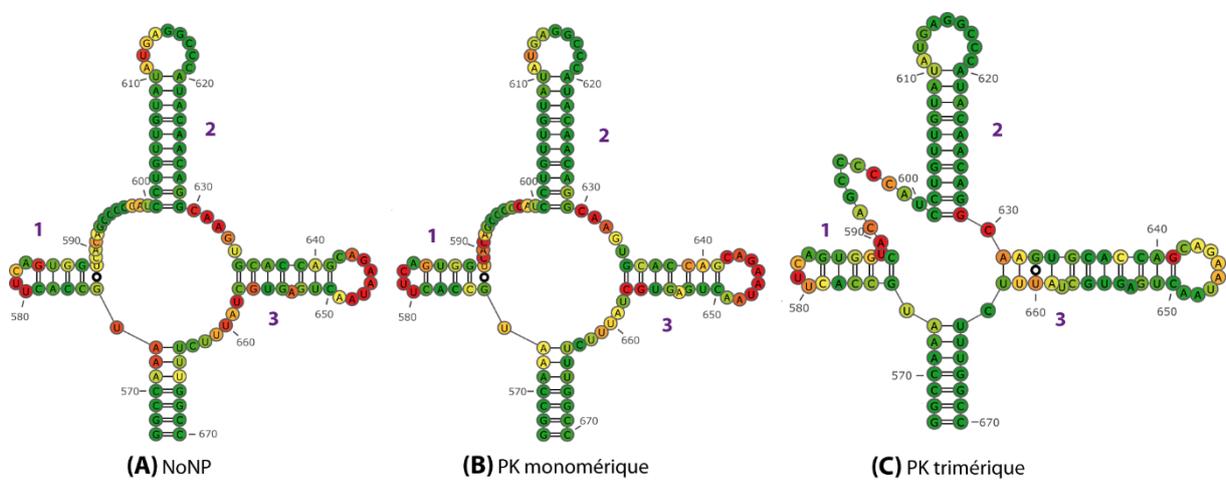


Figure 44 : Structure secondaire de la région 567-670 de l'ARN M de la souche WSN dans les conditions NoNP (A), PK monomérique (B) et PK trimérique (C). Les tiges-boucles sont numérotées.

En outre, la protéine NP trimérique ([Figure 44C](#)) modifie légèrement la structure, au niveau de la première et de la troisième tige boucle. La première tige boucle voit la première paire de base $G_{575} \cdot U_{588}$ être remplacée par un appariement $G_{575} - C_{589}$, tandis que le nucléotide U_{588} forme un bulge, absent dans les deux autres conditions. La troisième tige-boucle est composée de sept paires de base et d'un bulge dans les conditions NoNP et PK monomérique, tandis que dans la condition PK trimérique elle est composée de onze paires de base et de deux bulges. Sur chaque structure présentée on observe que les nucléotides non appariés, notamment au niveau des boucles apicales, sont réactifs, à l'exception des nucléotides 615 à 619 formant la séquence GGCCC. Une interaction intramoléculaire entre cette région et une autre région simple brin dans la structure pourrait expliquer cette absence de réactivité. De même, une

interaction intermoléculaire de quatre nucléotides est possible entre les séquences GGCCC (nucléotides 615 à 619) de deux ARN M.

Contrairement à la région présentée précédemment, peu impactée par la fixation de la protéine NP, d'autres régions présentent une structure radicalement différente, permettant de conclure à un effet chaperon de la protéine virale. La région comprenant les nucléotides 820 à 900 (**Figure 45**) en est un exemple représentatif. La structure de cette région en condition NoNP (**Figure 45A**) est formée de deux tige-boucles dont une très courte, tandis qu'en condition chaperonnée (**Figure 45B et C**), bien que deux tige-boucles soient également observées, leur structure est différente.

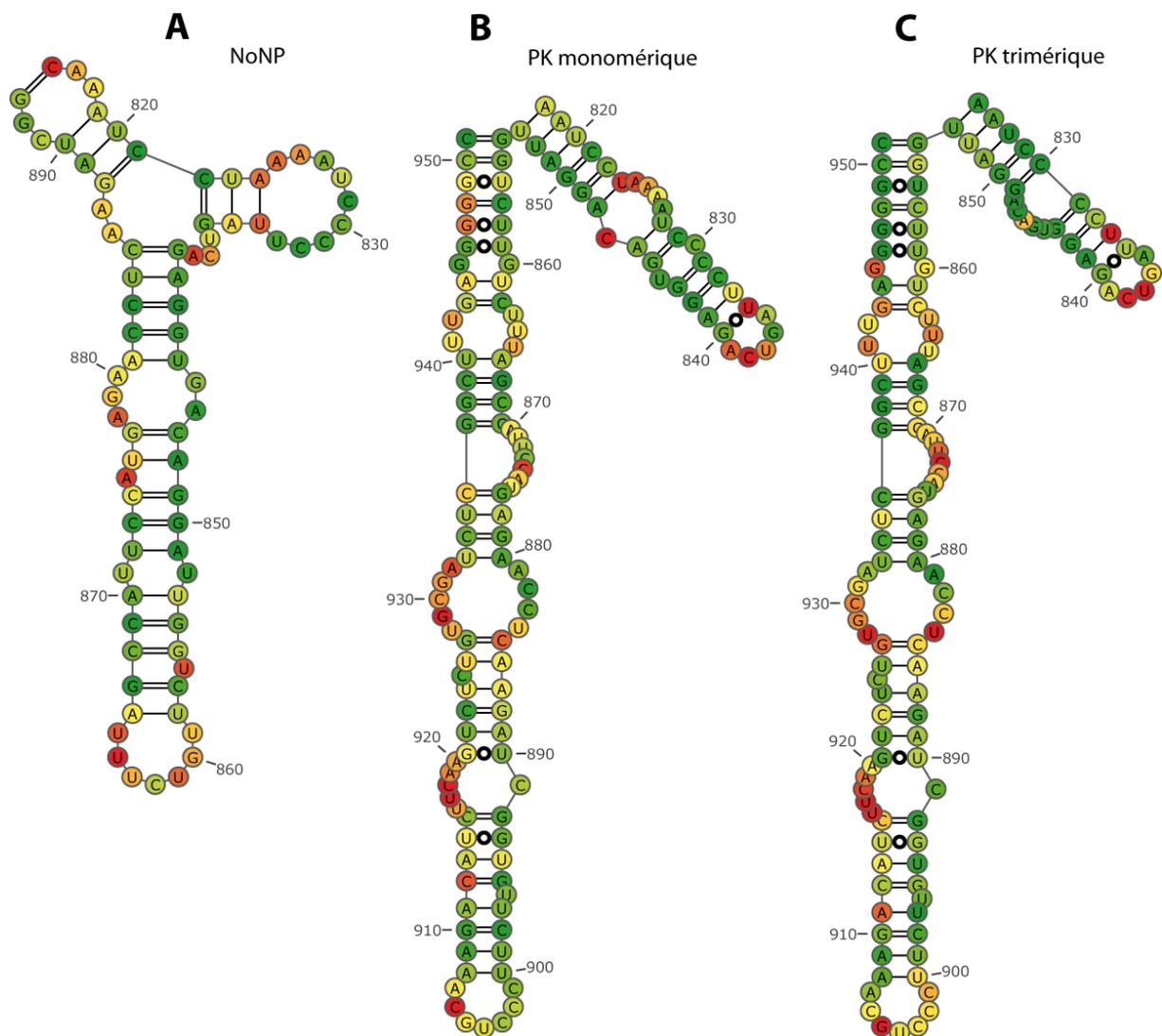


Figure 45 : Structure secondaire de la région 820-900 de l'ARN M de la souche WSN dans les conditions NoNP (**A**), PK monomérique (**B**) et PK trimérique (**C**).

Ces deux tige-boucles sont plus longues que dans la condition NoNP et les appariements impliqués sont totalement différents. Des régions doubles brin en

condition NoNP deviennent simples brin après fixation de la NP (nucléotides 870 à 877 et 880 à 885) et des nucléotides se repliant dans des boucles apicales en absence de protéines se retrouvent dans des boucles internes (régions 825-833 et 859-865), moins accessibles pour des interactions intermoléculaires éventuelles. On observe également que les structures de cette région après effet chaperon des protéines NP trimériques ou monomériques sont presque identiques.

Cependant, les protéines NP monomériques et trimériques n'induisent pas toujours les mêmes changements conformationnels. La région comprenant les nucléotides 20 à 130 présente une structure différente pour chaque condition testée (**Figure 46**). En absence d'effet chaperon de la protéine NP, peu de tige-boucles mais de nombreuses boucles internes sont observées et les régions double brin impliquent des nucléotides éloignés dans la séquence (**Figure 46A**). Le repliement en présence de protéine NP, indépendamment de son état d'oligomérisation, induit la formation de nombreuses tiges-boucles (**Figure 46B et C**).

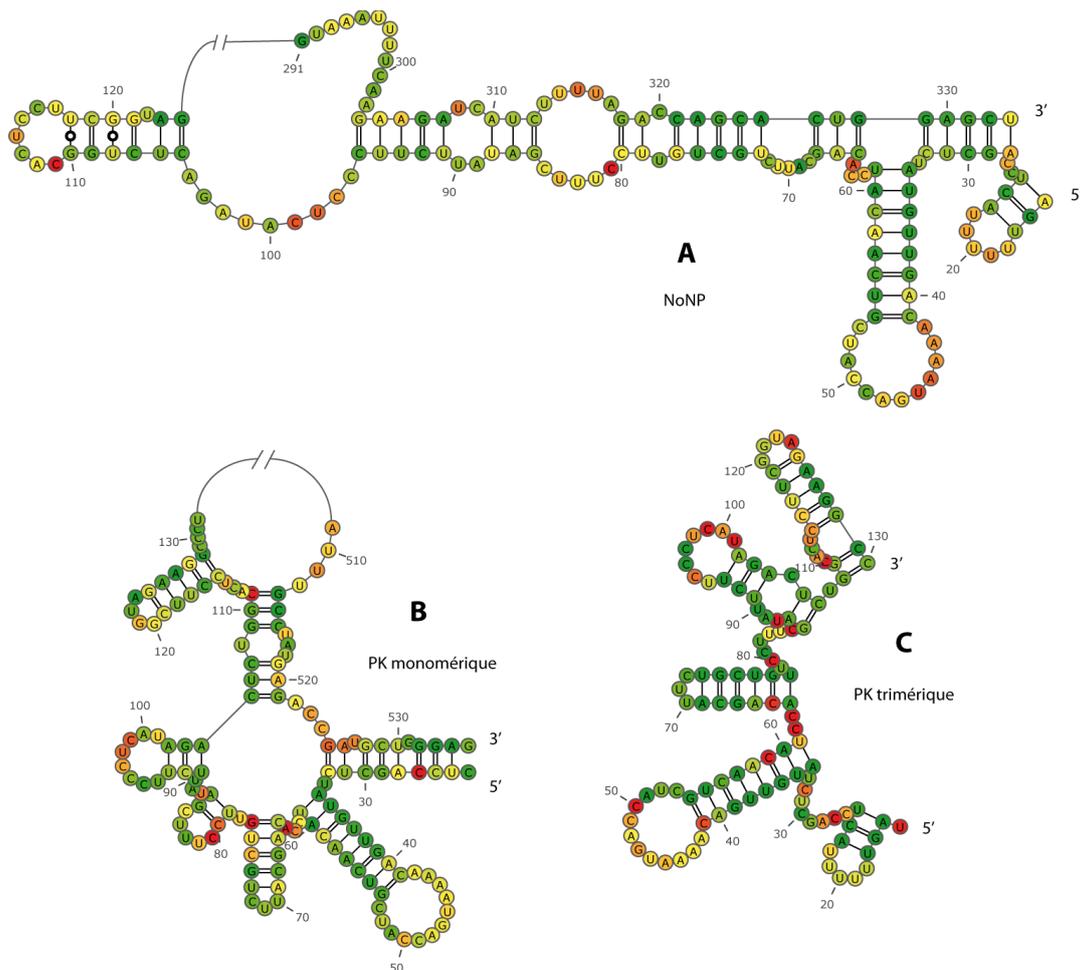


Figure 46 : Structure secondaire de la région 30-130 de l'ARN M de la souche WSN dans les conditions NoNP (A), PK monomérique (B) et PK trimérique (C).

Certaines de celles-ci sont communes entre les conditions PK monomérique et PK trimérique (entre les nucléotides 35 et 77) mais plusieurs différences majeures sont observées. Dans le cas de la condition PK monomérique (**Figure 46B**), les nucléotides 24 à 33 sont appariés tandis qu'ils sont en simple brin dans la condition PK trimérique (**Figure 46C**). La tige boucle formée par les nucléotides 79 à 87 dans la condition PK monomérique est remplacée par une région en simple brin dans la condition PK trimérique. De plus, les appariements mis en jeu dans la condition PK trimérique font intervenir des nucléotides proches dans la séquence tandis que plusieurs appariements entre nucléotides éloignés dans la séquence existent dans la condition PK monomérique.

J'ai réalisé la même analyse sur le segment NS et les mêmes cas de figures qu'avec le segment M sont observés (**Annexes 4, 7 et 8**). Certaines régions de l'ARN replié en présence de NP présentent la même structure que l'ARN libre, indiquant que ces régions sont insensibles à l'effet chaperon de la protéine, leur stabilité étant peut-être trop forte pour un réarrangement structural. D'autres régions présentent des changements structuraux après fixation de la NP, mais ces changements dépendent de l'état d'oligomérisation de la protéine NP.

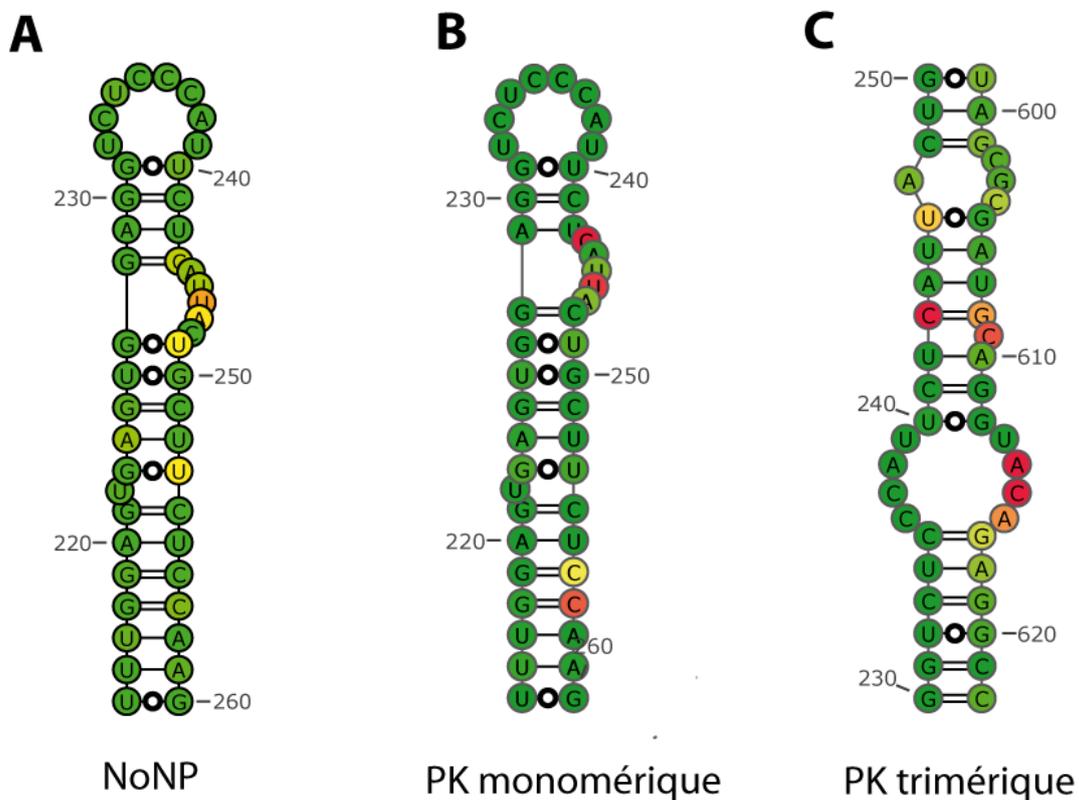


Figure 47: Structure secondaire de la région 215-260 de l'ARN NS de la souche WSN dans les conditions NoNP (A), PK monomérique (B) et PK trimérique (C).

La région 215-260 (**Figure 47**) forme une tige boucle dans la condition NoNP (**Figure 47A**) et cette structure n'est pas modifiée après fixation de la NP monomérique (**Figure 47B**). Cependant, l'ajout de la protéine NP trimérique (**Figure 47C**) réarrange totalement la structure de cette région, avec l'apparition d'appariements impliquant des nucléotides éloignés dans la séquence et la disparition de la boucle apicale. Malgré ces différences majeures, la région 236-239 reste en simple brin malgré l'absence de réactivité de tous les nucléotides comme indiqué par le code couleur. Ces nucléotides forment la séquence CCAU et peuvent former une interaction intramoléculaire de trois paires de base avec les nucléotides GUA (région 270-272) localisés dans une boucle apicale et dont les réactivités sont très faibles, expliquant l'absence de réactivité de ces nucléotides.

A l'inverse, la structure de la région 538-557 (**Figure 48**) est radicalement modifiée après fixation de la NP monomérique. On observe en effet que dans la condition NoNP cette région forme une tige-boucle (**Figure 48A**) tandis que dans la condition PK monomérique cette région est prise dans une boucle interne encadrée par deux hélices (**Figure 48B**).

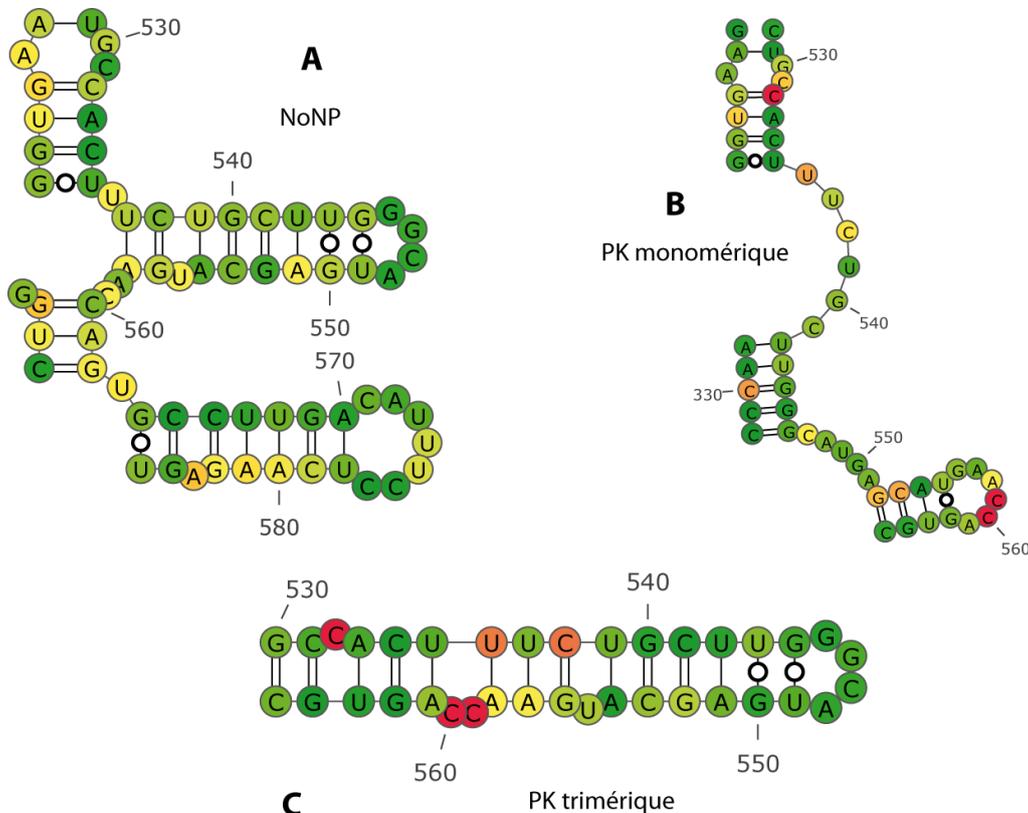


Figure 48 : Structure secondaire de la région 538-557 de l'ARN NS de la souche WSN dans les conditions NoNP (A), PK monomérique (B) et PK trimérique (C).

Après fixation de la NP trimérique, cette région garde sa structure en tige-boucle (**Figure 48C**) mais n'est pas rigoureusement identique à la condition NoNP. En effet, dans la condition PK trimérique, la tige boucle est plus longue, avec deux nouvelles hélices de deux et trois paires de bases séparées par un bulge d'un nucléotide.

Il y a donc un léger effet chaperon de la protéine NP trimérique sur cette région, qui permet de proposer une meilleure exposition de la tige-boucle à l'extérieur de la structure, mais la structure ne présente pas de changement drastique en comparaison à la condition NoNP. En outre, les différences de réactivité entre les conditions PK monomérique et trimérique sont très faibles, il est donc possible que la différence de structure observée ne soit due qu'au logiciel utilisé et que cette région soit repliée de la même façon dans ces deux conditions. A nouveau, la boucle apicale de la structure des conditions NoNP et PK trimérique ne présente pas de réactivité, ce qui peut s'expliquer par une interaction intramoléculaire. La séquence $_{545}\text{GGCA}_{548}$ de cette boucle peut interagir avec la séquence $_{697}\text{UGUU}_{700}$ (dont la réactivité est faible) qui est en configuration simple brin dans la condition NoNP ou en interaction proches d'une boucle interne dans la condition PK trimérique et présentant deux appariements GU. Dans cette dernière condition, une interaction avec la séquence $_{190}\text{UGUU}_{193}$ se situant à la base d'une boucle apicale et présentant un appariement GU, est aussi envisageable.

3. Sites de fixation de la protéine NP

Afin de mieux caractériser la fixation de la protéine NP à l'ARN, je me suis intéressé aux sites de fixation de la protéine, dans le but d'éventuellement trouver une séquence ou une structure consensus de fixation. J'ai dans un premier temps réalisé la même analyse que précédemment (voir §III.D.2.) afin de visualiser globalement l'évolution de la réactivité de chaque nucléotide entre les conditions NoNP ou PK et la condition NP (**Figure 40**), qui correspond à la modification par le réactif SHAPE du complexe ARN-NP (**Figures 49** et **50**). Pour la comparaison entre les conditions NoNP et NP, on observe avec ces graphiques que la réactivité des nucléotides dans la condition NP a tendance à diminuer par rapport à leur état initial dans la condition NoNP.

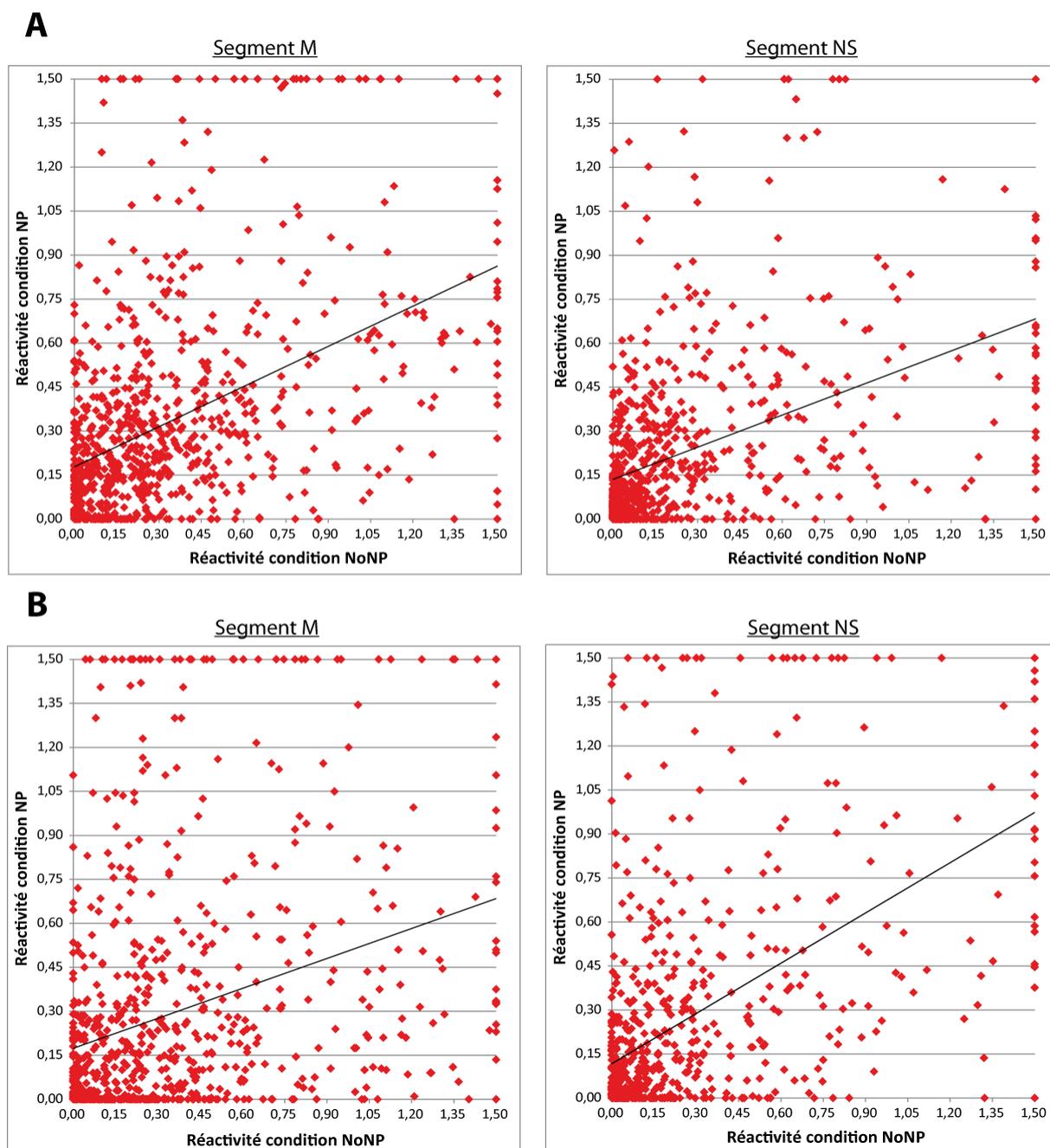


Figure 49 : Représentation de la réactivité de chaque nucléotide pour les ARN M et NS dans les conditions NoNP et NP. Chaque point représente un nucléotide en fonction de sa réactivité après repliement en absence de NP et en complexe avec (A) la NP monomérique ou (B) la NP trimérique. La courbe de tendance linéaire est représentée.

En présence de la protéine monomérique (**Figure 49A**), on observe l'effet le plus important pour le segment NS, avec un coefficient directeur de la courbe de tendance plus petit que pour le segment M. Ce résultat s'inverse en présence de la protéine trimérique (**Figure 49B**), avec un plus grand nombre de nucléotides dont la réactivité diminue en présence de NP pour le segment M comparé au segment NS.

La comparaison entre les conditions PK et NP montre une tendance similaire, avec une diminution de réactivité des nucléotides en présence de NP.

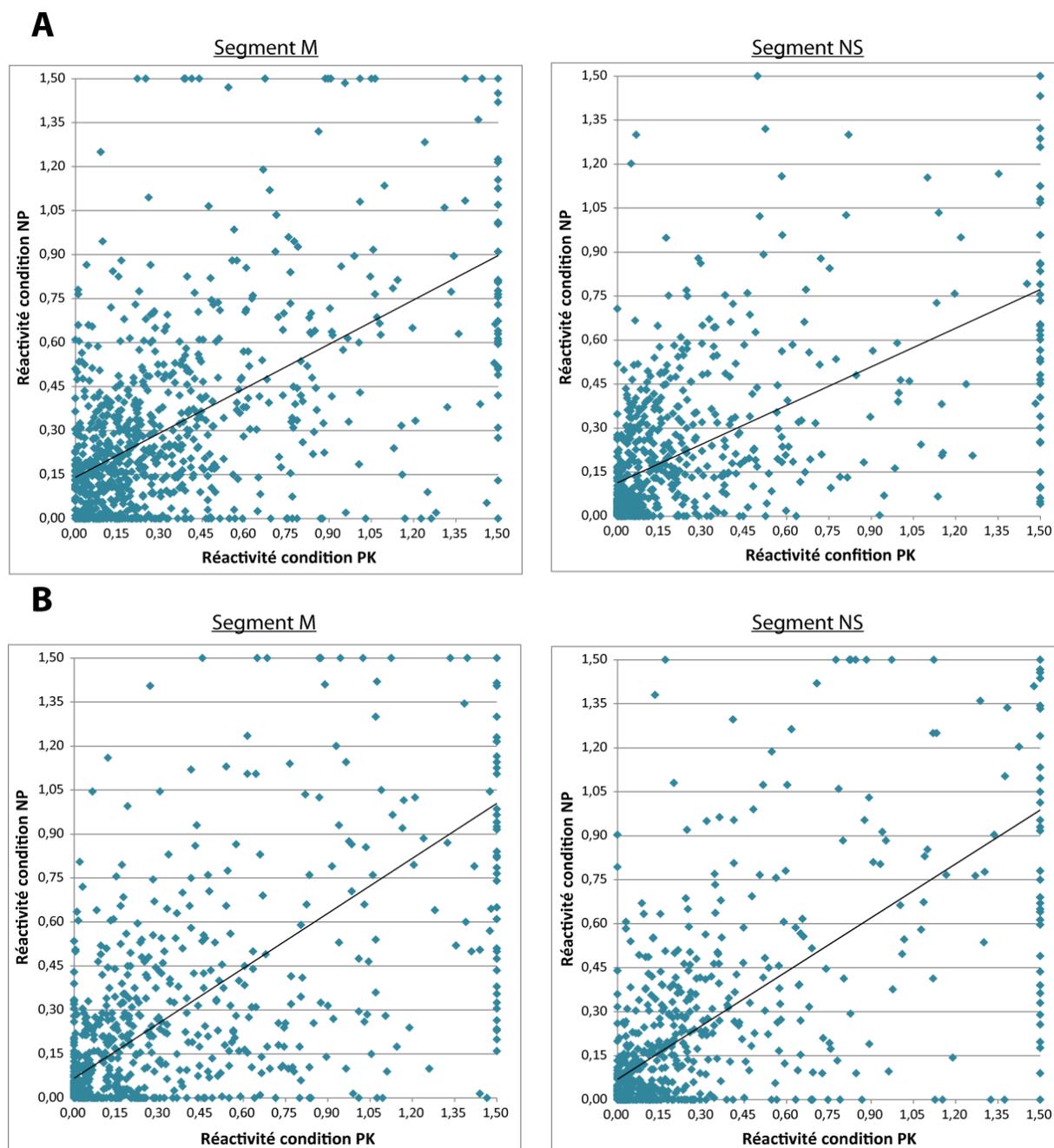


Figure 50 : Représentation de la réactivité de chaque nucléotide pour les ARN M et NS dans les conditions PK et NP. Chaque point représente un nucléotide en fonction de sa réactivité après (A) repliement en présence de NP monomérique (abscisse) et en complexe avec la NP monomérique (ordonnée) ou (B) repliement en présence de NP trimérique (abscisse) et en complexe avec la NP trimérique (ordonnée). La courbe de tendance linéaire est représentée.

En outre, les différences entre les segments M et NS en fonction de l'état d'oligomérisation de la protéine sont moins visibles. On observe tout de même un effet plus important de la protéine monomérique (**Figure 50A**) sur la baisse de réactivité

des nucléotides dans le segment NS comparé au segment M, mais en présence de protéine trimérique (**Figure 50B**) il n'y a pas de différence entre les deux segments étudiés. Ces résultats confirment l'effet de protection induit par la fixation de la protéine NP sur l'ARN. En comparant l'effet des protéines monomériques et trimériques on observe, dans la majorité des conditions comparées, que les diminutions de réactivité sont plus importantes avec la NP monomérique. L'analyse des coefficients de corrélation (**Figure 51**) des différentes courbes de tendance présentées précédemment montre que les réactivités dans la condition NP diffèrent moins des réactivités de la condition PK que de la condition NoNP.

	<u>NoNP vs NP</u>	<u>PK vs NP</u>
<u>NP monomérique</u>	M : 0,22 NS : 0,21	M : 0,35 NS : 0,37
<u>NP trimérique</u>	M : 0,1 NS : 0,3	M : 0,52 NS : 0,49

Figure 51 : Coefficients de corrélation (R^2) des courbes de tendance présentées dans les Figures 49 et 50.

Je me suis ensuite intéressé aux nucléotides dont la réactivité diminue en présence de NP (la liste des régions identifiées est présentée en **Annexe 9**). L'analyse a été réalisée indépendamment du segment étudié pour les deux états d'oligomérisation de la protéine en comparant, par rapport à la condition PK, la réactivité des nucléotides dans la condition NP, qui correspond au complexe ARN-NP modifié. J'ai choisi de comparer ces deux conditions afin de m'affranchir au maximum de la capacité de remodelage de la NP. En effet, la condition PK représentant l'ARN chaperonné, les diminutions de réactivité doivent correspondre pour la majorité à des sites de fixation de la protéine et non à des réarrangements structuraux. J'ai analysé la fréquence de chaque nucléotide dans les sites de fixation de la protéine NP (WSN et PR8) que j'ai comparé à la composition globale des ARNv NS et M des souches correspondantes (**Figure 52A**). On observe que les sites protégés par la NP monomérique de la souche WSN sont composés à 31% de nucléotides C alors que les ARNv NS et M sont composés à 25% de C, ce qui indique un enrichissement de ce nucléotide dans les sites de fixation de la NP monomérique (**Figure 52A**, en violet dans le panneau de gauche). La composition des sites de fixation de la protéine monomérique en nucléotides A et U est semblable à celle des ARNv M et NS. On remarque par contre un appauvrissement en nucléotide G dans les sites de protection de la protéine monomérique (12%) par

rapport à la composition des ARN (20%). Ces résultats indiquent que la protéine monomérique se fixe préférentiellement sur des régions riches en C et pauvres en G (**Figure 52A**, en violet dans le panel de gauche)). Ces résultats sont exacerbés lors de la l'analyse des sites de fixation de la protéine NP trimérique de la souche WSN (**Figure 52A**, en bleu dans le panel de gauche). On retrouve en effet un enrichissement plus important en nucléotide C dans les sites de fixation en comparaison à la répartition des nucléotides dans les ARN (35% contre 25%). L'appauvrissement en G est également plus marqué dans le cas dans la protéine trimérique, les sites de fixation étant composés à 9% de ce nucléotide, alors qu'il représente 21% des ARNv NS et M.

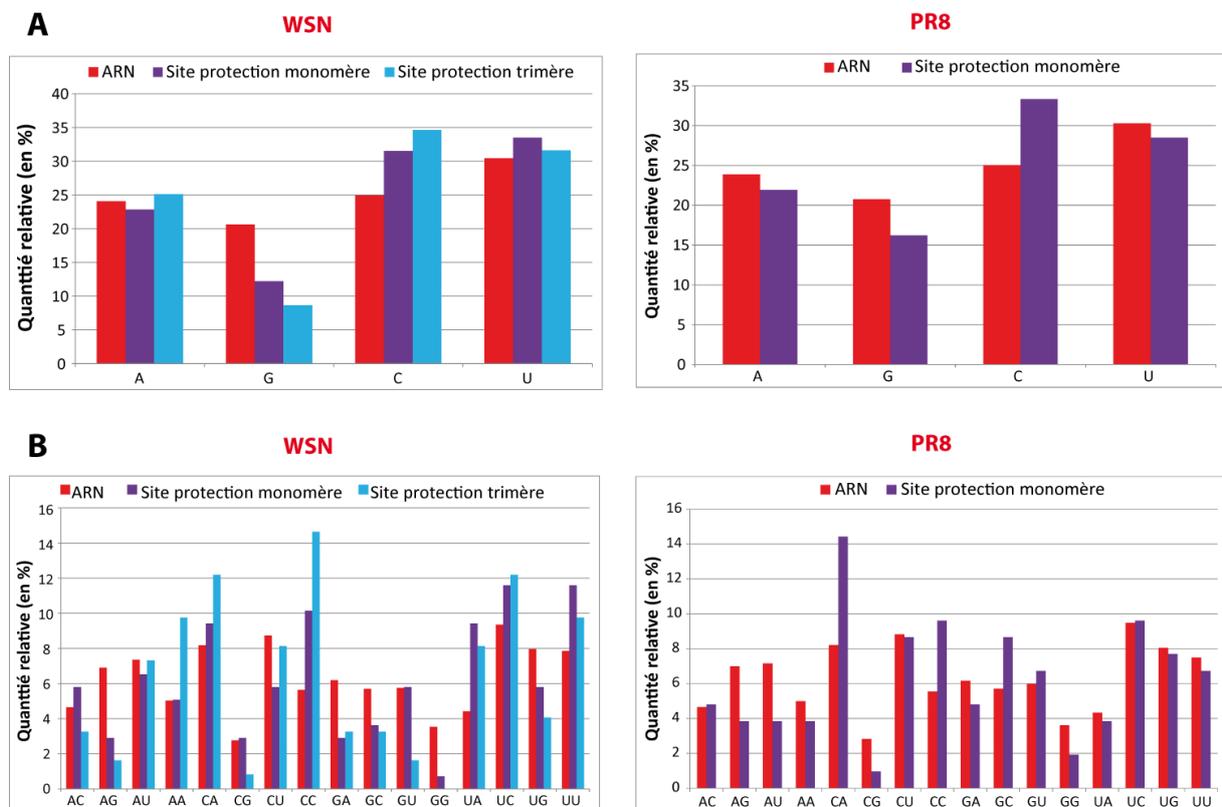


Figure 52 : (A) Fréquence relative des différents nucléotides dans les sites protégés par la protéine NP WSN sous forme monomérique ou trimérique (à gauche) et PR8 (à droite) comparée à la composition globale des ARNv M et NS. (B) Fréquence relative de tous les dinucléotides dans les sites protégés par la protéine NP WSN sous forme monomérique ou trimérique (à gauche) et PR8 (à droite) comparée à la composition globale des ARNv M et NS.

L'analyse réalisée avec la protéine NP PR8 monomérique permet d'obtenir des résultats similaires (**Figure 52B**, panel de droite). En effet, les sites protégés par la protéine NP de la souche PR8 sont enrichis en C (33% contre 25% dans les ARN) et appauvris en G (16% contre 21%). Ces effets sont légèrement plus faibles que pour la souche WSN, notamment l'appauvrissement en G. Cependant, les expériences de

SHAPE avec la protéine de la souche PR8 ayant été réalisées seulement dans les conditions NoNP et NP, les sites de protection relevés peuvent être des régions dont la structure a été réarrangée et qui sont devenues double brin après l'effet chaperon de la protéine NP. Pour confirmer ces résultats, il faudrait étudier la condition PK avec la protéine de la souche PR8, afin de pouvoir déterminer avec précisions les sites de fixation de la NP.

J'ai également analysé la fréquence relative de tous les dinucléotides entre les sites de protection et la totalité des segments NS et M pour les deux souches étudiées (**Figure 52B**). Cette analyse a pour but de déterminer un motif préférentiellement reconnu par la protéine. On observe que les dinucléotides AG et GG, sont peu retrouvés dans les sites de protection, et cela indépendamment de la souche de la protéine NP étudiée ou de son état d'oligomérisation. Le dinucléotide CG est également peu retrouvé avec la protéine NP PR8 (**Figure 52B**, panel de droite) et avec la protéine de la souche WSN mais uniquement sous forme trimérique (**Figure 52B**, en bleu dans le panel de gauche). Le dinucléotide UG est également peu représenté dans les sites protégés par la protéine NP WSN, le résultat étant plus marqué avec la protéine trimérique. Ces résultats sont corrélés à l'appauvrissement en nucléotide G des sites protégés par la protéine NP. En outre, les dinucléotides CA et CC sont beaucoup plus représentés dans les sites de protection pour les souches WSN et PR8. De même, pour la souche WSN, on observe une proportion plus importante des nucléotides UU et UA dans les sites protégés, mais ce résultat n'est pas observé pour la souche PR8.

Je me suis ensuite intéressé aux structures sur lesquelles la protéine se fixe le long de l'ARN. Pour cela j'ai comparé la réactivité des nucléotides entre les conditions PK et NP en reportant les réactivités de la condition NP sur la structure de l'ARN déterminée pour les conditions PK (**Annexes 10 à 13**). Je me suis ensuite intéressé à la comparaison entre les protéines NP monomériques et trimériques sur les deux segments M et NS étudiés. De façon intéressante, et contrairement à l'analyse effectuée pour l'effet chaperon, il semble que la fixation de la protéine NP à l'ARN soit indépendante de son état d'oligomérisation. Ainsi, on observe pour la région 854-951 de l'ARNv M que les sites de protection entre la protéine NP monomérique (**Figure 53A**) et trimérique (**Figure 53B**) sont les mêmes. Les premiers nucléotides de la boucle apicale deviennent moins réactifs, ainsi que les nucléotides formant les boucles internes de cette structure, à l'exception de celle la plus éloignée de la boucle apicale. Il est

intéressant de noter que la protection des premiers nucléotides de la boucle apicale entraîne une augmentation de la réactivité des nucléotides formant l'hélice adjacente, notamment avec la NP trimérique (**Figure 53B**). Des changements conformationnels, tels l'ouverture de régions double brin, peuvent être induits par la fixation de la NP mais sont réversibles une fois la NP enlevée, contrairement aux changements induits par son activité chaperonne. Dans la région étudiée, et dans le cas de la protéine trimérique (**Figure 53B**), les nucléotides 908 à 916 deviennent flexibles suite à la fixation de la protéine NP au niveau de la boucle apicale, et les nucléotides 892 à 899, initialement engagés dans l'hélice restent non réactifs car vraisemblablement protégés par la NP.

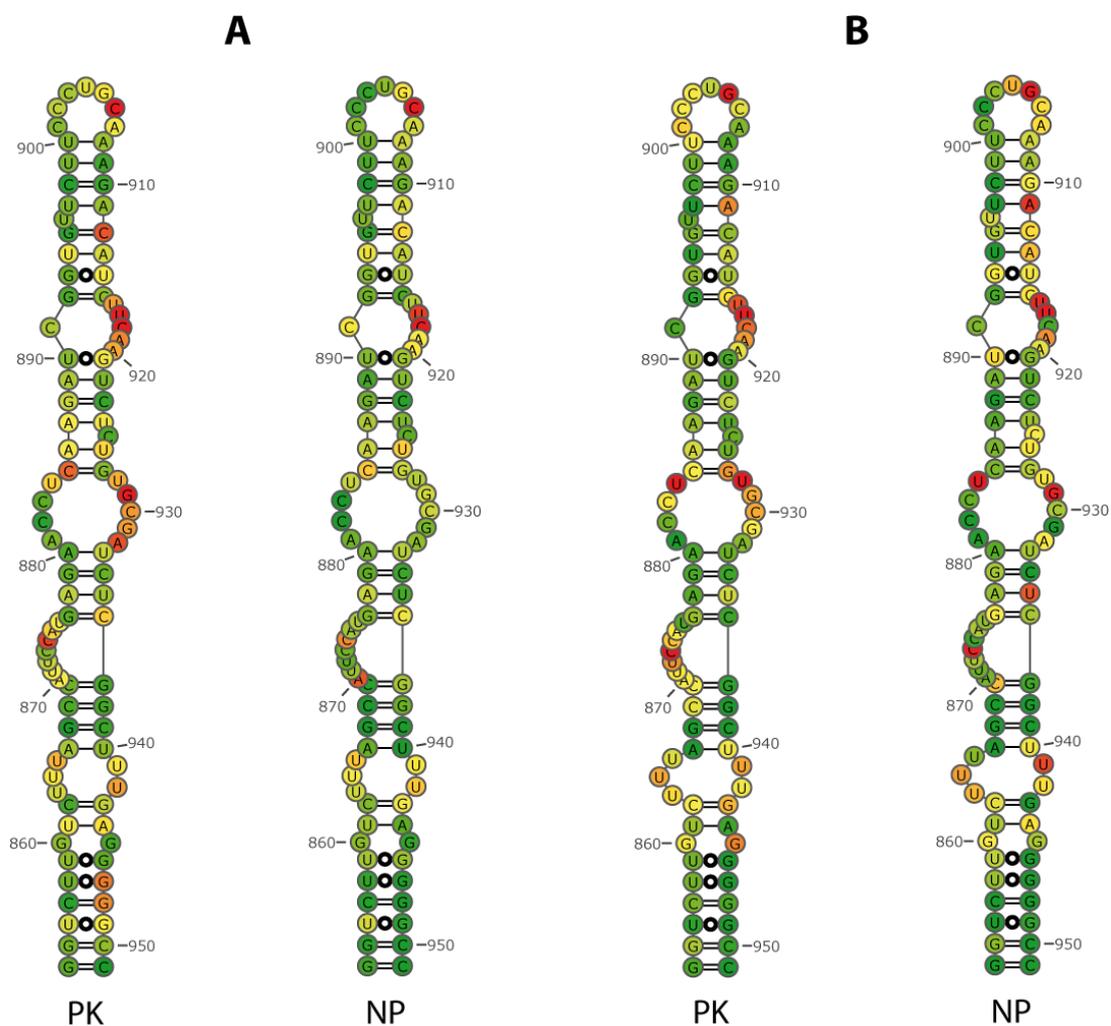


Figure 53 : Comparaison de la réactivité des nucléotides 854 à 951 dans les conditions NP monomérique (**A**) et NP trimérique (**B**). Ces réactivités sont reportées sur la structure secondaire de la région 854-951 de l'ARNv M de la souche WSN déterminée dans les conditions PK associées.

Dans le cas de la protéine NP monomérique (**Figure 53A**), une protection de la boucle apicale est visible sur trois nucléotides mais la fixation de la protéine n'entraîne pas d'ouverture de l'hélice. Les dernières données de la littérature indiquent que la NP se lie à environ douze nucléotides [272], la protection de seulement trois à quatre nucléotides telle que visible ici semble peu probable. Nous pouvons donc supposer que la NP peut également se fixer à des régions double brin adjacentes, mais cette hypothèse est difficile à vérifier si la fixation ne déstabilise pas l'hélice, car dans cette éventualité les régions en double brin seront peu ou pas réactives aussi bien en présence qu'en absence de protéine.

Les caractéristiques décrites précédemment sont retrouvées tout au long de la structure des ARNv NS et M de façon plus ou moins marquée. Dans la région 567-670 de l'ARNv M (**Figures 54** et **55**), précédemment étudiée pour l'effet chaperon de la protéine NP, on observe que la protéine NP monomérique protège, en se fixant, les nucléotides 658 à 663, ce qui entraîne une augmentation de la réactivité de l'hélice basale (**Figure 54C**) qui n'est pas observée pour la condition PK (**Figure 54B**). Pour cette même région, la fixation de la protéine NP trimérique n'est visible qu'au niveau de la boucle apicale comprenant les nucléotides 642 à 649 (**Figure 55B**) et ne semble pas induire l'ouverture de l'hélice adjacente.

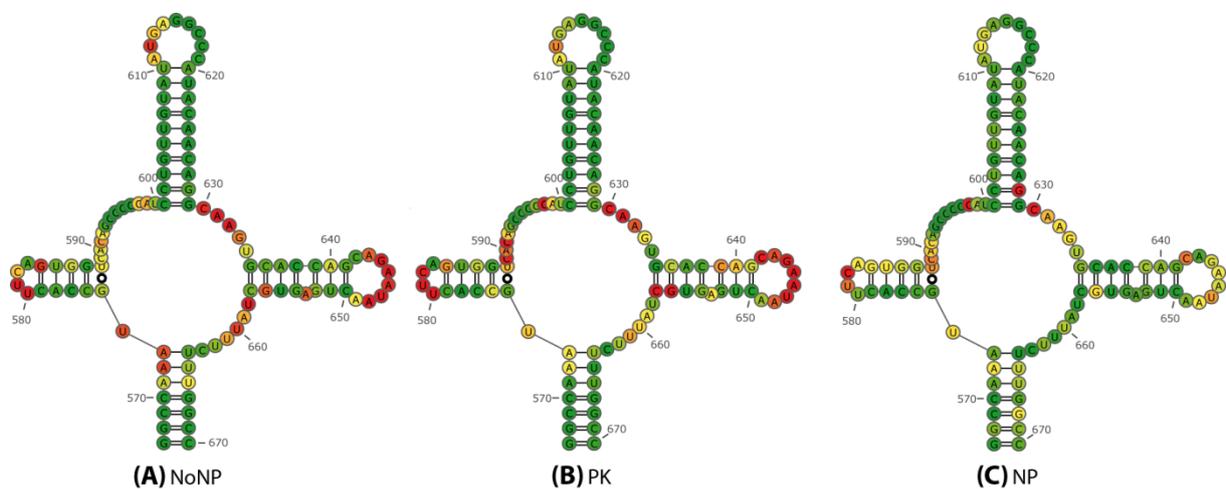


Figure 54 : Comparaison de la réactivité des nucléotides dans les conditions NoNP (A), PK monomérique (B) NP monomérique (C). Les réactivités de la condition NP sont reportées sur la structure secondaire de la région 567-670 de l'ARN M de la souche WSN déterminée dans la condition PK.

Il est intéressant de noter que sur une région d'une centaine de nucléotides un seul site de fixation de la NP est visible en comparant les conditions PK (**Figure 55A**) et

NP (**Figure 55B**), alors que la littérature décrit en moyenne un site de fixation tous les vingt-quatre nucléotides [101,186].

Deux hypothèses peuvent être émises pour expliquer cela : (i) comme proposé précédemment, la NP est capable de fixer les régions double brin et l'analyse effectuée ne nous permet pas de le détecter, ou (ii) cette région correspond à une région appauvrie en NP comme il a été récemment démontré dans la littérature [133]. On peut émettre l'hypothèse que cette région permet l'exposition de ces trois tiges-boucles à l'extérieur de la structure, ce qui augmente la capacité de l'ARNv à interagir avec des structures voisines en vue du mécanisme d'empaquetage sélectif, l'hypothèse d'une région appauvrie en NP pour permettre ces interactions n'est pas à écarter. Cependant, l'utilisation de la NP monomérique montre sur cette même région au moins trois sites de protection, ce qui indiquerait que les régions enrichies ou appauvries en NP dépendent de l'état d'oligomérisation de cette protéine.

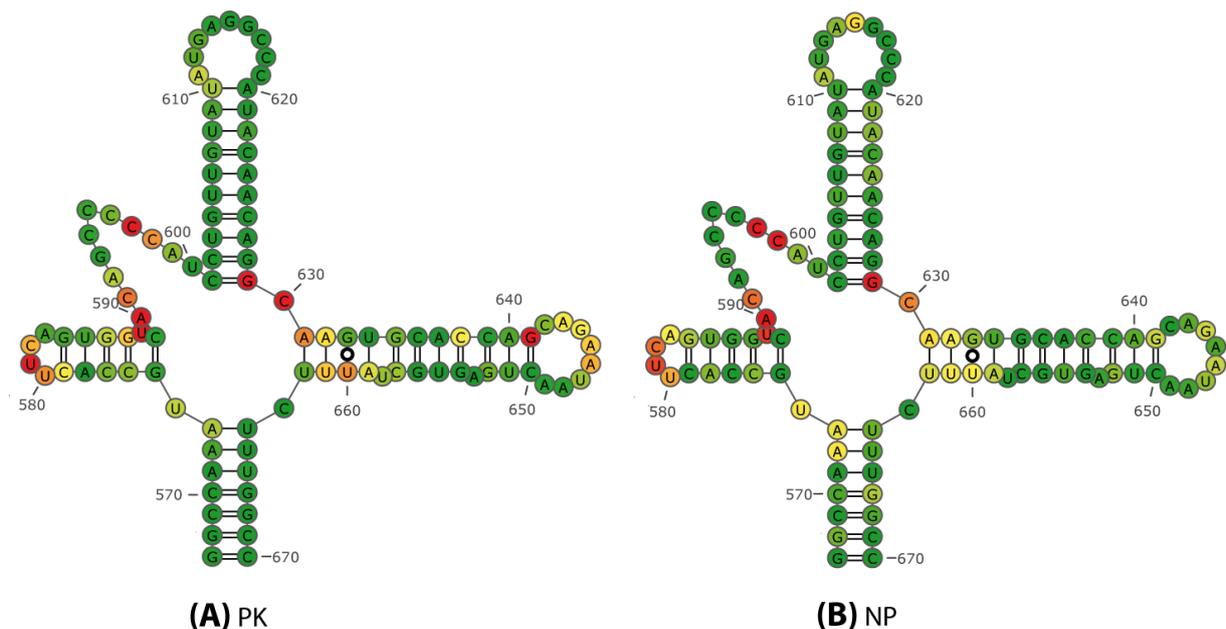


Figure 55 : Comparaison de la réactivité des nucléotides 567-670 dans les conditions PK trimérique (A) et NP trimérique (B). Les réactivités de la condition NP sont reportées sur la structure secondaire de la région 567-670 de l'ARNv M de la souche WSN déterminée dans la condition PK.

L'analyse de la diminution de réactivité de certaines régions des ARNv étudiés semble indiquer que la NP est capable de lier des régions doubles brin. L'analyse de la région 800-823 de l'ARNv NS (**Figure 56**), formant une tige boucle séparée en son centre par une boucle interne d'un nucléotide, montre, après fixation de la protéine NP monomérique (**Figure 56B**) une baisse de réactivité des nucléotides formant la boucle

ainsi que des nucléotides appariés encadrant cette boucle en comparaison avec la condition PK (**Figure 56A**). Il est malgré tout compliqué de conclure à une fixation sur des régions double brin avec ce type de résultat, les nucléotides impliqués dans des appariements étant fortement réactifs autour de la boucle interne, indiquant une flexibilité importante. Il est donc possible que ces appariements soient artificiellement créés par le logiciel RNAstructure en privilégiant les paramètres thermodynamiques par rapport aux contraintes SHAPE. J'ai également émis l'hypothèse d'une fixation de la protéine NP à des régions double brin de l'ARN après avoir obtenu les premiers résultats de cartographie chimique avec la protéine NP PR8. Ces résultats montrent que de nombreuses régions protégées sont très courtes (de un à quatre nucléotides), comparées à la taille minimale d'oligoribonucléotides permettant une fixation optimale de la NP [272]. Cependant, les résultats obtenus avec la protéine PR8 ne correspondent qu'aux conditions NoNP et NP, rendant difficile la différenciation entre les sites de fixation de la protéine et les réarrangements structuraux induits par l'activité chaperon de la NP. L'utilisation de la protéine WSN dans les conditions PK et NP m'ont permis de discriminer ces deux actions et me permettent de proposer la capacité de la protéine NP à lier les régions double brin de l'ARN.

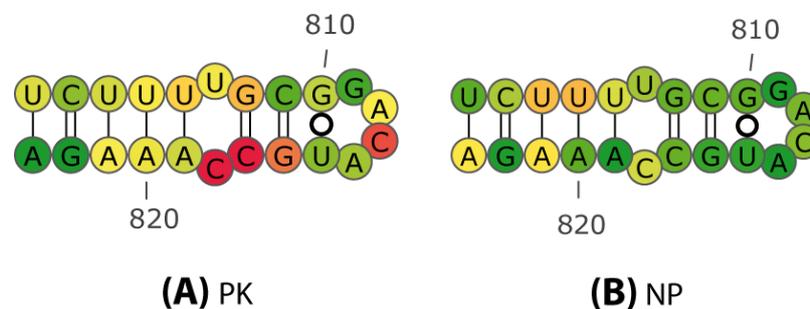


Figure 56 : Comparaison de la réactivité des nucléotides 800-823 dans les conditions PK monomérique (A) NP monomérique (B). Les réactivités de la condition NP sont reportées sur la structure secondaire de la région 800-823 de l'ARNv NS de la souche WSN déterminée dans la condition PK.

J'ai également comparé la réactivité des nucléotides entre les conditions NoNP et NP en reportant les réactivités de la condition NP sur la structure de l'ARN déterminée pour les conditions NoNP (**Annexes 14 à 17**). Avec cette analyse on observe à plusieurs endroits une augmentation de réactivité de régions double brin (**Figure 57**). De façon intéressante, cette augmentation de réactivité est beaucoup plus marquée avec la protéine monomérique que trimérique. Ainsi, la région 336-343 de l'ARNv M forme

une hélice avec la région 996-1003 bien que la réactivité de cette dernière région ne soit pas connue.

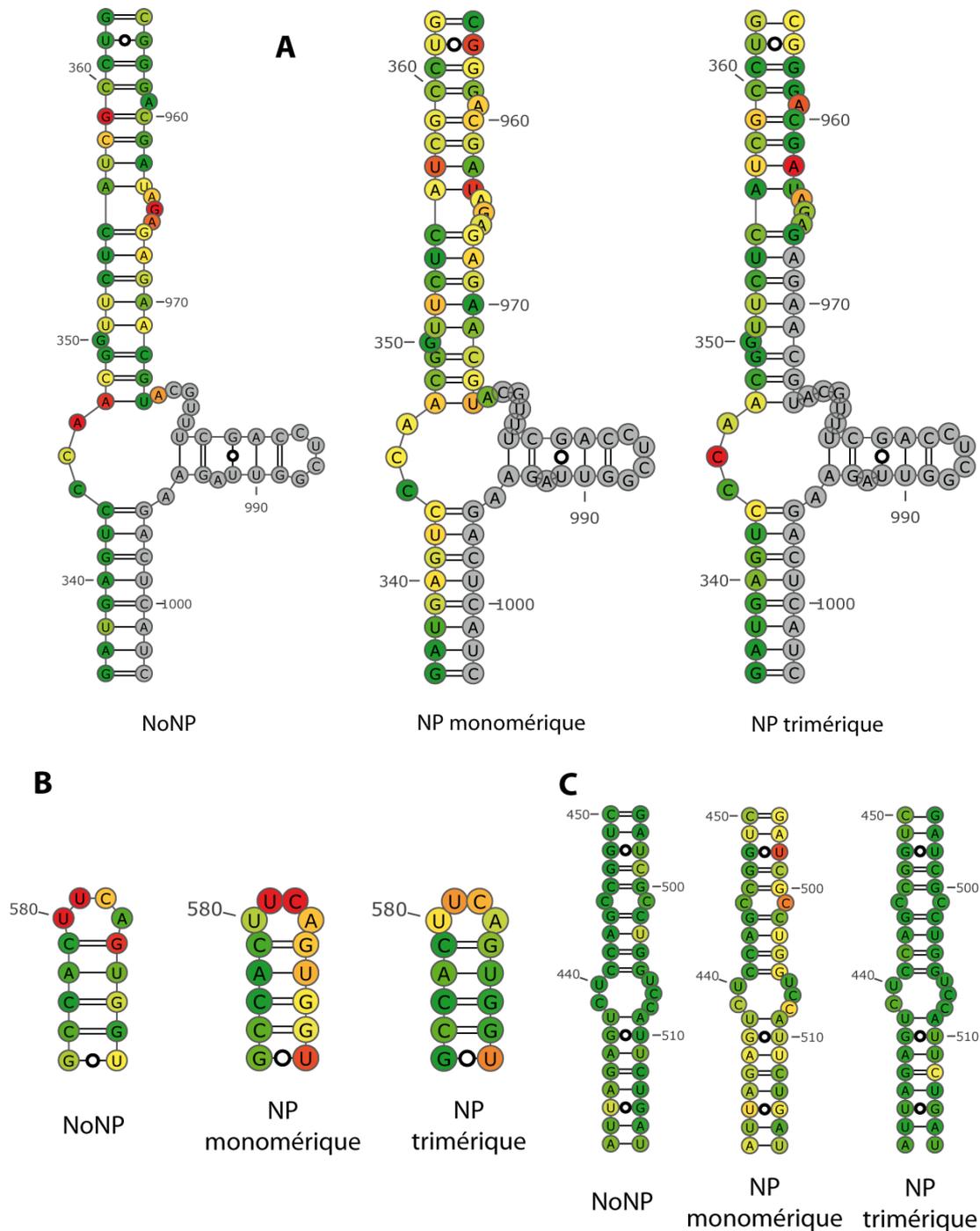


Figure 57 : Comparaison de la réactivité des nucléotides dans les conditions NoNP et NP monomérique et trimérique. Les réactivités de la condition NP sont reportées sur les structures secondaires des régions 336-363 (A) et 575-588 (B) de l'ARN M et 431-516 (C) de l'ARNv NS de la souche WSN déterminée dans la condition NoNP.

En présence de protéine NP monomérique on observe une augmentation de la réactivité des nucléotides 339 à 343 alors que l'utilisation de la protéine trimérique ne

modifie pas la réactivité de cette région (**Figure 57A**). D'autres exemples sont observés, aussi bien sur l'ARNv M (**Figure 57B**) que l'ARNv NS (**Figure 57C**), où seule la protéine NP monomérique induit une augmentation de la réactivité de la région alors qu'en présence de protéine trimérique le profil est presque identique à la condition NoNP. Contrairement aux résultats précédents, j'ai ici comparé la réactivité des nucléotides en condition NP monomérique et trimérique sur la même structure, permettant une analyse plus globale sur l'impact de l'état d'oligomérisation de la protéine vis-à-vis de sa fixation sur l'ARN. Ces nouveaux résultats permettent de conclure que la NP monomérique induit des changements de réactivité plus importants sur les ARNv NS et M que la protéine NP trimérique, indépendamment de leur propriété chaperonne.

CONCLUSION
&
DISCUSSION

Conclusions et Discussion

Au cours de ce travail de thèse, je me suis intéressé à la structure des ARN viraux des *Alphainfluenzavirus* en utilisant la technique de cartographie chimique en solution SHAPE. La majeure partie de mon travail a été focalisée sur la caractérisation de l'effet de la protéine virale NP sur la structure des ARNv transcrits *in vitro*. La première partie de ce travail m'a permis de définir les conditions d'induction, de purification et d'utilisation de la protéine NP. J'ai en effet déterminé que l'induction de l'expression de la protéine NP dépend de la souche virale. Mes premiers essais avec la protéine NP PR8 n'ont jamais donné lieu à une induction significative, tandis qu'avec la souche WSN, l'utilisation de bactéries capables de décoder les codons rares permet une induction massive de l'expression. Il est intéressant de noter que ces deux protéines présentent une identité de séquence de 96%, soit vingt résidus différents. Une différence d'induction entre ces deux protéines est donc surprenante. Parmi ces résidus différents, deux d'entre eux se trouvent dans la partie N-terminale de la protéine, aux positions 3 et 4. Or, les résultats sur gel SDS-PAGE, après purification de la protéine de la souche PR8, indiquent une dégradation de la protéine, vraisemblablement depuis sa partie N-terminale. Ces deux acides aminés de différence (SQ pour PR8 et TK pour WSN) pourraient être à l'origine de cette dégradation et dont l'absence d'induction. Pour confirmer cette hypothèse, il faudrait remplacer les deux résidus de la souche PR8 par les résidus de la souche WSN et tester l'induction de la production de la protéine.

Les données de la littérature quant à l'état d'oligomérisation de la protéine NP nécessaire pour la fixation à l'ARN n'étant pas claires, j'ai déterminé les conditions salines nécessaires à l'obtention de monomères ou de trimères de protéine. Pour la protéine de la souche WSN, mes résultats montrent qu'une faible concentration en NaCl (50 mM) permet de garder la protéine sous forme monomérique, tandis qu'une augmentation de cette concentration à 150 mM conduit à la formation de trimères de protéine, en accord avec la littérature [242]. Cependant, la protéine NP de la souche PR8 semble se comporter différemment, la forme monomérique étant majoritaire même à de fortes concentrations salines (300 mM NaCl), indiquant que l'oligomérisation de la NP est dépendante de la souche virale étudiée. La dernière optimisation que j'ai réalisée avant d'effectuer les expériences de SHAPE *in vitro* a été de déterminer les conditions me permettant de récupérer l'ARNv complexé à la protéine NP WSN. J'ai démontré au cours de cette optimisation, une incompatibilité

lors de l'étape de précipitation à l'éthanol de l'ARN entre le SDS et la protéinase K, menant à la perte de l'ARN après cette étape. L'explication de ces résultats reste compliquée, notamment du fait que les protocoles des fournisseurs de la protéinase K indiquent que l'utilisation de SDS jusqu'à 2% augmente son activité [96]. J'ai également démontré qu'en présence de NMIA, réactif utilisé pour réaliser les modifications de l'ARN, il est impossible de récupérer l'ARN complexé à la NP. Ces résultats pourraient être expliqués par le fait que le NMIA peut réagir avec les groupements SH, NH₂ et OH, portés par les acides aminés de la protéine NP, formant ainsi des thioesters, amides ou esters respectivement. La modification de ce groupement par le NMIA pourrait avoir pour effet de rendre la protéine insensible à la protéinase K et de séquestrer l'ARN qui serait retrouvé dans la phase organique après extraction phénol-chloroforme. Pour pallier à ce problème, j'ai démontré qu'une première étape de précipitation permet d'éliminer l'excès de réactif. Une fois ces conditions établies, j'ai cherché à mettre en évidence une éventuelle activité chaperonne d'ARN de la protéine NP, modifiant la structure de l'ARN de façon irréversible, ainsi qu'à identifier les sites préférentiels de fixation de la protéine sur l'ARNv et ses effets transitoires sur la structure.

J'ai caractérisé et comparé l'activité chaperonne des protéines NP monomériques et trimériques de la souche WSN sur les ARNv NS et M de cette même souche. J'ai pour cela modifié chimiquement, à l'aide du réactif NMIA, les ARNv transcrits *in vitro* seuls, ou après formation du complexe ribonucléoprotéique puis déprotéinisation. Les ARN repliés en présence de NP présentent globalement une baisse de réactivité, indépendamment de l'état d'oligomérisation de la NP. Ces résultats indiquent que la fixation de la NP influe sur la structure de l'ARN de façon irréversible. J'ai confirmé ce résultat en générant des modèles de structure 2D des ARNv dans ces deux conditions qui intègrent les paramètres thermodynamiques et les réactivités SHAPE déterminées expérimentalement. Les structures obtenues pour chacune des conditions sont différentes, permettant de conclure à un rôle de la protéine NP dans le repliement de l'ARN au sein des RNPv. Les structures des ARNv M et NS de la souche A/Vietnam/1203/2004 (H5N1) ont été récemment publiées dans la littérature [134,220] et sont différentes des structures que j'ai obtenues pour la souche WSN, bien que certaines régions présentent la même structure (nucléotides 261-289 et 312-327 pour NS, 34-61 et 788-809 pour M). Malgré les nombreuses différences structurales de ces segments entre les deux souches étudiées, certaines caractéristiques, telles que la présence de nombreuses boucles internes et de bulges, limitant la présence de très

longues hélices ou de tiges-boucles très courtes, sont retrouvées. Ces résultats sont cependant issus d'analyses réalisées sur des ARN transcrits *in vitro* et non complexés à la NP, rendant très peu probable la présence de ces structures au sein des particules virales.

Les structures chaperonnées sont différentes des structures des ARNv seuls, mais elles ne sont guère (ou pas) plus stables thermodynamiquement que celles-ci, ce qui suggère que le remodelage de la structure de l'ARNv par la protéine NP ne stabilise pas cette structure. Cependant, l'absence de stabilisation au niveau thermodynamique des ARN étudiés n'indique pas une absence d'activité chaperonne de la protéine NP. En effet, si dans les cas des chaperons de protéines, la structure finale est généralement plus stable thermodynamiquement, ce n'est généralement pas le cas avec les chaperons d'ARN [95]. De plus, en analysant la répartition de la taille des hélices entre les structures repliées avec et sans protéine, on observe que la proportion d'hélices de grande taille augmente dans les conditions PK. Ces résultats indiquent une restructuration globale de l'ARN en présence de NP, indépendamment de son état d'oligomérisation, qui tend à stabiliser la structure. Cette stabilisation invisible lorsque l'on considère les paramètres thermodynamiques suggère que la NP décroît la flexibilité structurale de l'ARN viral.

L'effet chaperon des protéines monomériques et trimériques n'est pas rigoureusement identique : certaines régions gardent la même structure que l'ARNv libre après fixation de la protéine trimérique mais sont totalement réarrangées par la protéine monomérique, l'inverse étant également observé. Au contraire, il n'y a que très peu de régions qui présentent une structure unique, qui ne sera retrouvée dans aucune autre condition testée. Il est important de noter que ces résultats ont été obtenus en absence du complexe polymérase viral, composant majeur de la RNPv. Cependant, la fixation de ce complexe étant localisée aux extrémités 5' et 3' de l'ARNv, sa fixation ne devrait pas avoir d'effet structural majeur sur l'ARNv.

J'ai par la suite déterminé les sites de fixation de la protéine NP, sous sa forme monomérique et trimérique, en identifiant les nucléotides présentant une baisse de réactivité dans la condition NP comparée à la condition PK, ce qui indique une protection de ces régions. En comparant la composition nucléotidique de ces sites de fixation et celle de l'ARN entier, j'ai remarqué que les sites de fixation sont enrichis en cytosine et appauvris en guanine pour la protéine NP de la souche WSN, de même que

pour la protéine de la souche PR8, ce qui suggère que les protéines se comportent similairement, indépendamment de leur souche d'origine. Ces résultats sont surprenants à la lumière des données de la littérature, également obtenues avec la protéine de la souche WSN. En effet, une étude a récemment montré, par HITS-CLiP (*high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation*), que les sites de fixation de la protéine NP de la souche WSN sont enrichis en guanosine et appauvris en uracile [133]. Cette différence de résultat peut s'expliquer par l'approche expérimentale. En effet, la cartographie chimique permet de distinguer les nucléotides flexibles des nucléotides engagés dans une interaction, que cette dernière soit avec une protéine ou avec un autre nucléotide. Les sites de fixation que j'ai identifiés peuvent donc être des régions qui, après fixation de la NP, sont engagées dans une interaction réversible avec une autre région (intra ou intermoléculaire). Le réactif SHAPE utilisé au cours de ma thèse modifie le squelette sucre-phosphate des nucléotides flexibles et qui est inaccessible dans le cas d'une interaction avec la protéine NP ou un autre nucléotide. La technique du HITS-CLiP identifie directement, grâce au pontage aux UV, les séquences en proximité étroite avec la protéine. Cependant, le pontage aux UV n'est pas homogène, plusieurs études ayant montré une préférence pour le pontage de certains nucléotides [236,285]. De plus, les techniques CLiP nécessitent l'utilisation d'une RNase pour isoler un fragment d'ARN ponté à la protéine et le séquencer. Ce clivage entraîne la présence de plusieurs nucléotides en plus du site de pontage, qui seront séquencés, et chaque RNase ayant des sites de clivage préférentiels [121], cela peut affecter la composition nucléotidique des sites protégés.

Pour remédier aux limites exposées de l'approche expérimentale que j'ai utilisée, il faudrait utiliser un réactif chimique qui, comme le DMS, modifie la face Watson-Crick des nucléotides et permettrait la modification des nucléotides en interaction avec la protéine NP. En comparant le profil SHAPE et le profil DMS, il serait alors possible de discriminer les régions protégées par la NP des régions restructurées. Les résultats préliminaires que j'ai obtenu pour le SHAPE *in vitro* montre effectivement des différences de réactivités entre les échantillons modifiés au DMS ou avec un réactif SHAPE. Certains nucléotides non réactifs en présence d' Ai-N_3 le sont en présence de DMS, ce qui indique que le nucléotide n'est pas impliqué dans une interaction ARN-ARN mais ARN-NP.

J'ai également observé que les dinucléotides CA et CC sont très présents dans les sites de protection de la NP, aussi bien pour la souche PR8 que WSN, tandis que les dinucléotides CG et GG sont sous-représentés. Ces résultats suggèrent une préférence de fixation des protéines NP sur certaines régions de l'ARN et sont à corrélés à des résultats récemment publiés qui montrent que toutes les régions de l'ARN ne sont pas équitablement fixées par la protéine NP [133,272]. Les sites enrichis en dinucléotides CA et CC pourraient être les régions enrichies en NP décrites dans la littérature et les régions riches en dinucléotides CG et GG seraient les régions appauvries en NP. Des différences entre les deux souches sont également visibles quant à la composition en dinucléotides des sites de protection, notamment pour le dinucléotide GC, fortement présent au niveau des sites de protection de la NP PR8 alors que pour la protéine NP WSN, ce dinucléotide est plus rarement retrouvé. Ces différences de reconnaissance entre les protéines des deux souches pourraient être à l'origine de repliements différents de l'ARN, les régions appauvries en NP contenant de nombreuses structures secondaires [272], ce qui pourrait expliquer des incompatibilités génétiques entre ces deux souches.

En représentant les réactivités obtenues dans les conditions NP monomérique ou trimérique sur la structure déterminée dans la condition PK correspondante, j'ai observé que l'état d'oligomérisation de la protéine NP ne semble avoir que peu d'influence sur les protections. Cette absence de différence pourrait s'expliquer par le protocole de formation des complexes ribonucléoprotéiques. En effet, les données de la littérature indiquent que la fixation de la protéine monomérique à l'ARN induit rapidement la formation de trimères de NP sur de longs ARN [242]. L'incubation de trente minutes pour la formation des complexes pourrait donc permettre à la protéine monomérique de former des trimères, ce qui empêcherait de différencier clairement les deux états d'oligomérisation obtenus après dialyse.

De façon intéressante, j'ai observé dans plusieurs régions que la fixation de la NP sur une région simple brin permet l'ouverture de la région double brin adjacente. La protéine NP, en plus de son activité chaperonne irréversible, permet donc de modifier la structure de l'ARN de façon transitoire, menant à l'exposition de nucléotides impliqués dans des appariements en son absence, comme cela a déjà été proposé dans la littérature [11]. L'ouverture de la structure crée donc de nouvelles régions accessibles pour des interactions intermoléculaires qui pourraient être impliquées

dans le processus d'empaquetage sélectif. De plus, la taille des régions protégées par la NP est fréquemment comprise entre deux et quatre nucléotides, alors qu'une protection sur une douzaine de nucléotides est attendue [272]. L'approche expérimentale que j'ai utilisée me permet d'identifier les sites de fixation de la NP s'ils sont en conformation simple brin, or les sites que j'ai identifiés sont plus courts que ce qui a été décrit dans la littérature, suggérant une fixation de la protéine NP sur des régions double brin. Afin de mieux comparer les deux états d'oligomérisation de la NP, j'ai reporté les réactivités obtenues dans les conditions NP sur la structure déterminée dans la condition NoNP. Cette analyse me permet de comparer directement ces deux états sur la même structure, tandis que l'analyse précédente est indirecte car la structure déterminée dans la condition PK est différente entre la forme monomérique et trimérique de la NP. Les résultats de cette analyse montrent que la protéine NP monomérique est celle qui induit le plus de changements de réactivités.

Les résultats obtenus dans la partie *in viro* de ma thèse sont principalement des mises au point qui m'ont permis d'affiner le protocole de purification des particules virales et de valider l'approche expérimentale de cartographie chimique comprenant la modification de l'ARN et la construction de banques de séquençage haut débit. La modification de l'ARN au sein des particules virales intactes est réalisable aussi bien avec les réactifs SHAPE (NMIA, Ai-N₃) qu'avec le DMS, et le traitement visant à désassembler les particules virales afin d'inhiber les interactions intermoléculaires est également efficace. En effet, la comparaison des conditions intactes et désassemblées met en évidence des différences de réactivité, notamment des augmentations de réactivité dans la condition désassemblée, indiquant que ces régions sont importantes dans le mécanisme d'empaquetage viral. Les régions impliquées dans des interactions intermoléculaires seront non modifiées dans les particules intactes, mais réactives dans les particules désassemblées. De plus, l'utilisation de deux sondes chimiques différentes met en évidence des différences dans les mêmes conditions expérimentales, ce qui justifie notre idée de combiner les approches de cartographie chimique par SHAPE-MaP et DMS-MaP, chacune apportant des informations complémentaires. En effet, comme expliqué précédemment, l'approche SHAPE ne permet pas de discriminer les interactions entre ARNv des interactions ARN-NP, tandis que la modification au DMS, réalisée au niveau des appariements Watson-Crick, n'est sensible qu'aux interactions ARN-ARN. La combinaison de ces deux approches permet donc de déterminer quelles régions sont double brin, simple brin, ou associées à la

protéine NP, ce qui permet d'affiner les résultats et de dessiner une structure la plus précise possible. Cependant, le DMS ne permet d'interroger la flexibilité que des nucléotides A et C, l'utilisation d'un réactif, tel que le CMCT, qui permet la modification des nucléotides U et G non appariés sera nécessaire pour compléter mon étude. Cependant le CMCT pénètre peu ou pas les membranes cellulaires, ce qui limiterait son utilisation aux particules désassemblées.

PERSPECTIVES

Perspectives

Afin de confirmer la capacité de la NP à chaperonner les ARN, il serait intéressant d'utiliser des protéines issues de différentes souches. De même, la réalisation de tests de repliement croisés, en utilisant la protéine d'une souche et l'ARN d'une autre souche ou un ARN non viral, nous permettrait de déterminer si l'activité chaperonne de la protéine NP est universelle ou nécessite la reconnaissance de signaux spécifiques présents sur les ARNv de toutes les souches, voire uniquement sur les ARNv homologues. De même, l'utilisation d'une protéine produite en système cellulaire eucaryote permettrait d'affiner les résultats en prenant en compte l'impact des modifications post-traductionnelles, notamment les phosphorylations [102] et acétylations [80], sur la propriété chaperonne de la protéine, d'autant que la phosphorylation du résidu S165 est importante pour l'oligomérisation de la NP [249]. Il serait intéressant d'approfondir les premiers résultats obtenus avec la NP PR8, souche couramment utilisée dans les laboratoires et pour les vaccins, mais il faut pour cela améliorer sa production et sa purification. Dans ce but, nous pourrions exprimer la protéine NP tronquée des sept premiers nucléotides, stratégie déjà utilisée dans la littérature pour déterminer la structure cristallographique de la protéine NP de la souche WSN [278]. Comme présenté précédemment, la présence de trois bandes après purification de la NP PR8 est probablement due à une dégradation de l'extrémité N-terminale de cette protéine, qui pourrait être évitée en utilisant une protéine tronquée.

Dans le but de mieux caractériser l'effet de l'oligomérisation de la protéine NP sur son activité chaperonne, il serait intéressant d'utiliser un mutant monomérique, qui n'est pas capable de former des trimères, comme les mutants R416A et S165D, décrits et caractérisés dans la littérature [36]. Cette stratégie permettrait de s'affranchir des effets d'une incubation longue pour former les complexes qui entraîne une oligomérisation de la NP monomérique. L'utilisation de réactifs chimiques plus réactifs tels que le BzCN ou le DMS permettrait également de réduire le temps d'incubation et d'améliorer l'homogénéité des structures repliées.

La partie *in viro* ouvre les perspectives les plus intéressantes quant au mécanisme d'empaquetage sélectif. La combinaison des techniques de SHAPE-MaP et DMS-MaP sur les particules intactes et désassemblées permettra d'identifier les régions impliquées dans des interactions intermoléculaires. Ces séquences seront alors

analysées le logiciel GUUGle (<https://bibiserv2.cebitec.uni-bielefeld.de/guugle>) qui prédira les partenaires d'interaction de cette séquence dans le génome viral. Parmi ces partenaires potentiels, nous rechercherons les séquences ayant également été identifiées comme non appariées dans les particules désassemblées et appariées dans les particules natives. Les interactions seront alors validées par mutagenèse dirigée, qui abolira les interactions prédites, et la capacité des virus mutants à se répliquer et à produire des particules virales infectieuses sera testée. Des mutants compensatoires, restaurant l'interaction, seront également étudiés afin de valider l'importance de l'interaction pour l'empaquetage sélectif. En plus de l'approche par mutagenèse dirigée, nous prévoyons de réaliser les études de SHAPE-MaP et DMS-MaP sur des virus possédant des mutations dans un ou plusieurs signaux d'empaquetage et dont les défauts d'empaquetage ont déjà été caractérisés par nos collaborateurs à Freiburg.

A long terme, nous développerons un logiciel de prédiction des réseaux d'interaction pour une souche donnée. Ce logiciel combinera les données de prédiction d'interaction fournies par l'outil GUUGle et les données expérimentales obtenues sur différentes souches virales afin de prédire les interactions d'une nouvelle souche, qui pourront être confirmées par mutagenèse dirigée. Ce logiciel pourrait alors être utilisé pour améliorer l'efficacité de production des vaccins en modifiant spécifiquement le fond génétique PR8 pour diminuer l'incorporation des segments HA et NA de cette souche tout en facilitant l'incorporation des segments HA et NA de la souche d'intérêt vaccinal.

BIBLIOGRAPHIE

Bibliographie

- [1] Abd El-Wahab, E.W., Smyth, R.P., Mailler, E., Bernacchi, S., Vivet-Boudou, V., Hijnen, M., Jossinet, F., Mak, J., Paillart, J.-C., Marquet, R. (2014) Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat. Commun.* 5, 4304, Doi: 10.1038/ncomms5304.
- [2] Akarsu, H., Burmeister, W.P., Petosa, C., Petit, I., Müller, C.W., Ruigrok, R.W.H., Baudin, F. (2003) Crystal structure of the M1 protein-binding domain of the influenza A virus nuclear export protein (NEP/NS2). *EMBO J.* 22(18), 4646–55, Doi: 10.1093/emboj/cdg449.
- [3] Albo, C., Valencia, A., Portela, A. (1995) Identification of an RNA binding region within the N-terminal third of the influenza A virus nucleoprotein. *J. Virol.* 69(6), 3799–806.
- [4] Ali, A., Avalos, R.T., Ponimaskin, E., Nayak, D.P. (2000) Influenza Virus Assembly: Effect of Influenza Virus Glycoproteins on the Membrane Association of M1 Protein. *J. Virol.* 74(18), 8709–19.
- [5] Amorim, M.J., Bruce, E.A., Read, E.K.C., Foeglein, A., Mahen, R., Stuart, A.D., Digard, P. (2011) A Rab11- and microtubule-dependent mechanism for cytoplasmic transport of influenza A virus viral RNA. *J. Virol.* 85(9), 4143–56, Doi: 10.1128/JVI.02606-10.
- [6] Amorim, M.-J., Read, E.K., Dalton, R.M., Medcalf, L., Digard, P. (2007) Nuclear export of influenza A virus mRNAs requires ongoing RNA polymerase II activity. *Traffic Cph. Den.* 8(1), 1–11, Doi: 10.1111/j.1600-0854.2006.00507.x.
- [7] Arranz, R., Coloma, R., Chichón, F.J., Conesa, J.J., Carrascosa, J.L., Valpuesta, J.M., Ortín, J., Martín-Benito, J. (2012) The Structure of Native Influenza Virion Ribonucleoproteins. *Science* 338(6114), 1634–7, Doi: 10.1126/science.1228172.
- [8] Avilov, S.V., Moisy, D., Naffakh, N., Cusack, S. (2012) Influenza A virus progeny vRNP trafficking in live infected cells studied with the virus-encoded fluorescently tagged PB2 protein. *Vaccine* 30(51), 7411–7, Doi: 10.1016/j.vaccine.2012.09.077.
- [9] Bae, W., Xia, B., Inouye, M., Severinov, K. (2000) Escherichia coli CspA-family RNA chaperones are transcription antiterminators. *Proc. Natl. Acad. Sci. U. S. A.* 97(14), 7784–9.
- [10] Batisse, J., Guerrero, S., Bernacchi, S., Sleiman, D., Gabus, C., Darlix, J.-L., Marquet, R., Tisné, C., Paillart, J.-C. (2012) The role of Vif oligomerization and RNA chaperone activity in HIV-1 replication. *Virus Res.* 169(2), 361–76, Doi: 10.1016/j.virusres.2012.06.018.
- [11] Baudin, F., Bach, C., Cusack, S., Ruigrok, R.W. (1994) Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO J.* 13(13), 3158–65.
- [12] Bergmann, M., Muster, T. (1995) The relative amount of an influenza A virus segment present in the viral particle is not affected by a reduction in replication of that segment. *J. Gen. Virol.* 76 (Pt 12), 3211–5, Doi: 10.1099/0022-1317-76-12-3211.

- [13] Biswas, S.K., Boutz, P.L., Nayak, D.P. (1998) Influenza virus nucleoprotein interacts with influenza virus polymerase proteins. *J. Virol.* 72(7), 5493–501.
- [14] Biswas, S.K., Nayak, D.P. (1994) Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. *J. Virol.* 68(3), 1819–26.
- [15] Bouloy, M., Plotch, S.J., Krug, R.M. (1978) Globin mRNAs are primers for the transcription of influenza viral RNA in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 75(10), 4886–90.
- [16] Boyce, M., McCrae, M.A., Boyce, P., Kim, J.T. (2016) Inter-segment complementarity in orbiviruses: a driver for co-ordinated genome packaging in the Reoviridae? *J. Gen. Virol.* 97(5), 1145–57, Doi: 10.1099/jgv.0.000400.
- [17] Brownlee, G.G., Sharps, J.L. (2002) The RNA Polymerase of Influenza A Virus Is Stabilized by Interaction with Its Viral RNA Promoter. *J. Virol.* 76(14), 7103–13, Doi: 10.1128/JVI.76.14.7103-7113.2002.
- [18] Brunotte, L., Flies, J., Bolte, H., Reuther, P., Vreede, F., Schwemmle, M. (2014) The nuclear export protein of H5N1 influenza A viruses recruits Matrix 1 (M1) protein to the viral ribonucleoprotein to mediate nuclear export. *J. Biol. Chem.* 289(29), 20067–77, Doi: 10.1074/jbc.M114.569178.
- [19] Bui, M., Wills, E.G., Helenius, A., Whittaker, G.R. (2000) Role of the influenza virus M1 protein in nuclear export of viral ribonucleoproteins. *J. Virol.* 74(4), 1781–6.
- [20] Bulletin Santé Publique France. (2018) Bulletin épidémiologique grippe, semaine 16. Bilan préliminaire. Saison 2017-2018. / Données de surveillance / Grippe : généralités / Grippe / Maladies à prévention vaccinale / Maladies infectieuses / Dossiers thématiques / Accueil. <http://invs.santepubliquefrance.fr/Dossiers-thematiques/Maladies-infectieuses/Maladies-a-prevention-vaccinale/Grippe/Grippe-generalites/Donnees-de-surveillance/Bulletin-epidemiologique-grippe-semaine-16.-Bilan-preliminaire.-Saison-2017-2018>. [accessed August 14, 2018].
- [21] Bullido, R., Gómez-Puertas, P., Albo, C., Portela, A. (2000) Several protein regions contribute to determine the nuclear and cytoplasmic localization of the influenza A virus nucleoprotein. *J. Gen. Virol.* 81(Pt 1), 135–42, Doi: 10.1099/0022-1317-81-1-135.
- [22] Bullough, P.A., Hughson, F.M., Skehel, J.J., Wiley, D.C. (1994) Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* 371(6492), 37–43, Doi: 10.1038/371037a0.
- [23] Calder, L.J., Wasilewski, S., Berriman, J.A., Rosenthal, P.B. (2010) Structural organization of a filamentous influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 107(23), 10685–90, Doi: 10.1073/pnas.1002123107.
- [24] Cao, S., Liu, X., Yu, M., Li, J., Jia, X., Bi, Y., Sun, L., Gao, G.F., Liu, W. (2012) A nuclear export signal in the matrix protein of Influenza A virus is required for efficient virus replication. *J. Virol.* 86(9), 4883–91, Doi: 10.1128/JVI.06586-11.
- [25] Carr, C.M., Kim, P.S. (1993) A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* 73(4), 823–32.
- [26] Chaimayo, C., Hayashi, T., Underwood, A., Hodges, E., Takimoto, T. (2017) Selective incorporation of vRNP into influenza A virions determined by its

- specific interaction with M1 protein. *Virology* 505, 23–32, Doi: 10.1016/j.virol.2017.02.008.
- [27] Chambers, B.S., Parkhouse, K., Ross, T.M., Alby, K., Hensley, S.E. (2015) Identification of hemagglutinin residues responsible for H3N2 antigenic drift during the 2014–2015 influenza season. *Cell Rep.* 12(1), 1–6, Doi: 10.1016/j.celrep.2015.06.005.
- [28] Chan, W.-H., Ng, A.K.-L., Robb, N.C., Lam, M.K.-H., Chan, P.K.-S., Au, S.W.-N., Wang, J.-H., Fodor, E., Shaw, P.-C. (2010) Functional analysis of the influenza virus H5N1 nucleoprotein tail loop reveals amino acids that are crucial for oligomerization and ribonucleoprotein activities. *J. Virol.* 84(14), 7337–45, Doi: 10.1128/JVI.02474-09.
- [29] Chang, S., Sun, D., Liang, H., Wang, J., Li, J., Guo, L., Wang, X., Guan, C., Boruah, B.M., Yuan, L., Feng, F., Yang, M., Wang, L., Wang, Y., Wojdyla, J., Li, L., Wang, J., Wang, M., Cheng, G., Wang, H.-W., Liu, Y. (2015) Cryo-EM Structure of Influenza Virus RNA Polymerase Complex at 4.3 Å Resolution. *Mol. Cell* 57(5), 925–35, Doi: 10.1016/j.molcel.2014.12.031.
- [30] Chase, G.P., Rameix-Welti, M.-A., Zvirbliene, A., Zvirblis, G., Götz, V., Wolff, T., Naffakh, N., Schwemmle, M. (2011) Influenza Virus Ribonucleoprotein Complexes Gain Preferential Access to Cellular Export Machinery through Chromatin Targeting. *PLOS Pathog.* 7(9), e1002187, Doi: 10.1371/journal.ppat.1002187.
- [31] Chauhan, S., Woodson, S.A. (2008) Tertiary interactions determine the accuracy of RNA folding. *J. Am. Chem. Soc.* 130(4), 1296–303, Doi: 10.1021/ja076166i.
- [32] Chen, B.J., Leser, G.P., Jackson, D., Lamb, R.A. (2008) The influenza virus M2 protein cytoplasmic tail interacts with the M1 protein and influences virus assembly at the site of virus budding. *J. Virol.* 82(20), 10059–70, Doi: 10.1128/JVI.01184-08.
- [33] Chen, B.J., Leser, G.P., Morita, E., Lamb, R.A. (2007) Influenza virus hemagglutinin and neuraminidase, but not the matrix protein, are required for assembly and budding of plasmid-derived virus-like particles. *J. Virol.* 81(13), 7111–23, Doi: 10.1128/JVI.00361-07.
- [34] Chen, J., Skehel, J.J., Wiley, D.C. (1999) N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *Proc. Natl. Acad. Sci. U. S. A.* 96(16), 8967–72.
- [35] Chen, R., Holmes, E.C. (2006) Avian influenza virus exhibits rapid evolutionary dynamics. *Mol. Biol. Evol.* 23(12), 2336–41, Doi: 10.1093/molbev/msl102.
- [36] Chenavas, S., Estrozi, L.F., Slama-Schwok, A., Delmas, B., Di Primo, C., Baudin, F., Li, X., Crépin, T., Ruigrok, R.W.H. (2013) Monomeric nucleoprotein of influenza A virus. *PLoS Pathog.* 9(3), e1003275, Doi: 10.1371/journal.ppat.1003275.
- [37] Chenavas, S., Monod, A., Ruigrok, R.W.H., Crépin, T. (2013) Le complexe de transcription et de réplication des virus influenza, une cible pour de nouvelles molécules antivirales. *Virologie* 17(1), 6–16, Doi: 10.1684/vir.2013.0475.
- [38] Chou, Y., Heaton, N.S., Gao, Q., Palese, P., Singer, R.H., Singer, R., Lionnet, T. (2013) Colocalization of different influenza viral RNA segments in the cytoplasm

before viral budding as shown by single-molecule sensitivity FISH analysis. *PLoS Pathog.* 9(5), e1003358, Doi: 10.1371/journal.ppat.1003358.

- [39] Chou, Y., Vafabakhsh, R., Doğanay, S., Gao, Q., Ha, T., Palese, P. (2012) One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis. *Proc. Natl. Acad. Sci. U. S. A.* 109(23), 9101–6, Doi: 10.1073/pnas.1206069109.
- [40] Chu, C.M., Dawson, I.M., Elford, W.J. (1949) Filamentous forms associated with newly isolated influenza virus. *Lancet Lond. Engl.* 1(6554), 602.
- [41] Coloma, R., Valpuesta, J.M., Arranz, R., Carrascosa, J.L., Ortín, J., Martín-Benito, J. (2009) The Structure of a Biologically Active Influenza Virus Ribonucleoprotein Complex. *PLOS Pathog.* 5(6), e1000491, Doi: 10.1371/journal.ppat.1000491.
- [42] Compans, R.W., Content, J., Duesberg, P.H. (1972) Structure of the Ribonucleoprotein of Influenza Virus. *J. Virol.* 10(4), 795–800.
- [43] Cristofari, G., Darlix, J.-L. (2002) The ubiquitous nature of RNA chaperone proteins. *Prog. Nucleic Acid Res. Mol. Biol.* 72, 223–68.
- [44] Cros, J.F., Palese, P. (2003) Trafficking of viral genomic RNA into and out of the nucleus: influenza, Thogoto and Borna disease viruses. *Virus Res.* 95(1–2), 3–12.
- [45] Dadonaite, B., Barilaite, E., Fodor, E., Laederach, A., Bauer, D.L. (2017) The structure of the influenza A virus genome. *BioRxiv*, 236620, Doi: 10.1101/236620.
- [46] Danev, R., Tegunov, D., Baumeister, W. (2017) Using the Volta phase plate with defocus for cryo-EM single particle analysis. *ELife* 6, e23006, Doi: 10.7554/eLife.23006.
- [47] Daniels, R.S., Downie, J.C., Hay, A.J., Knossow, M., Skehel, J.J., Wang, M.L., Wiley, D.C. (1985) Fusion mutants of the influenza virus hemagglutinin glycoprotein. *Cell* 40(2), 431–9.
- [48] Darty, K., Denise, A., Ponty, Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25(15), 1974–5, Doi: 10.1093/bioinformatics/btp250.
- [49] Das, S.C., Watanabe, S., Hatta, M., Noda, T., Neumann, G., Ozawa, M., Kawaoka, Y. (2012) The Highly Conserved Arginine Residues at Positions 76 through 78 of Influenza A Virus Matrix Protein M1 Play an Important Role in Viral Replication by Affecting the Intracellular Localization of M1. *J. Virol.* 86(3), 1522–30, Doi: 10.1128/JVI.06230-11.
- [50] Davey, J., Dimmock, N.J., Colman, A. (1985) Identification of the sequence responsible for the nuclear accumulation of the influenza virus nucleoprotein in *Xenopus* oocytes. *Cell* 40(3), 667–75.
- [51] Deng, T., Engelhardt, O.G., Thomas, B., Akoulitchev, A.V., Brownlee, G.G., Fodor, E. (2006) Role of ran binding protein 5 in nuclear import and assembly of the influenza virus RNA polymerase complex. *J. Virol.* 80(24), 11911–9, Doi: 10.1128/JVI.01565-06.
- [52] Dias, A., Bouvier, D., Crépin, T., McCarthy, A.A., Hart, D.J., Baudin, F., Cusack, S., Ruigrok, R.W.H. (2009) The cap-snatching endonuclease of influenza virus

- polymerase resides in the PA subunit. *Nature* 458(7240), 914–8, Doi: 10.1038/nature07745.
- [53] Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., Assmann, S.M. (2014) *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505(7485), 696–700, Doi: 10.1038/nature12756.
- [54] Diot, C., Fournier, G., Santos, M.D., Magnus, J., Komarova, A., Werf, S. van der., Munier, S., Naffakh, N. (2016) Influenza A Virus Polymerase Recruits the RNA Helicase DDX19 to Promote the Nuclear Export of Viral mRNAs. *Sci. Rep.* 6, 33763, Doi: 10.1038/srep33763.
- [55] Drake, J.W., Holland, J.J. (1999) Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci.* 96(24), 13910–3, Doi: 10.1073/pnas.96.24.13910.
- [56] Dubois, N., Marquet, R., Paillart, J.-C., Bernacchi, S. (2018) Retroviral RNA Dimerization: From Structure to Functions. *Front. Microbiol.* 9, 527, Doi: 10.3389/fmicb.2018.00527.
- [57] Durrer, P., Galli, C., Hoenke, S., Corti, C., Glück, R., Vorherr, T., Brunner, J. (1996) H⁺-induced membrane insertion of influenza virus hemagglutinin involves the HA2 amino-terminal fusion peptide but not the coiled coil region. *J. Biol. Chem.* 271(23), 13417–21.
- [58] Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.P., Ehresmann, B. (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res.* 15(22), 9109–28.
- [59] Eisfeld, A.J., Kawakami, E., Watanabe, T., Neumann, G., Kawaoka, Y. (2011) RAB11A is essential for transport of the influenza virus genome to the plasma membrane. *J. Virol.* 85(13), 6117–26, Doi: 10.1128/JVI.00378-11.
- [60] Eisfeld, A.J., Neumann, G., Kawaoka, Y. (2011) Human immunodeficiency virus rev-binding protein is essential for influenza a virus replication and promotes genome trafficking in late-stage infection. *J. Virol.* 85(18), 9588–98, Doi: 10.1128/JVI.05064-11.
- [61] Elton, D., Medcalf, L., Bishop, K., Harrison, D., Digard, P. (1999) Identification of Amino Acid Residues of Influenza Virus Nucleoprotein Essential for RNA Binding. *J. Virol.* 73(9), 7357–67.
- [62] Elton, D., Simpson-Holley, M., Archer, K., Medcalf, L., Hallam, R., McCauley, J., Digard, P. (2001) Interaction of the influenza virus nucleoprotein with the cellular CRM1-mediated nuclear export pathway. *J. Virol.* 75(1), 408–19, Doi: 10.1128/JVI.75.1.408-419.2001.
- [63] Engelhardt, O.G., Smith, M., Fodor, E. (2005) Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II. *J. Virol.* 79(9), 5812–8, Doi: 10.1128/JVI.79.9.5812-5818.2005.
- [64] Fajardo, T., Sung, P.-Y., Celma, C.C., Roy, P. (2017) Rotavirus Genomic RNA Complex Forms via Specific RNA-RNA Interactions: Disruption of RNA Complex Inhibits Virus Infectivity. *Viruses* 9(7), Doi: 10.3390/v9070167.
- [65] Ferhadian, D., Contrant, M., Printz-Schweigert, A., Smyth, R.P., Paillart, J.-C., Marquet, R. (2018) Structural and Functional Motifs in Influenza Virus RNAs. *Front. Microbiol.* 9, 559, Doi: 10.3389/fmicb.2018.00559.

- [66] Fodor, E., Seong, B.L., Brownlee, G.G. (1993) Photochemical cross-linking of influenza A polymerase to its virion RNA promoter defines a polymerase binding site at residues 9 to 12 of the promoter. *J. Gen. Virol.* 74 (Pt 7), 1327–33, Doi: 10.1099/0022-1317-74-7-1327.
- [67] Fodor, E., Smith, M. (2004) The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex. *J. Virol.* 78(17), 9144–53, Doi: 10.1128/JVI.78.17.9144-9153.2004.
- [68] Fournier, E., Moules, V., Essere, B., Paillart, J.-C., Sirbat, J.-D., Cavalier, A., Rolland, J.-P., Thomas, D., Lina, B., Isel, C., Marquet, R. (2012) Interaction network linking the human H3N2 influenza A virus genomic RNA segments. *Vaccine* 30(51), 7359–67, Doi: 10.1016/j.vaccine.2012.09.079.
- [69] Fournier, E., Moules, V., Essere, B., Paillart, J.-C., Sirbat, J.-D., Isel, C., Cavalier, A., Rolland, J.-P., Thomas, D., Lina, B., Marquet, R. (2012) A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic Acids Res.* 40(5), 2197–209, Doi: 10.1093/nar/gkr985.
- [70] Fujiyoshi, Y., Kume, N.P., Sakata, K., Sato, S.B. (1994) Fine structure of influenza A virus observed by electron cryo-microscopy. *EMBO J.* 13(2), 318–26.
- [71] Galloway, S.E., Reed, M.L., Russell, C.J., Steinhauer, D.A. (2013) Influenza HA subtypes demonstrate divergent phenotypes for cleavage activation and pH of fusion: implications for host range and adaptation. *PLoS Pathog.* 9(2), e1003151, Doi: 10.1371/journal.ppat.1003151.
- [72] Gamblin, S.J., Skehel, J.J. (2010) Influenza Hemagglutinin and Neuraminidase Membrane Glycoproteins. *J. Biol. Chem.* 285(37), 28403–9, Doi: 10.1074/jbc.R110.129809.
- [73] Gao, S., Wang, S., Cao, S., Sun, L., Li, J., Bi, Y., Gao, G.F., Liu, W. (2014) Characteristics of Nucleocytoplasmic Transport of H1N1 Influenza A Virus Nuclear Export Protein. *J. Virol.* 88(13), 7455–63, Doi: 10.1128/JVI.00257-14.
- [74] Garmann, R.F., Gopal, A., Athavale, S.S., Knobler, C.M., Gelbart, W.M., Harvey, S.C. (2015) Visualizing the global secondary structure of a viral RNA genome with cryo-electron microscopy. *RNA N. Y. N* 21(5), 877–86, Doi: 10.1261/rna.047506.114.
- [75] Gavazzi, C., Isel, C., Fournier, E., Moules, V., Cavalier, A., Thomas, D., Lina, B., Marquet, R. (2013) An in vitro network of intermolecular interactions between viral RNA segments of an avian H5N2 influenza A virus: comparison with a human H3N2 virus. *Nucleic Acids Res.* 41(2), 1241–54, Doi: 10.1093/nar/gks1181.
- [76] Gavazzi, C., Yver, M., Isel, C., Smyth, R.P., Rosa-Calatrava, M., Lina, B., Moulès, V., Marquet, R. (2013) A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proc. Natl. Acad. Sci. U. S. A.* 110(41), 16604–9, Doi: 10.1073/pnas.1314419110.
- [77] Gerber, M., Moulès, V., Gavazzi, C., Marquet, R., Isel, C. (2014) Empaquetage sélectif du génome des influenza virus A. *Virologie* 18(5), 278–96, Doi: 10.1684/vir.2014.0580.
- [78] Getie-Kebtie, M., Sultana, I., Eichelberger, M., Alterman, M. (2013) Label-free mass spectrometry-based quantification of hemagglutinin and neuraminidase in

- influenza virus preparations and vaccines. *Influenza Other Respir. Viruses* 7(4), 521–30, Doi: 10.1111/irv.12001.
- [79] Giese, S., Bolte, H., Schwemmle, M. (2016) The Feat of Packaging Eight Unique Genome Segments. *Viruses* 8(6), Doi: 10.3390/v8060165.
- [80] Giese, S., Ciminski, K., Bolte, H., Moreira, É.A., Lakdawala, S., Hu, Z., David, Q., Kolesnikova, L., Götz, V., Zhao, Y., Dengjel, J., Chin, Y.E., Xu, K., Schwemmle, M. (2017) Role of influenza A virus NP acetylation on viral growth and replication. *Nat. Commun.* 8(1), 1259, Doi: 10.1038/s41467-017-01112-3.
- [81] Godet, J., Boudier, C., Humbert, N., Ivanyi-Nagy, R., Darlix, J.-L., Mély, Y. (2012) Comparative nucleic acid chaperone properties of the nucleocapsid protein NCp7 and Tat protein of HIV-1. *Virus Res.* 169(2), 349–60, Doi: 10.1016/j.virusres.2012.06.021.
- [82] Gómez-Puertas, P., Albo, C., Pérez-Pastrana, E., Vivo, A., Portela, A. (2000) Influenza virus matrix protein is the major driving force in virus budding. *J. Virol.* 74(24), 11538–47.
- [83] Gopal, A., Zhou, Z.H., Knobler, C.M., Gelbart, W.M. (2012) Visualizing large RNA molecules in solution. *RNA* 18(2), 284–99, Doi: 10.1261/rna.027557.111.
- [84] Grantham, M.L., Stewart, S.M., Lalime, E.N., Pekosz, A. (2010) Tyrosines in the influenza A virus M2 protein cytoplasmic tail are critical for production of infectious virus particles. *J. Virol.* 84(17), 8765–76, Doi: 10.1128/JVI.00853-10.
- [85] Guan, Y., Vijaykrishna, D., Bahl, J., Zhu, H., Wang, J., Smith, G.J.D. (2010) The emergence of pandemic influenza viruses. *Protein Cell* 1(1), 9–13, Doi: 10.1007/s13238-010-0008-z.
- [86] Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R.W.H., Ortin, J., Hart, D.J., Cusack, S. (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.* 15(5), 500–6, Doi: 10.1038/nsmb.1421.
- [87] Hanada, K., Suzuki, Y., Gojobori, T. (2004) A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Mol. Biol. Evol.* 21(6), 1074–80, Doi: 10.1093/molbev/msh109.
- [88] Hao, L., Sakurai, A., Watanabe, T., Sorensen, E., Nidom, C.A., Newton, M.A., Ahlquist, P., Kawaoka, Y. (2008) *Drosophila* RNAi screen identifies host genes important for influenza virus replication. *Nature* 454(7206), 890–3, Doi: 10.1038/nature07151.
- [89] Harris, A., Cardone, G., Winkler, D.C., Heymann, J.B., Brecher, M., White, J.M., Steven, A.C. (2006) Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proc. Natl. Acad. Sci. U. S. A.* 103(50), 19123–7, Doi: 10.1073/pnas.0607614103.
- [90] Hatada, E., Hasegawa, M., Mukaigawa, J., Shimizu, K., Fukuda, R. (1989) Control of influenza virus gene expression: quantitative analysis of each viral RNA species in infected cells. *J. Biochem. (Tokyo)* 105(4), 537–46.
- [91] Hause, B.M., Collin, E.A., Liu, R., Huang, B., Sheng, Z., Lu, W., Wang, D., Nelson, E.A., Li, F. (2014) Characterization of a Novel Influenza Virus in Cattle and

- Swine: Proposal for a New Genus in the Orthomyxoviridae Family. *MBio* 5(2), e00031-14, Doi: 10.1128/mBio.00031-14.
- [92] He, X., Zhou, J., Bartlam, M., Zhang, R., Ma, J., Lou, Z., Li, X., Li, J., Joachimiak, A., Zeng, Z., Ge, R., Rao, Z., Liu, Y. (2008) Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature* 454(7208), 1123–6, Doi: 10.1038/nature07120.
- [93] Hengrung, N., El Omari, K., Serna Martin, I., Vreede, F.T., Cusack, S., Rambo, R.P., Vonrhein, C., Bricogne, G., Stuart, D.I., Grimes, J.M., Fodor, E. (2015) Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. *Nature* 527(7576), 114–7, Doi: 10.1038/nature15525.
- [94] Henriët, S., Sinck, L., Bec, G., Gorelick, R.J., Marquet, R., Paillart, J.-C. (2007) Vif is a RNA chaperone that could temporally regulate RNA dimerization and the early steps of HIV-1 reverse transcription. *Nucleic Acids Res.* 35(15), 5141–53, Doi: 10.1093/nar/gkm542.
- [95] Herschlag, D. (1995) RNA chaperones and the RNA folding problem. *J. Biol. Chem.* 270(36), 20871–4.
- [96] Hilz, H., Wiegers, U., Adamietz, P. (1975) Stimulation of proteinase K action by denaturing agents: application to the isolation of nucleic acids and the degradation of “masked” proteins. *Eur. J. Biochem.* 56(1), 103–8.
- [97] Hirst, G.K. (1962) Genetic recombination with Newcastle disease virus, polioviruses, and influenza. *Cold Spring Harb. Symp. Quant. Biol.* 27, 303–9.
- [98] Holbrook, S.R., Kim, S.H. (1997) RNA crystallography. *Biopolymers* 44(1), 3–21, Doi: 10.1002/(SICI)1097-0282(1997)44:1<3::AID-BIP2>3.0.CO;2-Z.
- [99] Hsu, M.T., Parvin, J.D., Gupta, S., Krystal, M., Palese, P. (1987) Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. *Proc. Natl. Acad. Sci. U. S. A.* 84(22), 8140–4.
- [100] Hu, Y., Gor, V., Morikawa, K., Nagata, K., Kawaguchi, A. (2017) Cellular splicing factor UAP56 stimulates trimeric NP formation for assembly of functional influenza viral ribonucleoprotein complexes. *Sci. Rep.* 7(1), 14053, Doi: 10.1038/s41598-017-13784-4.
- [101] Hutchinson, E.C., Charles, P.D., Hester, S.S., Thomas, B., Trudgian, D., Martínez-Alonso, M., Fodor, E. (2014) Conserved and host-specific features of influenza virion architecture. *Nat. Commun.* 5, 4816, Doi: 10.1038/ncomms5816.
- [102] Hutchinson, E.C., Denham, E.M., Thomas, B., Trudgian, D.C., Hester, S.S., Ridlova, G., York, A., Turell, L., Fodor, E. (2012) Mapping the Phosphoproteome of Influenza A and B Viruses by Mass Spectrometry. *PLOS Pathog.* 8(11), e1002993, Doi: 10.1371/journal.ppat.1002993.
- [103] Hutchinson, E.C., von Kirchbach, J.C., Gog, J.R., Digard, P. (2010) Genome packaging in influenza A virus. *J. Gen. Virol.* 91(Pt 2), 313–28, Doi: 10.1099/vir.0.017608-0.
- [104] Hutchinson, E.C., Orr, O.E., Man Liu, S., Engelhardt, O.G., Fodor, E. (2011) Characterization of the interaction between the influenza A virus polymerase subunit PB1 and the host nuclear import factor Ran-binding protein 5. *J. Gen. Virol.* 92(Pt 8), 1859–69, Doi: 10.1099/vir.0.032813-0.

- [105] Imai, M., Watanabe, S., Odagiri, T. (2003) Influenza B virus NS2, a nuclear export protein, directly associates with the viral ribonucleoprotein complex. *Arch. Virol.* 148(10), 1873–84, Doi: 10.1007/s00705-003-0166-x.
- [106] Inagaki, A., Goto, H., Kakugawa, S., Ozawa, M., Kawaoka, Y. (2012) Competitive Incorporation of Homologous Gene Segments of Influenza A Virus into Virions. *J. Virol.* 86(18), 10200–2, Doi: 10.1128/JVI.01204-12.
- [107] Inglis, S.C., Brown, C.M. (1981) Spliced and unspliced RNAs encoded by virion RNA segment 7 of influenza virus. *Nucleic Acids Res.* 9(12), 2727–40.
- [108] Iwatsuki-Horimoto, K., Horimoto, T., Fujii, Y., Kawaoka, Y. (2004) Generation of Influenza A Virus NS2 (NEP) Mutants with an Altered Nuclear Export Signal Sequence. *J. Virol.* 78(18), 10149–55, Doi: 10.1128/JVI.78.18.10149-10155.2004.
- [109] Iwatsuki-Horimoto, K., Horimoto, T., Noda, T., Kiso, M., Maeda, J., Watanabe, S., Muramoto, Y., Fujii, K., Kawaoka, Y. (2006) The cytoplasmic tail of the influenza A virus M2 protein plays a role in viral assembly. *J. Virol.* 80(11), 5233–40, Doi: 10.1128/JVI.00049-06.
- [110] Jackson, D., Elderfield, R.A., Barclay, W.S. (2011) Molecular studies of influenza B virus in the reverse genetics era. *J. Gen. Virol.* 92(1), 1–17, Doi: 10.1099/vir.0.026187-0.
- [111] Jorba, N., Coloma, R., Ortín, J. (2009) Genetic trans-Complementation Establishes a New Model for Influenza Virus RNA Transcription and Replication. *PLOS Pathog.* 5(5), e1000462, Doi: 10.1371/journal.ppat.1000462.
- [112] Kadaré, G., Haenni, A.L. (1997) Virus-encoded RNA helicases. *J. Virol.* 71(4), 2583–90.
- [113] Kauffmann, A.D., Campagna, R.J., Bartels, C.B., Childs-Disney, J.L. (2009) Improvement of RNA secondary structure prediction using RNase H cleavage and randomized oligonucleotides. *Nucleic Acids Res.* 37(18), e121, Doi: 10.1093/nar/gkp587.
- [114] Kawaguchi, A., Matsumoto, K., Nagata, K. (2012) YB-1 functions as a porter to lead influenza virus ribonucleoprotein complexes to microtubules. *J. Virol.* 86(20), 11086–95, Doi: 10.1128/JVI.00453-12.
- [115] Kemler, I., Whittaker, G., Helenius, A. (1994) Nuclear import of microinjected influenza virus ribonucleoproteins. *Virology* 202(2), 1028–33, Doi: 10.1006/viro.1994.1432.
- [116] Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y., Segal, E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467(7311), 103–7, Doi: 10.1038/nature09322.
- [117] Khoshouei, M., Radjainia, M., Baumeister, W., Danev, R. (2017) Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat. Commun.* 8, 16099, Doi: 10.1038/ncomms16099.
- [118] Kilbourne, E.D., Murphy, J.S. (1960) Genetic studies of influenza viruses. I. Viral morphology and growth capacity as exchangeable genetic traits. Rapid in ovo adaptation of early passage Asian strain isolates by combination with PR8. *J. Exp. Med.* 111, 387–406.
- [119] King, A.M.Q., Adams, M.J., Carstens, E.B., Elliot J. Lefkowitz (2012) Family - Orthomyxoviridae. *Virus Taxonomy*, Elsevier, San Diego, pp. 749–61.

- [120] Klumpp, K., Ruigrok, R.W.H., Baudin, F. (1997) Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. *EMBO J.* 16(6), 1248–57, Doi: 10.1093/emboj/16.6.1248.
- [121] Knapp, G. (1989) Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol.* 180, 192–212.
- [122] Kobayashi, M., Toyoda, T., Adyshev, D.M., Azuma, Y., Ishihama, A. (1994) Molecular dissection of influenza virus nucleoprotein: deletion mapping of the RNA binding domain. *J. Virol.* 68(12), 8433–6.
- [123] König, R., Stertz, S., Zhou, Y., Inoue, A., Hoffmann, H.-H., Bhattacharyya, S., Alamares, J.G., Tscherne, D.M., Ortigoza, M.B., Liang, Y., Gao, Q., Andrews, S.E., Bandyopadhyay, S., Jesus, P.D., Tu, B.P., Pache, L., Shih, C., Orth, A., Bonamy, G., Miraglia, L., Ideker, T., García-Sastre, A., Young, J.A.T., Palese, P., Shaw, M.L., Chanda, S.K. (2010) Human Host Factors Required for Influenza Virus Replication. *Nature* 463(7282), 813–7, Doi: 10.1038/nature08699.
- [124] Kwok, C.K. (2016) Dawn of the in vivo RNA structurome and interactome. *Biochem. Soc. Trans.* 44(5), 1395–410, Doi: 10.1042/BST20160075.
- [125] Labaronne, A., Swale, C., Monod, A., Schoehn, G., Crépin, T., Ruigrok, R.W.H. (2016) Binding of RNA by the Nucleoproteins of Influenza Viruses A and B. *Viruses* 8(9), Doi: 10.3390/v8090247.
- [126] Lai, J.C.C., Chan, W.W.L., Kien, F., Nicholls, J.M., Peiris, J.S.M., Garcia, J.-M. (2010) Formation of virus-like particles from human cell lines exclusively expressing influenza neuraminidase. *J. Gen. Virol.* 91(Pt 9), 2322–30, Doi: 10.1099/vir.0.019935-0.
- [127] Lakadamyali, M., Rust, M.J., Babcock, H.P., Zhuang, X. (2003) Visualizing infection of individual influenza viruses. *Proc. Natl. Acad. Sci. U. S. A.* 100(16), 9280–5, Doi: 10.1073/pnas.0832269100.
- [128] Lakdawala, S.S., Wu, Y., Wawrzusin, P., Kabat, J., Broadbent, A.J., Lamirande, E.W., Fodor, E., Altan-Bonnet, N., Shroff, H., Subbarao, K. (2014) Influenza a virus assembly intermediates fuse in the cytoplasm. *PLoS Pathog.* 10(3), e1003971, Doi: 10.1371/journal.ppat.1003971.
- [129] Lamb, R.A., Lai, C.J. (1980) Sequence of interrupted and uninterrupted mRNAs and cloned DNA coding for the two overlapping nonstructural proteins of influenza virus. *Cell* 21(2), 475–85.
- [130] Lamb, R.A., Lai, C.J., Choppin, P.W. (1981) Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. *Proc. Natl. Acad. Sci. U. S. A.* 78(7), 4170–4.
- [131] Lamb, R.A., Zebedee, S.L., Richardson, C.D. (1985) Influenza virus M2 protein is an integral membrane protein expressed on the infected-cell surface. *Cell* 40(3), 627–33.
- [132] Larsen, S., Bui, S., Perez, V., Mohammad, A., Medina-Ramirez, H., Newcomb, L.L. (2014) Influenza polymerase encoding mRNAs utilize atypical mRNA nuclear export. *Virol. J.* 11, 154, Doi: 10.1186/1743-422X-11-154.
- [133] Lee, N., Le Sage, V., Nanni, A.V., Snyder, D.J., Cooper, V.S., Lakdawala, S.S. (2017) Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucleic Acids Res.* 45(15), 8968–77, Doi: 10.1093/nar/gkx584.

- [134] Lenartowicz, E., Keszy, J., Ruskowska, A., Soszynska-Jozwiak, M., Michalak, P., Moss, W.N., Turner, D.H., Kierzek, R., Kierzek, E. (2016) Self-Folding of Naked Segment 8 Genomic RNA of Influenza A Virus. *PloS One* 11(2), e0148281, Doi: 10.1371/journal.pone.0148281.
- [135] Levin, J.G., Guo, J., Rouzina, Ioulia., Musier-Forsyth, K. (2005) Nucleic Acid Chaperone Activity of HIV-1 Nucleocapsid Protein: Critical Role in Reverse Transcription and Molecular Mechanism. *Progress in Nucleic Acid Research and Molecular Biology*, vol. 80, Academic Press, pp. 217–86.
- [136] Levin, J.G., Mitra, M., Mascarenhas, A., Musier-Forsyth, K. (2010) Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol.* 7(6), 754–74.
- [137] Li, J., Yu, M., Zheng, W., Liu, W. (2015) Nucleocytoplasmic Shuttling of Influenza A Virus Proteins. *Viruses* 7(5), 2668–82, Doi: 10.3390/v7052668.
- [138] Lu, K., Heng, X., Garyu, L., Monti, S., Garcia, E.L., Kharytonchyk, S., Dorjsuren, B., Kulandaivel, G., Jones, S., Hiremath, A., Divakaruni, S.S., LaCotti, C., Barton, S., Tummillo, D., Hosis, A., Edme, K., Albrecht, S., Telesnitsky, A., Summers, M.F. (2011) NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science* 334(6053), 242–5, Doi: 10.1126/science.1210460.
- [139] de Lucas, S., Peredo, J., Marión, R.M., Sánchez, C., Ortín, J. (2010) Human Staufen1 protein interacts with influenza virus ribonucleoproteins and is required for efficient virus multiplication. *J. Virol.* 84(15), 7603–12, Doi: 10.1128/JVI.00504-10.
- [140] Lucks, J.B., Mortimer, S.A., Trapnell, C., Luo, S., Aviran, S., Schroth, G.P., Pachter, L., Doudna, J.A., Arkin, A.P. (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci.* 108(27), 11063–8, Doi: 10.1073/pnas.1106501108.
- [141] Luque, D., Rivas, G., Alfonso, C., Carrascosa, J.L., Rodríguez, J.F., Castón, J.R. (2009) Infectious bursal disease virus is an icosahedral polyploid dsRNA virus. *Proc. Natl. Acad. Sci.* 106(7), 2148–52, Doi: 10.1073/pnas.0808498106.
- [142] Ma, K., Roy, A.M., Whittaker, G.R. (2001) Nuclear export of influenza virus ribonucleoproteins: identification of an export intermediate at the nuclear periphery. *Virology* 282(2), 215–20, Doi: 10.1006/viro.2001.0833.
- [143] Maeda, T., Ohnishi, S. (1980) Activation of influenza virus by acidic media causes hemolysis and fusion of erythrocytes. *FEBS Lett.* 122(2), 283–7.
- [144] Mailler, E., Bernacchi, S., Marquet, R., Paillart, J.-C., Vivet-Boudou, V., Smyth, R.P. (2016) The Life-Cycle of the HIV-1 Gag-RNA Complex. *Viruses* 8(9), Doi: 10.3390/v8090248.
- [145] Mark, G.E., Taylor, J.M., Broni, B., Krug, R.M. (1979) Nuclear accumulation of influenza viral RNA transcripts and the effects of cycloheximide, actinomycin D, and alpha-amanitin. *J. Virol.* 29(2), 744–52.
- [146] Marklund, J.K., Ye, Q., Dong, J., Tao, Y.J., Krug, R.M. (2012) Sequence in the influenza A virus nucleoprotein required for viral polymerase binding and RNA synthesis. *J. Virol.* 86(13), 7292–7, Doi: 10.1128/JVI.00014-12.
- [147] Markovic, I., Leikina, E., Zhukovsky, M., Zimmerberg, J., Chernomordik, L.V. (2001) Synchronized activation and refolding of influenza hemagglutinin in

- multimeric fusion machines. *J. Cell Biol.* 155(5), 833–44, Doi: 10.1083/jcb.200103005.
- [148] Martin, I.F. de C., Fournier, G., Sachse, M., Pizarro-Cerda, J., Risco, C., Naffakh, N. (2017) Influenza virus genome reaches the plasma membrane via a modified endoplasmic reticulum and Rab11-dependent vesicles. *Nat. Commun.* 8(1), 1396, Doi: 10.1038/s41467-017-01557-6.
- [149] Martin, K., Helenius, A. (1991) Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. *Cell* 67(1), 117–30.
- [150] Martín-Benito, J., Area, E., Ortega, J., Llorca, O., Valpuesta, J.M., Carrascosa, J.L., Ortín, J. (2001) Three-dimensional reconstruction of a recombinant influenza virus ribonucleoprotein particle. *EMBO Rep.* 2(4), 313–7, Doi: 10.1093/embo-reports/kve063.
- [151] Martínez-Alonso, M., Hengrung, N., Fodor, E. (2016) RNA-Free and Ribonucleoprotein-Associated Influenza Virus Polymerases Directly Bind the Serine-5-Phosphorylated Carboxyl-Terminal Domain of Host RNA Polymerase II. *J. Virol.* 90(13), 6014–21, Doi: 10.1128/JVI.00494-16.
- [152] Matlin, K.S., Reggio, H., Helenius, A., Simons, K. (1981) Infectious entry pathway of influenza virus in a canine kidney cell line. *J. Cell Biol.* 91(3 Pt 1), 601–13.
- [153] Matsuoka, Y., Matsumae, H., Katoh, M., Eisfeld, A.J., Neumann, G., Hase, T., Ghosh, S., Shoemaker, J.E., Lopes, T.J., Watanabe, T., Watanabe, S., Fukuyama, S., Kitano, H., Kawaoka, Y. (2013) A comprehensive map of the influenza A virus replication cycle. *BMC Syst. Biol.* 7, 97, Doi: 10.1186/1752-0509-7-97.
- [154] McCown, M.F., Pekosz, A. (2005) The Influenza A Virus M2 Cytoplasmic Tail Is Required for Infectious Virus Production and Efficient Genome Packaging. *J. Virol.* 79(6), 3595–605, Doi: 10.1128/JVI.79.6.3595-3605.2005.
- [155] McCown, M.F., Pekosz, A. (2006) Distinct domains of the influenza a virus M2 protein cytoplasmic tail mediate binding to the M1 protein and facilitate infectious virus production. *J. Virol.* 80(16), 8178–89, Doi: 10.1128/JVI.00627-06.
- [156] McDonald, S.M., Nelson, M.I., Turner, P.E., Patton, J.T. (2016) Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nat. Rev. Microbiol.* 14(7), 448–60, Doi: 10.1038/nrmicro.2016.46.
- [157] McDonald, S.M., Patton, J.T. (2011) Assortment and packaging of the segmented rotavirus genome. *Trends Microbiol.* 19(3), 136–44, Doi: 10.1016/j.tim.2010.12.002.
- [158] McGeoch, D., Fellner, P., Newton, C. (1976) Influenza virus genome consists of eight distinct RNA species. *Proc. Natl. Acad. Sci. U. S. A.* 73(9), 3045–9.
- [159] Medcalf, L., Poole, E., Elton, D., Digard, P. (1999) Temperature-sensitive lesions in two influenza A viruses defective for replicative transcription disrupt RNA binding by the nucleoprotein. *J. Virol.* 73(9), 7349–56.
- [160] Melen, K., Fagerlund, R., Franke, J., Kohler, M., Kinnunen, L., Julkunen, I. (2003) Importin alpha nuclear localization signal binding sites for STAT1, STAT2, and influenza A virus nucleoprotein. *J. Biol. Chem.* 278(30), 28193–200, Doi: 10.1074/jbc.M303571200.

- [161] Mindich, L. (2012) Packaging in dsRNA viruses. *Adv. Exp. Med. Biol.* 726, 601–8, Doi: 10.1007/978-1-4614-0980-9_26.
- [162] Mir, M.A., Panganiban, A.T. (2006) The bunyavirus nucleocapsid protein is an RNA chaperone: possible roles in viral RNA panhandle formation and genome replication. *RNA N. Y. N* 12(2), 272–82, Doi: 10.1261/rna.2101906.
- [163] Moeller, A., Kirchdoerfer, R.N., Potter, C.S., Carragher, B., Wilson, I.A. (2012) Organization of the influenza virus replication machinery. *Science* 338(6114), 1631–4, Doi: 10.1126/science.1227270.
- [164] Moisy, D., Naffakh, N. (2013) Interactions moléculaires entre les ribonucléoprotéines des virus influenza de type A et la cellule hôte. *Virologie* 17(4), 278–93, Doi: 10.1684/vir.2013.0510.
- [165] Momose, F., Kikuchi, Y., Komase, K., Morikawa, Y. (2007) Visualization of microtubule-mediated transport of influenza viral progeny ribonucleoprotein. *Microbes Infect.* 9(12), 1422–33, Doi: 10.1016/j.micinf.2007.07.007.
- [166] Momose, F., Sekimoto, T., Ohkura, T., Jo, S., Kawaguchi, A., Nagata, K., Morikawa, Y. (2011) Apical transport of influenza A virus ribonucleoprotein requires Rab11-positive recycling endosome. *PloS One* 6(6), e21123, Doi: 10.1371/journal.pone.0021123.
- [167] Moreira, É.A., Weber, A., Bolte, H., Kolesnikova, L., Giese, S., Lakdawala, S., Beer, M., Zimmer, G., García-Sastre, A., Schwemmler, M., Juozapaitis, M. (2016) A conserved influenza A virus nucleoprotein code controls specific viral genome packaging. *Nat. Commun.* 7, 12861, Doi: 10.1038/ncomms12861.
- [168] Mosley, Y.-Y.C., Wu, C.C., Lin, T.L. (2016) IBDV particles packaged with only segment A dsRNA. *Virology* 488, 68–72, Doi: 10.1016/j.virol.2015.11.001.
- [169] Mosley, Y.-Y.C., Wu, C.C., Lin, T.L. (2013) Infectious bursal disease virus rescued efficiently with 3' authentic RNA sequence induces humoral immunity without bursal atrophy. *Vaccine* 31(4), 704–10, Doi: 10.1016/j.vaccine.2012.11.040.
- [170] Mosley, Y.-Y.C., Wu, C.C., Lin, T.L. (2017) Infectious bursal disease virus as a replication-incompetent viral vector expressing green fluorescent protein. *Arch. Virol.* 162(1), 23–32, Doi: 10.1007/s00705-016-3066-6.
- [171] Moulès, V., Terrier, O., Yver, M., Riteau, B., Moriscot, C., Ferraris, O., Julien, T., Giudice, E., Rolland, J.-P., Erny, A., Bouscambert-Duchamp, M., Frobert, E., Rosa-Calatrava, M., Pu Lin, Y., Hay, A., Thomas, D., Schoehn, G., Lina, B. (2011) Importance of viral genomic composition in modulating glycoprotein content on the surface of influenza virus particles. *Virology* 414(1), 51–62, Doi: 10.1016/j.virol.2011.03.011.
- [172] Muraki, Y., Hongo, S. (2010) The molecular virology and reverse genetics of influenza C virus. *Jpn. J. Infect. Dis.* 63(3), 157–65.
- [173] Musier-Forsyth, K. (2010) RNA remodeling by chaperones and helicases. *RNA Biol.* 7(6), 632–3.
- [174] Nakatsu, S., Murakami, S., Shindo, K., Horimoto, T., Sagara, H., Noda, T., Kawaoka, Y. (2018) Influenza C and D Viruses Package Eight Organized Ribonucleoprotein Complexes. *J. Virol.* 92(6), Doi: 10.1128/JVI.02084-17.

- [175] Nakatsu, S., Sagara, H., Sakai-Tagawa, Y., Sugaya, N., Noda, T., Kawaoka, Y. (2016) Complete and Incomplete Genome Packaging of Influenza A and B Viruses. *MBio* 7(5), Doi: 10.1128/mBio.01248-16.
- [176] Neumann, G., Castrucci, M.R., Kawaoka, Y. (1997) Nuclear import and export of influenza virus nucleoprotein. *J. Virol.* 71(12), 9690–700.
- [177] Neumann, G., Hughes, M.T., Kawaoka, Y. (2000) Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *EMBO J.* 19(24), 6751–8, Doi: 10.1093/emboj/19.24.6751.
- [178] Ng, A.K.-L., Zhang, H., Tan, K., Li, Z., Liu, J., Chan, P.K.-S., Li, S.-M., Chan, W.-Y., Au, S.W.-N., Joachimiak, A., Walz, T., Wang, J.-H., Shaw, P.-C. (2008) Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* 22(10), 3638–47, Doi: 10.1096/fj.08-112110.
- [179] Nilsen, T.W. (2013) RNA Structure Determination Using Nuclease Digestion. *Cold Spring Harb. Protoc.* 2013(4), pdb.proto72330, Doi: 10.1101/pdb.proto72330.
- [180] Noda, T., Murakami, S., Nakatsu, S., Imai, H., Muramoto, Y., Shindo, K., Sagara, H., Kawaoka, Y. (2018) Importance of the 1+7 configuration of ribonucleoprotein complexes for influenza A virus genome packaging. *Nat. Commun.* 9(1), 54, Doi: 10.1038/s41467-017-02517-w.
- [181] Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R.H., Kawaoka, Y. (2006) Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature* 439(7075), 490–2, Doi: 10.1038/nature04378.
- [182] Noda, T., Sugita, Y., Aoyama, K., Hirase, A., Kawakami, E., Miyazawa, A., Sagara, H., Kawaoka, Y. (2012) Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nat. Commun.* 3, 639, Doi: 10.1038/ncomms1647.
- [183] OMS. (2018) Grippe saisonnière. World Health Organization. [http://www.who.int/fr/news-room/fact-sheets/detail/influenza-\(seasonal\)](http://www.who.int/fr/news-room/fact-sheets/detail/influenza-(seasonal)). [accessed August 16, 2018].
- [184] O'Neill, R.E., Jaskunas, R., Blobel, G., Palese, P., Moroianu, J. (1995) Nuclear import of influenza virus RNA can be mediated by viral nucleoprotein and transport factors required for protein import. *J. Biol. Chem.* 270(39), 22701–4.
- [185] O'Neill, R.E., Talon, J., Palese, P. (1998) The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. *EMBO J.* 17(1), 288–96, Doi: 10.1093/emboj/17.1.288.
- [186] Ortega, J., Martín-Benito, J., Zürcher, T., Valpuesta, J.M., Carrascosa, J.L., Ortín, J. (2000) Ultrastructural and functional analyses of recombinant influenza virus ribonucleoproteins suggest dimerization of nucleoprotein during virus amplification. *J. Virol.* 74(1), 156–63.
- [187] Palese, P. (1977) The genes of influenza virus. *Cell* 10(1), 1–10, Doi: 10.1016/0092-8674(77)90133-7.
- [188] Palese, P., Compans, R.W. (1976) Inhibition of influenza virus replication in tissue culture by 2-deoxy-2,3-dehydro-N-trifluoroacetylneuraminic acid

- (FANA): mechanism of action. *J. Gen. Virol.* 33(1), 159–63, Doi: 10.1099/0022-1317-33-1-159.
- [189] Palese, P., Tobita, K., Ueda, M., Compans, R.W. (1974) Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. *Virology* 61(2), 397–410, Doi: 10.1016/0042-6822(74)90276-1.
- [190] Paterson, D., Fodor, E. (2012) Emerging Roles for the Influenza A Virus Nuclear Export Protein (NEP). *PLOS Pathog.* 8(12), e1003019, Doi: 10.1371/journal.ppat.1003019.
- [191] Patterson, S., Oxford, J.S., Dourmashkin, R.R. (1979) Studies on the mechanism of influenza virus entry into cells. *J. Gen. Virol.* 43(1), 223–9, Doi: 10.1099/0022-1317-43-1-223.
- [192] Pattnaik, A.K., Brown, D.J., Nayak, D.P. (1986) Formation of influenza virus particles lacking hemagglutinin on the viral envelope. *J. Virol.* 60(3), 994–1001.
- [193] Perez, J.T., Varble, A., Sachidanandam, R., Zlatev, I., Manoharan, M., García-Sastre, A., tenOever, B.R. (2010) Influenza A virus-generated small RNAs regulate the switch from transcription to replication. *Proc. Natl. Acad. Sci. U. S. A.* 107(25), 11525–30, Doi: 10.1073/pnas.1001984107.
- [194] Pflug, A., Guilligay, D., Reich, S., Cusack, S. (2014) Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* 516(7531), 355–60, Doi: 10.1038/nature14008.
- [195] Pflug, A., Lukarska, M., Resa-Infante, P., Reich, S., Cusack, S. (2017) Structural insights into RNA synthesis by the influenza virus transcription-replication machine. *Virus Res.* 234, 103–17, Doi: 10.1016/j.virusres.2017.01.013.
- [196] Pielak, R.M., Chou, J.J. (2011) Influenza M2 proton channels. *Biochim. Biophys. Acta* 1808(2), 522–9, Doi: 10.1016/j.bbamem.2010.04.015.
- [197] Plotch, S.J., Bouloy, M., Krug, R.M. (1979) Transfer of 5'-terminal cap of globin mRNA to influenza viral complementary RNA during transcription in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 76(4), 1618–22.
- [198] Plotch, S.J., Bouloy, M., Ulmanen, I., Krug, R.M. (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* 23(3), 847–58.
- [199] Pons, M.W. (1970) On the nature of the influenza virus genome. *Curr. Top. Microbiol. Immunol.* 52, 142–57.
- [200] Pons, M.W., Schulze, I.T., Hirst, G.K. (1969) Isolation and characterization of the ribonucleoprotein of influenza virus. *Virology* 39(2), 250–9, Doi: 10.1016/0042-6822(69)90045-2.
- [201] Poole, E., Elton, D., Medcalf, L., Digard, P. (2004) Functional domains of the influenza A virus PB2 protein: identification of NP- and PB1-binding sites. *Virology* 321(1), 120–33, Doi: 10.1016/j.virol.2003.12.022.
- [202] Poon, L.L., Pritlove, D.C., Fodor, E., Brownlee, G.G. (1999) Direct evidence that the poly(A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. *J. Virol.* 73(4), 3473–6.

- [203] Qiao, X., Qiao, J., Mindich, L. (2003) Analysis of Specific Binding Involved in Genomic Packaging of the Double-Stranded-RNA Bacteriophage ϕ 6. *J. Bacteriol.* 185(21), 6409–14, Doi: 10.1128/JB.185.21.6409-6414.2003.
- [204] Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M.F., Konrat, R., Bläsi, U., Schroeder, R. (2007) RNA chaperones, RNA annealers and RNA helicases. *RNA Biol.* 4(3), 118–30.
- [205] Rajkowitsch, L., Semrad, K., Mayer, O., Schroeder, R. (2005) Assays for the RNA chaperone activity of proteins. *Biochem. Soc. Trans.* 33(Pt 3), 450–6, Doi: 10.1042/BST0330450.
- [206] Reed, R. (2003) Coupling transcription, splicing and mRNA export. *Curr. Opin. Cell Biol.* 15(3), 326–31, Doi: 10.1016/S0955-0674(03)00048-6.
- [207] Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crépin, T., Hart, D., Lunardi, T., Nanao, M., Ruigrok, R.W.H., Cusack, S. (2014) Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* 516(7531), 361–6, Doi: 10.1038/nature14009.
- [208] Resa-Infante, P., Jorba, N., Coloma, R., Ortin, J. (2011) The influenza virus RNA synthesis machine: advances in its structure and function. *RNA Biol.* 8(2), 207–15.
- [209] Reuter, J.S., Mathews, D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11, 129, Doi: 10.1186/1471-2105-11-129.
- [210] Reyes, F.E., Garst, A.D., Batey, R.T. (2009) Strategies in RNA crystallography. *Methods Enzymol.* 469, 119–39, Doi: 10.1016/S0076-6879(09)69006-6.
- [211] Richardson, J.C., Akkina, R.K. (1991) NS2 protein of influenza virus is found in purified virus and phosphorylated in infected cells. *Arch. Virol.* 116(1–4), 69–80.
- [212] Robb, N.C., Chase, G., Bier, K., Vreede, F.T., Shaw, P.-C., Naffakh, N., Schwemmle, M., Fodor, E. (2011) The influenza A virus NS1 protein interacts with the nucleoprotein of viral ribonucleoprotein complexes. *J. Virol.* 85(10), 5228–31, Doi: 10.1128/JVI.02562-10.
- [213] Robertson, J.S., Schubert, M., Lazzarini, R.A. (1981) Polyadenylation sites for influenza virus mRNA. *J. Virol.* 38(1), 157–63.
- [214] Rossman, J.S., Jing, X., Leser, G.P., Balannik, V., Pinto, L.H., Lamb, R.A. (2010) Influenza Virus M2 Ion Channel Protein Is Necessary for Filamentous Virion Formation. *J. Virol.* 84(10), 5078–88, Doi: 10.1128/JVI.00119-10.
- [215] Rossman, J.S., Jing, X., Leser, G.P., Lamb, R.A. (2010) Influenza virus M2 protein mediates ESCRT-independent membrane scission. *Cell* 142(6), 902–13, Doi: 10.1016/j.cell.2010.08.029.
- [216] Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., Weissman, J.S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505(7485), 701–5, Doi: 10.1038/nature12894.
- [217] Roy, P. (2017) Bluetongue virus structure and assembly. *Curr. Opin. Virol.* 24, 115–23, Doi: 10.1016/j.coviro.2017.05.003.
- [218] Ruigrok, R.W.H., Baudin, F. (1995) Structure of influenza virus ribonucleoprotein particles. II. Purified RNA-free influenza virus

- ribonucleoprotein forms structures that are indistinguishable from the intact influenza virus ribonucleoprotein particles. *J. Gen. Virol.* 76(4), 1009–14, Doi: 10.1099/0022-1317-76-4-1009.
- [219] Rust, M.J., Lakadamyali, M., Zhang, F., Zhuang, X. (2004) Assembly of endocytic machinery around individual influenza viruses during viral entry. *Nat. Struct. Mol. Biol.* 11(6), 567–73, Doi: 10.1038/nsmb769.
- [220] Ruskowska, A., Lenartowicz, E., Moss, W.N., Kierzek, R., Kierzek, E. (2016) Secondary structure model of the naked segment 7 influenza A virus genomic RNA. *Biochem. J.* 473(23), 4327–48, Doi: 10.1042/BCJ20160651.
- [221] Sakaguchi, A., Hirayama, E., Hiraki, A., Ishida, Y., Kim, J. (2003) Nuclear export of influenza viral ribonucleoprotein is temperature-dependently inhibited by dissociation of viral matrix protein. *Virology* 306(2), 244–53.
- [222] Schmitt, A.P., Lamb, R.A. (2005) Influenza virus assembly and budding at the viral budzone. *Adv. Virus Res.* 64, 383–416, Doi: 10.1016/S0065-3527(05)64012-2.
- [223] Selman, M., Dankar, S.K., Forbes, N.E., Jia, J.-J., Brown, E.G. (2012) Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. *Emerg. Microbes Infect.* 1(11), e42, Doi: 10.1038/emi.2012.38.
- [224] Semrad, K., Green, R., Schroeder, R. (2004) RNA chaperone activity of large ribosomal subunit proteins from *Escherichia coli*. *RNA N. Y. N* 10(12), 1855–60, Doi: 10.1261/rna.7121704.
- [225] Sharma, K. kant., Didier, P., Darlix, J.L., de Rocquigny, H., Bensikaddour, H., Lavergne, J.-P., Pénin, F., Lessinger, J.-M., Mély, Y. (2010) Kinetic analysis of the nucleic acid chaperone activity of the hepatitis C virus core protein. *Nucleic Acids Res.* 38(11), 3632–42, Doi: 10.1093/nar/gkq094.
- [226] Shaw, M.L., Palese, P. (2013) Orthomyxoviridae. *Fields Virology* eds Knipe DM, Howley PM (Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, PA), pp. 1151–1185.
- [227] Shaw, M.L., Stone, K.L., Colangelo, C.M., Gulcicek, E.E., Palese, P. (2008) Cellular proteins in influenza virus particles. *PLoS Pathog.* 4(6), e1000085, Doi: 10.1371/journal.ppat.1000085.
- [228] Shimizu, T., Takizawa, N., Watanabe, K., Nagata, K., Kobayashi, N. (2011) Crucial role of the influenza virus NS2 (NEP) C-terminal domain in M1 binding and nuclear export of vRNP. *FEBS Lett.* 585(1), 41–6, Doi: 10.1016/j.febslet.2010.11.017.
- [229] Sieczkarski, S.B., Whittaker, G.R. (2002) Influenza virus can enter and infect cells in the absence of clathrin-mediated endocytosis. *J. Virol.* 76(20), 10455–64.
- [230] Skehel, J.J., Wiley, D.C. (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* 69, 531–69, Doi: 10.1146/annurev.biochem.69.1.531.
- [231] Smith, G.L., Hay, A.J. (1982) Replication of the influenza virus genome. *Virology* 118(1), 96–108, Doi: 10.1016/0042-6822(82)90323-3.
- [232] Smola, M.J., Rice, G.M., Busan, S., Siegfried, N.A., Weeks, K.M. (2015) Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling

- (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* 10(11), 1643–69, Doi: 10.1038/nprot.2015.103.
- [233] Smola, M.J., Weeks, K.M. (2018) In-cell RNA structure probing with SHAPE-MaP. *Nat. Protoc.* 13(6), 1181–95, Doi: 10.1038/nprot.2018.010.
- [234] Smyth, R.P., Despons, L., Huili, G., Bernacchi, S., Hijnen, M., Mak, J., Jossinet, F., Weixi, L., Paillart, J.-C., von Kleist, M., Marquet, R. (2015) Mutational interference mapping experiment (MIME) for studying RNA structure and function. *Nat. Methods* 12(9), 866–72, Doi: 10.1038/nmeth.3490.
- [235] Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T., Chang, H.Y. (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* 519(7544), 486–90, Doi: 10.1038/nature14263.
- [236] Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., Ule, J. (2012) Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.* 13(8), R67, Doi: 10.1186/gb-2012-13-8-r67.
- [237] Sugiyama, K., Obayashi, E., Kawaguchi, A., Suzuki, Y., Tame, J.R.H., Nagata, K., Park, S.-Y. (2009) Structural insight into the essential PB1-PB2 subunit contact of the influenza virus RNA polymerase. *EMBO J.* 28(12), 1803–11, Doi: 10.1038/emboj.2009.138.
- [238] Suzuki, Y. (2014) A possible packaging signal in the rotavirus genome. *Genes Genet. Syst.* 89(2), 81–6, Doi: 10.1266/ggs.89.81.
- [239] Swale, C., Monod, A., Tengo, L., Labaronne, A., Garzoni, F., Bourhis, J.-M., Cusack, S., Schoehn, G., Berger, I., Ruigrok, R.W.H., Crépin, T. (2016) Structural characterization of recombinant IAV polymerase reveals a stable complex between viral PA-PB1 heterodimer and host RanBP5. *Sci. Rep.* 6, 24727, Doi: 10.1038/srep24727.
- [240] Talkish, J., May, G., Lin, Y., Woolford, J.L., McManus, C.J. (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA N. Y.* N 20(5), 713–20, Doi: 10.1261/rna.042218.113.
- [241] Tarendeau, F., Crepin, T., Guilligay, D., Ruigrok, R.W.H., Cusack, S., Hart, D.J. (2008) Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog.* 4(8), e1000136, Doi: 10.1371/journal.ppat.1000136.
- [242] Tarus, B., Bakowicz, O., Chenavas, S., Duchemin, L., Estrozi, L.F., Bourdieu, C., Lejal, N., Bernard, J., Moudjou, M., Chevalier, C., Delmas, B., Ruigrok, R.W.H., Di Primo, C., Slama-Schwok, A. (2012) Oligomerization paths of the nucleoprotein of influenza A virus. *Biochimie* 94(3), 776–85, Doi: 10.1016/j.biochi.2011.11.009.
- [243] Taubenberger, J.K., Kash, J.C. (2010) Influenza Virus Evolution, Host Adaptation and Pandemic Formation. *Cell Host Microbe* 7(6), 440–51, Doi: 10.1016/j.chom.2010.05.009.
- [244] Tewawong, N., Prachayangprecha, S., Vichiwattana, P., Korkong, S., Klinfueng, S., Vongpunsawad, S., Thongmee, T., Theamboonlers, A., Poovorawan, Y. (2015) Assessing Antigenic Drift of Seasonal Influenza A(H3N2) and A(H1N1)pdm09 Viruses. *PLOS ONE* 10(10), e0139958, Doi: 10.1371/journal.pone.0139958.

- [245] Tobita, K. (1997) [Classification and nomenclature of influenza viruses]. *Nihon Rinsho Jpn. J. Clin. Med.* 55(10), 2512–4.
- [246] Tong, S., Li, Y., Rivaller, P., Conrardy, C., Castillo, D.A.A., Chen, L.-M., Recuenco, S., Ellison, J.A., Davis, C.T., York, I.A., Turmelle, A.S., Moran, D., Rogers, S., Shi, M., Tao, Y., Weil, M.R., Tang, K., Rowe, L.A., Sammons, S., Xu, X., Frace, M., Lindblade, K.A., Cox, N.J., Anderson, L.J., Rupprecht, C.E., Donis, R.O. (2012) A distinct lineage of influenza A virus from bats. *Proc. Natl. Acad. Sci. U. S. A.* 109(11), 4269–74, Doi: 10.1073/pnas.1116200109.
- [247] Tong, S., Zhu, X., Li, Y., Shi, M., Zhang, J., Bourgeois, M., Yang, H., Chen, X., Recuenco, S., Gomez, J., Chen, L.-M., Johnson, A., Tao, Y., Dreyfus, C., Yu, W., McBride, R., Carney, P.J., Gilbert, A.T., Chang, J., Guo, Z., Davis, C.T., Paulson, J.C., Stevens, J., Rupprecht, C.E., Holmes, E.C., Wilson, I.A., Donis, R.O. (2013) New World Bats Harbor Diverse Influenza A Viruses. *PLOS Pathog.* 9(10), e1003657, Doi: 10.1371/journal.ppat.1003657.
- [248] Tsurudome, M., Glück, R., Graf, R., Falchetto, R., Schaller, U., Brunner, J. (1992) Lipid interactions of the hemagglutinin HA2 NH₂-terminal segment during influenza virus-induced membrane fusion. *J. Biol. Chem.* 267(28), 20225–32.
- [249] Turell, L., Hutchinson, E.C., Vreede, F.T., Fodor, E. (2015) Regulation of influenza A virus nucleoprotein oligomerization by phosphorylation. *J. Virol.* 89(2), 1452–5, Doi: 10.1128/JVI.02332-14.
- [250] Vale-Costa, S., Alenquer, M., Sousa, A.L., Kellen, B., Ramalho, J., Tranfield, E.M., Amorim, M.J. (2016) Influenza A virus ribonucleoproteins modulate host recycling by competing with Rab11 effectors. *J. Cell Sci.* 129(8), 1697–710, Doi: 10.1242/jcs.188409.
- [251] Vasin, A.V., Temkina, O.A., Egorov, V.V., Klotchenko, S.A., Plotnikova, M.A., Kiselev, O.I. (2014) Molecular mechanisms enhancing the proteome of influenza A viruses: an overview of recently discovered proteins. *Virus Res.* 185, 53–63, Doi: 10.1016/j.virusres.2014.03.015.
- [252] Vidic, J., Noiray, M., Bagchi, A., Slama-Schwok, A. (2016) Identification of a Novel Complex between the Nucleoprotein and PA(1-27) of Influenza A Virus Polymerase. *Biochemistry (Mosc.)* 55(31), 4259–62, Doi: 10.1021/acs.biochem.6b00514.
- [253] Vogel, U., Scholtissek, C. (1995) Inhibition of the intracellular transport of influenza viral RNA by actinomycin D. *Arch. Virol.* 140(10), 1715–23.
- [254] Vreede, F.T., Gifford, H., Brownlee, G.G. (2008) Role of initiating nucleoside triphosphate concentrations in the regulation of influenza virus replication and transcription. *J. Virol.* 82(14), 6902–10, Doi: 10.1128/JVI.00627-08.
- [255] Vreede, F.T., Jung, T.E., Brownlee, G.G. (2004) Model Suggesting that Replication of Influenza Virus Is Regulated by Stabilization of Replicative Intermediates. *J. Virol.* 78(17), 9568–72, Doi: 10.1128/JVI.78.17.9568-9572.2004.
- [256] de Vries, E., Tscherne, D.M., Wienholts, M.J., Cobos-Jiménez, V., Scholte, F., García-Sastre, A., Rottier, P.J.M., de Haan, C.A.M. (2011) Dissection of the influenza A virus endocytic routes reveals macropinocytosis as an alternative entry pathway. *PLoS Pathog.* 7(3), e1001329, Doi: 10.1371/journal.ppat.1001329.

- [257] Wan, Y., Qu, K., Ouyang, Z., Chang, H.Y. (2013) Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nat. Protoc.* 8(5), 849–69, Doi: 10.1038/nprot.2013.045.
- [258] Wang, D., Harmon, A., Jin, J., Francis, D.H., Christopher-Hennings, J., Nelson, E., Montelaro, R.C., Li, F. (2010) The lack of an inherent membrane targeting signal is responsible for the failure of the matrix (M1) protein of influenza A virus to bud into virus-like particles. *J. Virol.* 84(9), 4673–81, Doi: 10.1128/JVI.02306-09.
- [259] Wang, P., Palese, P., O'Neill, R.E. (1997) The NPI-1/NPI-3 (karyopherin alpha) binding site on the influenza A virus nucleoprotein NP is a nonconventional nuclear localization signal. *J. Virol.* 71(3), 1850–6.
- [260] Wang, W., Cui, Z.-Q., Han, H., Zhang, Z.-P., Wei, H.-P., Zhou, Y.-F., Chen, Z., Zhang, X.-E. (2008) Imaging and characterizing influenza A virus mRNA transport in living cells. *Nucleic Acids Res.* 36(15), 4913–28, Doi: 10.1093/nar/gkn475.
- [261] Watanabe, K., Takizawa, N., Katoh, M., Hoshida, K., Kobayashi, N., Nagata, K. (2001) Inhibition of nuclear export of ribonucleoprotein complexes of influenza virus by leptomycin B. *Virus Res.* 77(1), 31–42.
- [262] Weber, F., Kochs, G., Gruber, S., Haller, O. (1998) A classical bipartite nuclear localization signal on Thogoto and influenza A virus nucleoproteins. *Virology* 250(1), 9–18, Doi: 10.1006/viro.1998.9329.
- [263] Weber, T., Paesold, G., Galli, C., Mischler, R., Semenza, G., Brunner, J. (1994) Evidence for H(+)-induced insertion of influenza hemagglutinin HA2 N-terminal segment into viral membrane. *J. Biol. Chem.* 269(28), 18353–8.
- [264] Westhof, E. (2015) Twenty years of RNA crystallography. *RNA* 21(4), 486–7, Doi: 10.1261/rna.049726.115.
- [265] Wharton, S.A., Belshe, R.B., Skehel, J.J., Hay, A.J. (1994) Role of virion M2 protein in influenza virus uncoating: specific reduction in the rate of membrane fusion between virus and liposomes by amantadine. *J. Gen. Virol.* 75 (Pt 4), 945–8, Doi: 10.1099/0022-1317-75-4-945.
- [266] White, J.M., Wilson, I.A. (1987) Anti-peptide antibodies detect steps in a protein conformational change: low-pH activation of the influenza virus hemagglutinin. *J. Cell Biol.* 105(6 Pt 2), 2887–96.
- [267] Whittaker, G., Bui, M., Helenius, A. (1996) Nuclear trafficking of influenza virus ribonucleoproteins in heterokaryons. *J. Virol.* 70(5), 2743–56.
- [268] WHO. (1980) A revision of the system of nomenclature for influenza viruses: a WHO Memorandum. *Bull. World Health Organ.* 58(4), 585–91.
- [269] Wickramasinghe, V.O., Laskey, R.A. (2015) Control of mammalian gene expression by selective mRNA export. *Nat. Rev. Mol. Cell Biol.* 16(7), 431–42, Doi: 10.1038/nrm4010.
- [270] Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C., Weeks, K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* 6(4), e96, Doi: 10.1371/journal.pbio.0060096.

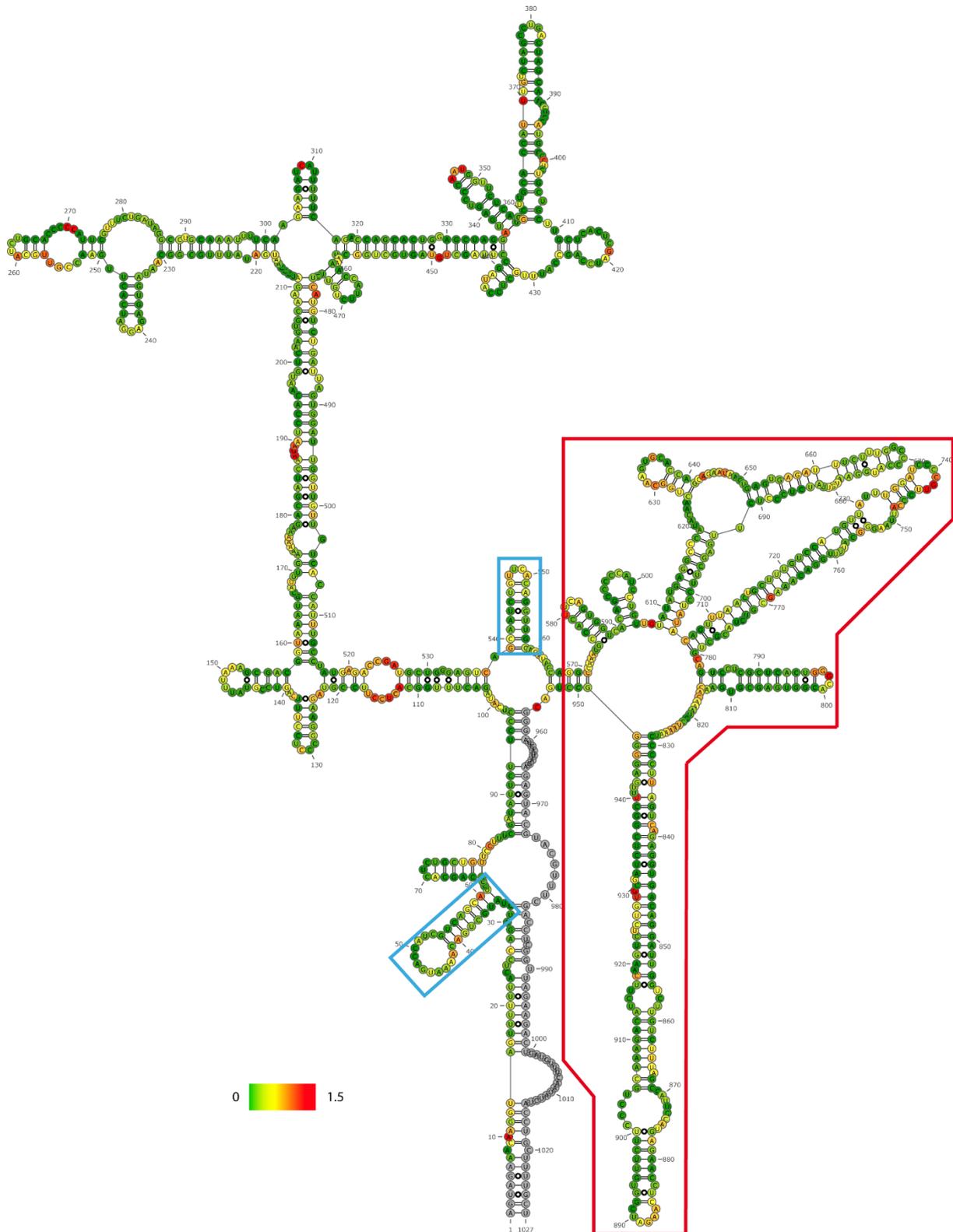
- [271] Wilkinson, K.A., Merino, E.J., Weeks, K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* 1(3), 1610–6, Doi: 10.1038/nprot.2006.249.
- [272] Williams, G.D., Townsend, D., Wylie, K.M., Kim, P.J., Amarasinghe, G.K., Kutluay, S.B., Boon, A.C.M. (2018) Nucleotide resolution mapping of influenza A virus nucleoprotein-RNA interactions reveals RNA features required for replication. *Nat. Commun.* 9(1), 465, Doi: 10.1038/s41467-018-02886-w.
- [273] Wu, C.-Y., Jeng, K.-S., Lai, M.M.-C. (2011) The SUMOylation of Matrix Protein M1 Modulates the Assembly and Morphogenesis of Influenza A Virus. *J. Virol.* 85(13), 6618–28, Doi: 10.1128/JVI.02401-10.
- [274] Yamanaka, K., Ishihama, A., Nagata, K. (1990) Reconstitution of influenza virus RNA-nucleoprotein complexes structurally resembling native viral ribonucleoprotein cores. *J. Biol. Chem.* 265(19), 11151–5.
- [275] Yang, J., Xia, H., Qian, Q., Zhou, X. (2015) RNA chaperones encoded by RNA viruses. *Virol. Sin.* 30(6), 401–9, Doi: 10.1007/s12250-015-3676-2.
- [276] Yang, Y., Wang, S. (2017) RNA Characterization by Solid-State NMR Spectroscopy. *Chem. Weinh. Bergstr. Ger.*, Doi: 10.1002/chem.201705583.
- [277] Yasuda, J., Nakada, S., Kato, A., Toyoda, T., Ishihama, A. (1993) Molecular assembly of influenza virus: association of the NS2 protein with virion matrix. *Virology* 196(1), 249–55, Doi: 10.1006/viro.1993.1473.
- [278] Ye, Q., Krug, R.M., Tao, Y.J. (2006) The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* 444(7122), 1078–82, Doi: 10.1038/nature05379.
- [279] Ye, Z., Robinson, D., Wagner, R.R. (1995) Nucleus-targeting domain of the matrix protein (M1) of influenza virus. *J. Virol.* 69(3), 1964–70.
- [280] Yondola, M.A., Fernandes, F., Belicha-Villanueva, A., Uccellini, M., Gao, Q., Carter, C., Palese, P. (2011) Budding Capability of the Influenza Virus Neuraminidase Can Be Modulated by Tetherin. *J. Virol.* 85(6), 2480–91, Doi: 10.1128/JVI.02188-10.
- [281] York, A., Hengrung, N., Vreede, F.T., Huiskonen, J.T., Fodor, E. (2013) Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* 110(45), E4238–4245, Doi: 10.1073/pnas.1315068110.
- [282] Yoshimura, A., Kuroda, K., Kawasaki, K., Yamashina, S., Maeda, T., Ohnishi, S.-I. (1982) Infectious Cell Entry Mechanism of Influenza Virus. *J. Virol.* 43(1), 284–93.
- [283] Yu, M., Liu, X., Cao, S., Zhao, Z., Zhang, K., Xie, Q., Chen, C., Gao, S., Bi, Y., Sun, L., Ye, X., Gao, G.F., Liu, W. (2012) Identification and characterization of three novel nuclear export signals in the influenza A virus nucleoprotein. *J. Virol.* 86(9), 4970–80, Doi: 10.1128/JVI.06159-11.
- [284] Yuan, P., Bartlam, M., Lou, Z., Chen, S., Zhou, J., He, X., Lv, Z., Ge, R., Li, X., Deng, T., Fodor, E., Rao, Z., Liu, Y. (2009) Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature* 458(7240), 909–13, Doi: 10.1038/nature07720.

- [285] Zhang, C., Darnell, R.B. (2011) Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* 29(7), 607–14, Doi: 10.1038/nbt.1873.
- [286] Zhang, K., Keane, S.C., Su, Z., Irobalieva, R.N., Chen, M., Van, V., Sciandra, C.A., Marchant, J., Heng, X., Schmid, M.F., Case, D.A., Ludtke, S.J., Summers, M.F., Chiu, W. (2018) Structure of the 30 kDa HIV-1 RNA Dimerization Signal by a Hybrid Cryo-EM, NMR, and Molecular Dynamics Approach. *Structure* 26(3), 490-498.e3, Doi: 10.1016/j.str.2018.01.001.
- [287] Zhirnov, O.P. (1990) Solubilization of matrix protein M1/M from virions occurs at different pH for orthomyxo- and paramyxoviruses. *Virology* 176(1), 274–9.
- [288] Zhirnov, O.P., Klenk, H.D. (1997) Histones as a target for influenza virus matrix protein M1. *Virology* 235(2), 302–10, Doi: 10.1006/viro.1997.8700.
- [289] Zubradt, M., Gupta, P., Persad, S., Lambowitz, A.M., Weissman, J.S., Rouskin, S. (2017) DMS-MaPseq for genome-wide or targeted RNA structure probing *in vivo*. *Nat. Methods* 14(1), 75–82, Doi: 10.1038/nmeth.4057.

ANNEXES

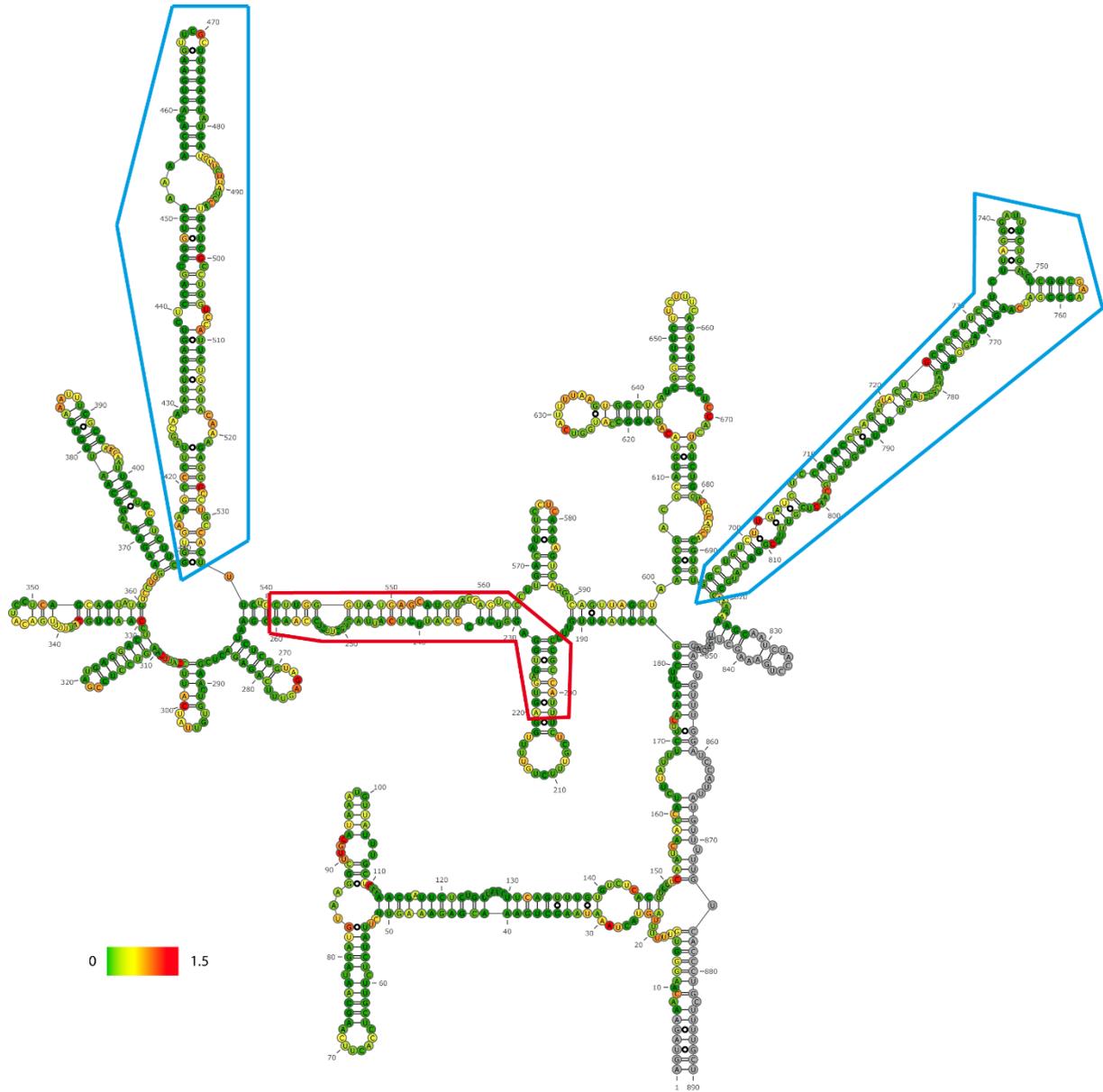
Annexe 1

La réactivité des nucléotides de l'ARNv M de la souche PR8 dans la condition NoNP est reportée sur la structure de l'ARN en condition NoNP



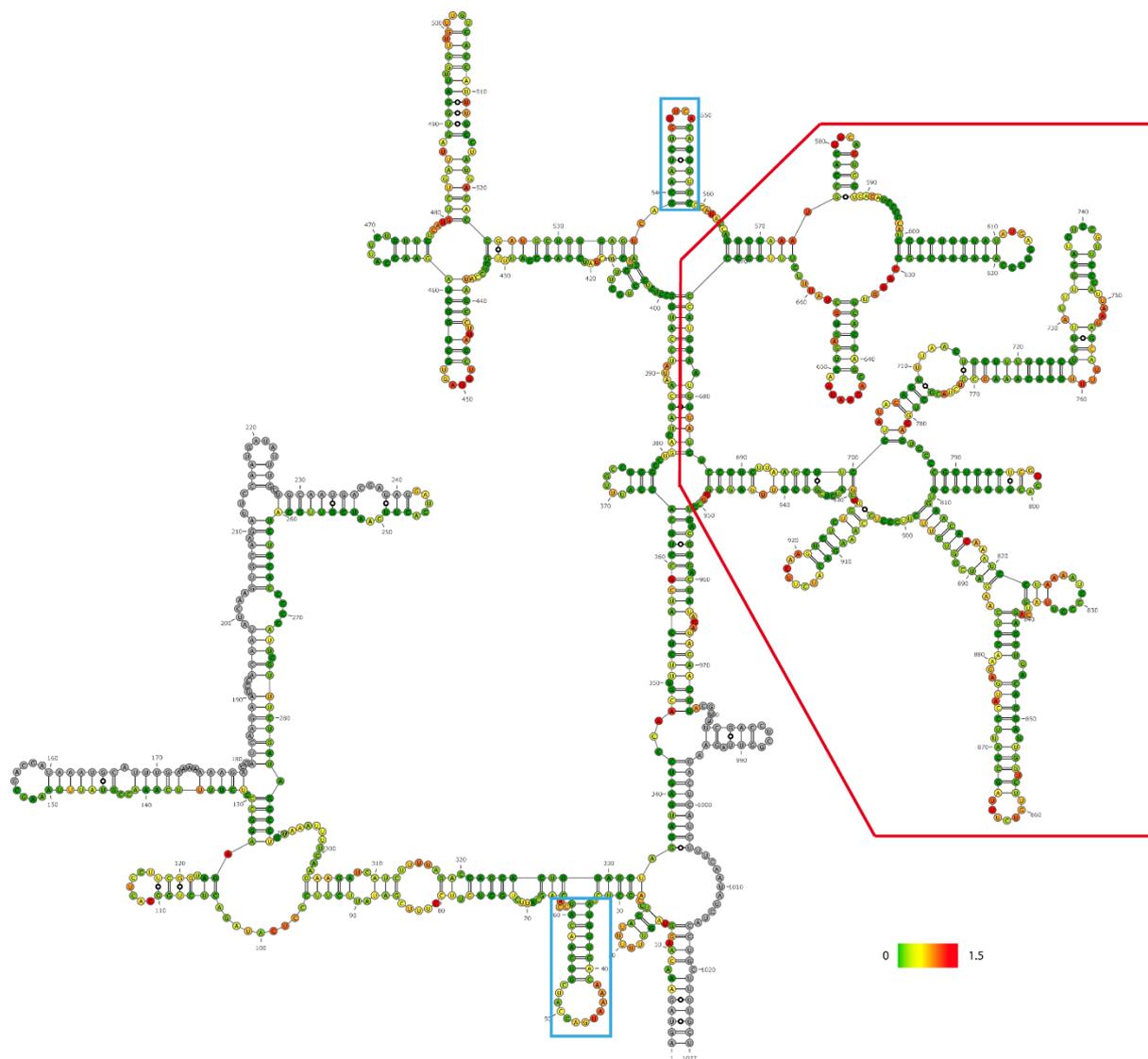
Annexe 2

La réactivité des nucléotides de l'ARNv NS de la souche PR8 dans la condition NoNP est reportée sur la structure de l'ARN en condition NoNP



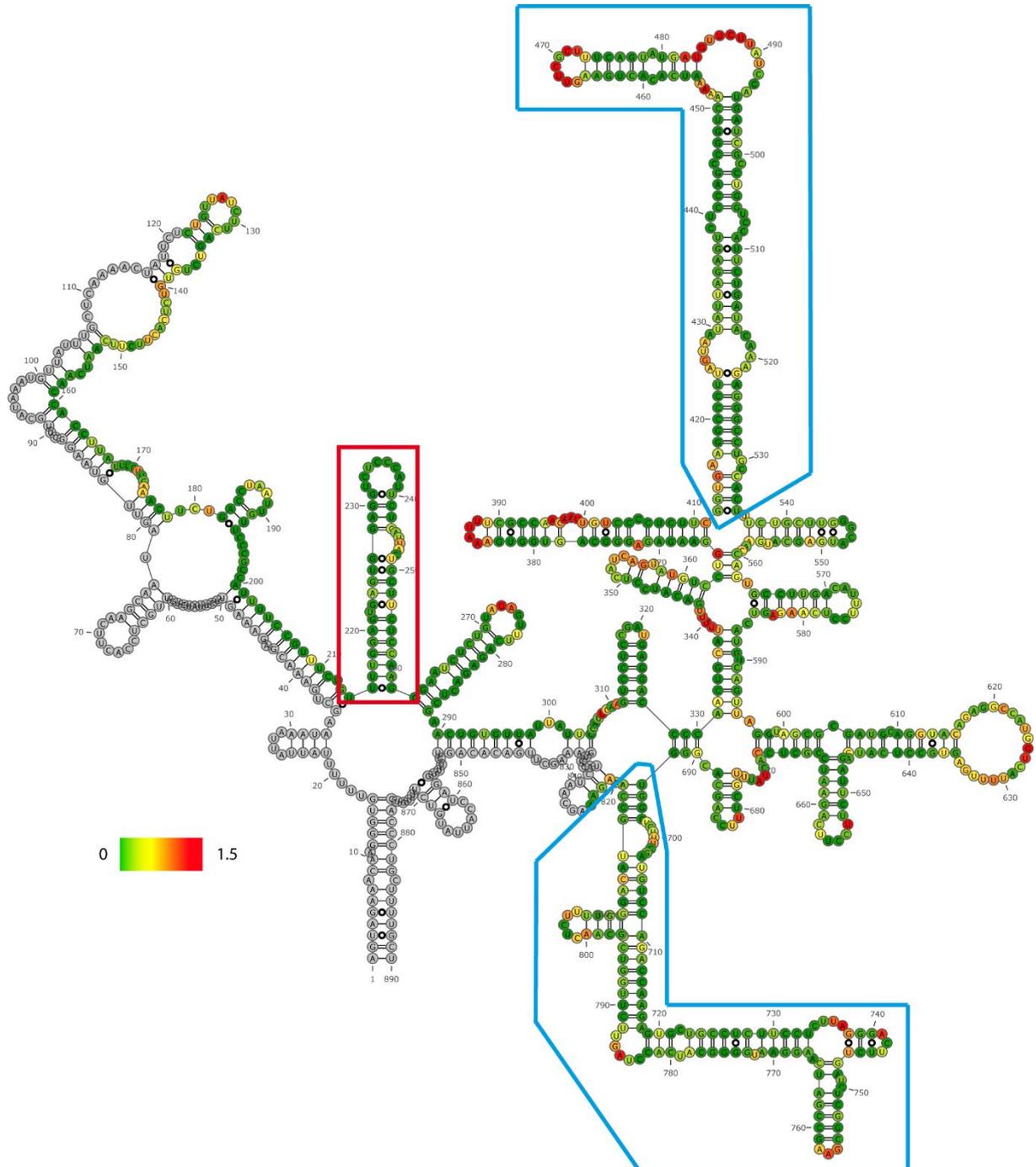
Annexe 3

La réactivité des nucléotides de l'ARNv M de la souche WSN dans la condition NoNP est reportée sur la structure de l'ARN en condition NoNP



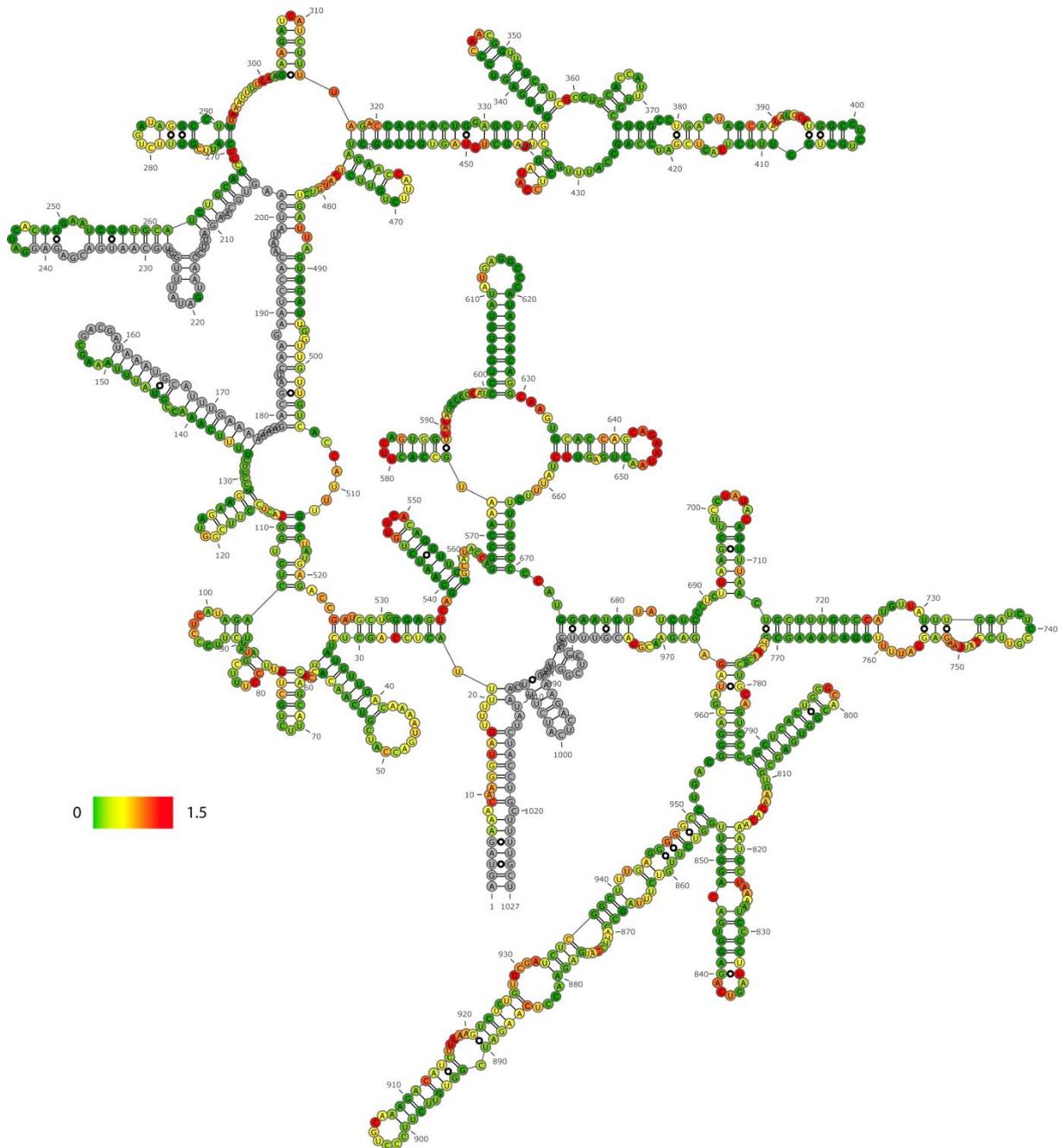
Annexe 4

La réactivité des nucléotides de l'ARNv NS de la souche WSN dans la condition NoNP est reportée sur la structure de l'ARN en condition NoNP



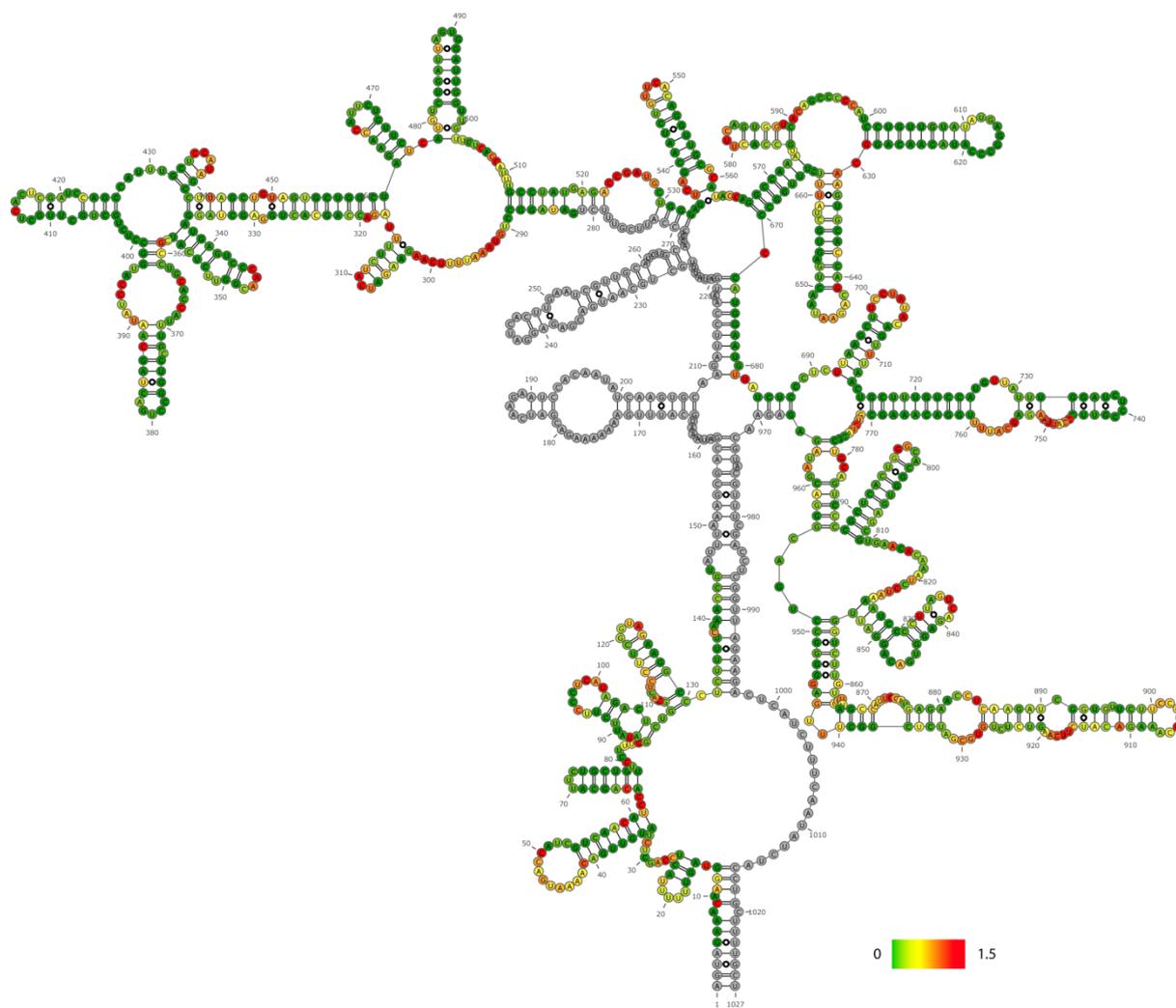
Annexe 5

La réactivité des nucléotides de l'ARNv M de la souche WSN dans la condition PK monomérique est reportée sur la structure de l'ARN en condition PK monomérique



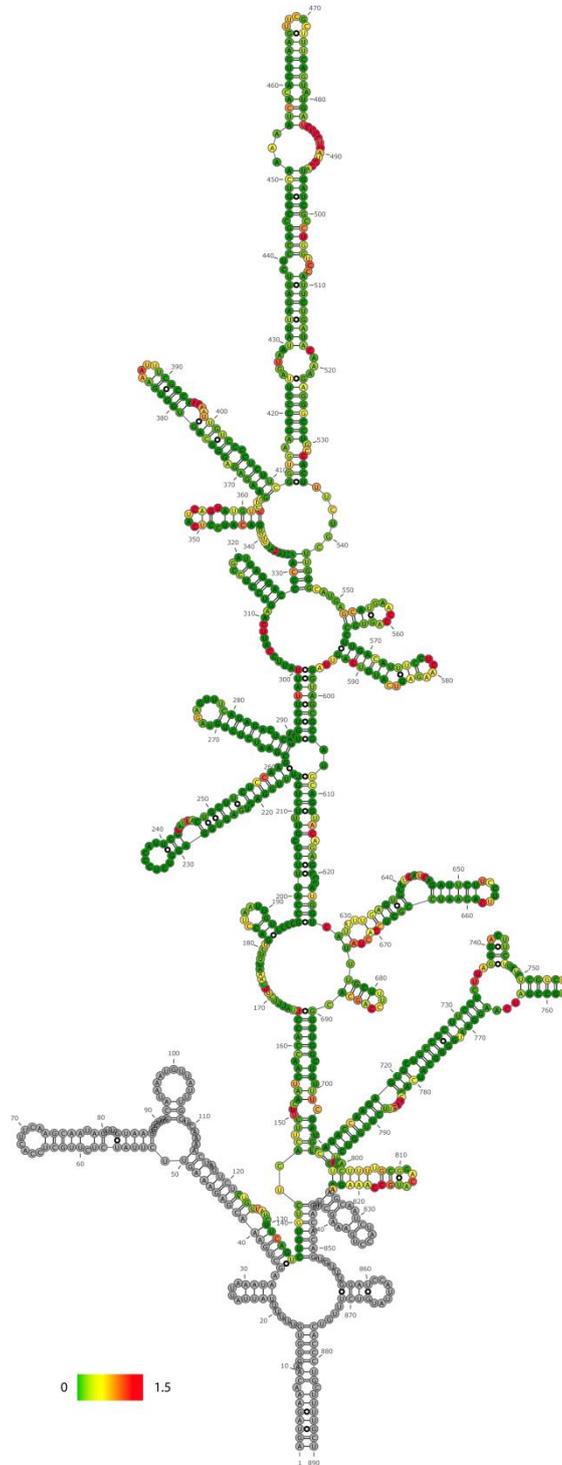
Annexe 6

La réactivité des nucléotides de l'ARNv M de la souche WSN dans la condition PK trimérique est reportée sur la structure de l'ARN en condition PK trimérique



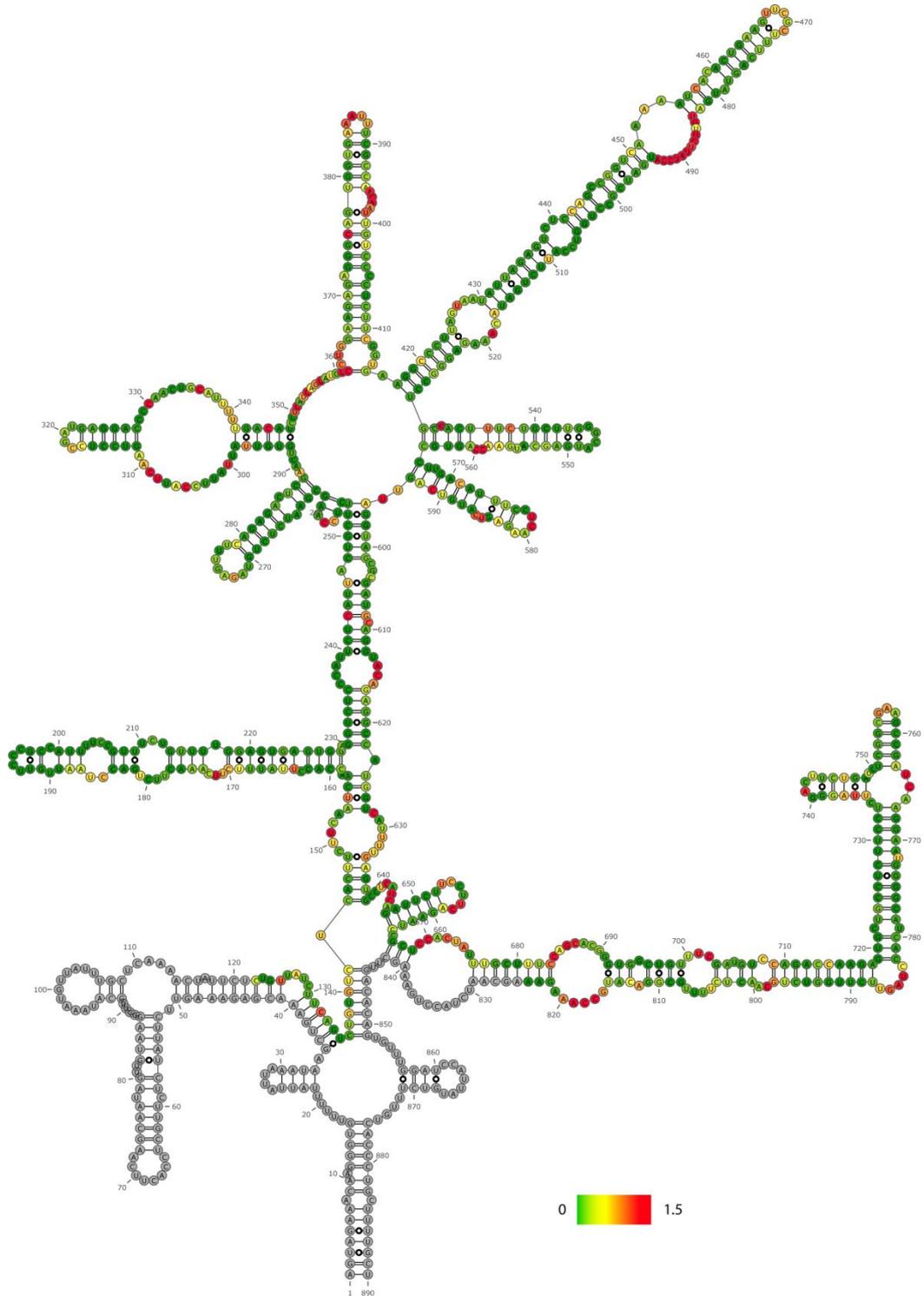
Annexe 7

La réactivité des nucléotides de l'ARNv NS de la souche WSN dans la condition PK monomérique est reportée sur la structure de l'ARN en condition PK monomérique



Annexe 8

La réactivité des nucléotides de l'ARNv NS de la souche WSN dans la condition PK trimérique est reportée sur la structure de l'ARN en condition PK trimérique



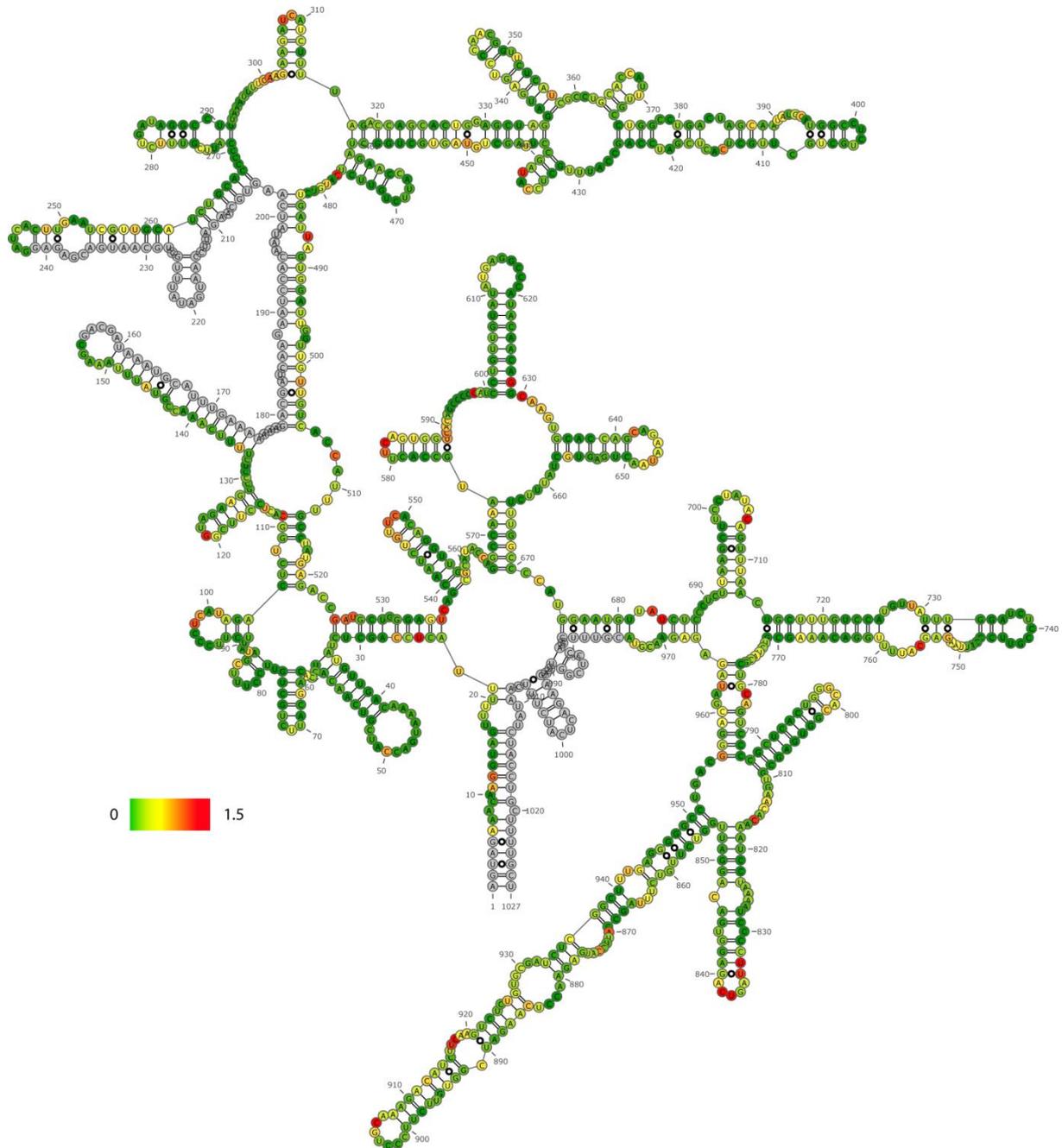
Annexe 9

Liste des régions protégées par la NP WSN (monomérique et trimérique) et PR8. En violet sont indiquées les positions des nucléotides sur l'ARN dans le sens 5' -> 3'

WSN		WSN		PR8	
Protéine NP monomérique		Protéine NP trimérique		Protéine NP monomérique	
M :	NS :	M :	NS :	M :	NS :
97 C	140 G	11 A	138 G	10 A	9 C
136 U	143 U	14 U	164 U	27 C	22 U
305 A	181 U	32 C	181 U	131 C	28 U
310 A	185 C	59 C	185 C	168 U	71 C
384 U	243 C	80 C	243 C	176 A	132 C
486 U	246 U	88 U	246 U	220 A	152 C
509 A	273 G	100 A	297 U	290 U	156 C
538 A	297 U	114 U	300 U	337 A	160 C
550 A	300 U	284 U	330 C	395 A	165 U
560 C	305 C	310 A	341 U	409 U	173 C
580 U	330 C	506 A	420 C	419 G	200 A
690 U	341 U	529 C	494 A	426 C	205 U
798 G	345 C	539 G	510 U	448 U	208 U
814 C	469 C	639 C	519 A	458 G	236 C
847 C	488 U	672 C	532 C	506 A	243 C
319-320 AC	547 C	798 G	538 C	580 U	258 A
358-359 CG	591 C	944 A	571 C	635 G	271 U
389-390 AU	624 U	946 G	578 U	643 A	305 C
394-395 CA	628 C	94-95 UC	594 U	715 U	330 C
497-498 GU	714 C	116-117 CU	628 C	725 A	351 C
591-592 CA	741 A	318-319 GA	644 U	839 A	405 C
639-640 CA	754 G	346-347 AA	654 C	863 U	419 G
656-657 GC	798 C	393-394 CC	709 C	940 U	421 C
692-693 UU	824 A	448-449 UG	714 C	946 G	448 G
724-725 CA	471-472 CU	465-466 CC	741 A	948 G	493 C
965-966 GA	502-503 CU	479-480 UG	765 U	39-40 GA	500 G
7-10 ACA	531-532 CC	560-561 CA	812 A	42-43 AA	519 A
62-64 CCA	552-553 GC	565-566 CA	169-170 UC	100-101 AU	533 A
268-270 CCC	585-586 UC	587-588 UC	256-257 CC	134-135 CU	546 G
284-286 UAG	594-595 UU	729-730 UA	317-318 CC	159-160 GU	560 C
315-317 UUA	614-615 AC	928-929 UG	442-443 CA	186-187 AA	595 U
345-347 CAA	641-642 UC	931-932 GA	489-490 UA	400-401 CU	615 C
417-419 UCG	644-645 UC	26-28 CCA	585-586 UU	539-540 GC	652 U
435-437 CCA	653-654 UC	61-63 UCC	608-609 GC	560-561 CA	673 U
448-450 UGU	657-658 UC	83-85 UUC	614-615 AC	606-607 UG	720 U
478-480 AUG	683-684 CC	435-437 CCA	641-642 UC	617-618 CC	724 G
522-524 ACC	686-687 GC	508-510 CAU	657-658 UC	657-658 AG	798 C
659-661 AUU	701-702 UC	523-525 CCG	686-687 GC	747-748 CA	801 C
702-704 UAU	764-765 UC	580-582 UUC	783-784 CU	754-755 GG	809 C
758-760 UUU	812-813 AC	631-633 AAG	797-798 GC	769-770 GC	24-25 GU
772-774 UCU	150-151 UUU	691-694 CUUA	304-306 CCA	814-815 CA	52-53 CU
823-825 UAA	256-257 CC	772-775 UCUA	140-143 GUCU	823-824 UA	109-110 UC
947-949 GGG	308-309 UC	884-886 UCAA	361-364 UCCU	874-875 CA	143-144 UC
973-975 GUA	350-351 UC	901-904 CCCU	557-560 AACU	924-925 UC	308-309 UC
442-445 CUUA	361-363 UCC	658-662 UAUUU	631-634 UUUU	928-929 UG	336-337 CA
466-469 CAUU	395-397 ACA	700-704 CCUAU	816-820 GCCAA	58-60 GCA	394-395 AA
13-17 GUAGU	506-508 UCC	757-761 AUUUU	669-674 CCACUA	259-261 CAU	470-471 GC
928-932 UGCGA	558-560 ACC	747-752 CAUUA	-	477-479 CAU	488-489 UU
292-297 UAAAUU	578-580 UCA	820-825 UCCUAA	-	569-571 CCA	628-629 CA
643-648 AGAAUA	735-737 UUA	871-876 UCCAU	-	741-743 CGU	669-670 CC
42-48 AAAAUGA	806-808 UUG	642-647 CAGAAU	-	953-955 GAC	683-684 CC
75-82 CUGUCCU	816-818 GCC	41-50 CAAAAUGACC	-	627-630 UGGC	82-84 UGU
741-752 CGUCCCAUUA	483-486 UGUU	291-302 GUAAAAUUUCAA	-	520-524 AGACC	299-301 UCA
	631-634 UUUU			702-706 UAUAC	344-346 ACA
	783-786 CUAG			885-889 CAAGA	414-416 UGA
	669-674 CCACUA			112-117 ACUCCU	526-528 GCC
				76-82 UGUCCU	551-553 AGC
					577-579 CUC
					632-634 UUA
					686-688 GCA
					700-702 CUU
					16-19 GUUU
					90-93 UUGC
					361-364 UCCU
					483-486 UGUU
					506-509 UCCA

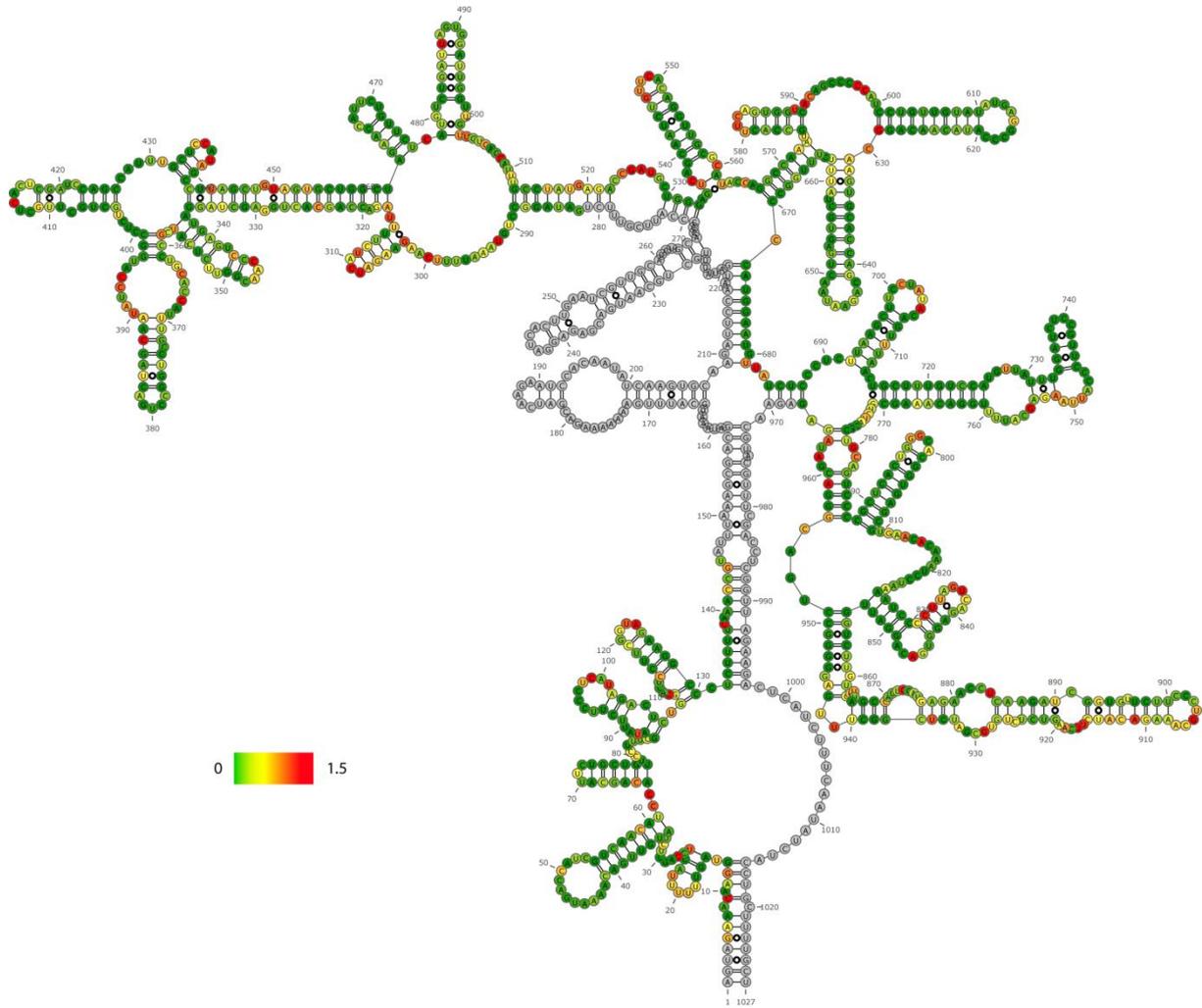
Annexe 10

La réactivité des nucléotides de l'ARNv M de la souche WSN dans la condition NP monomérique est reportée sur la structure de l'ARN en condition PK monomérique



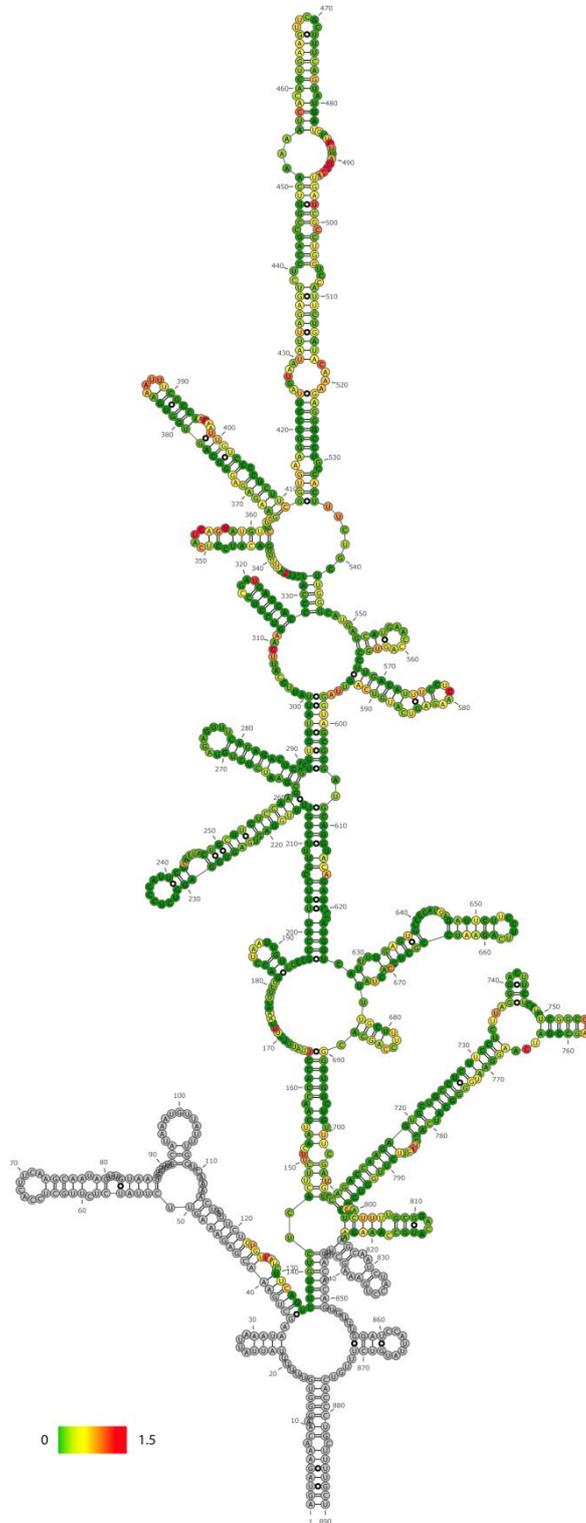
Annexe 11

La réactivité des nucléotides de l'ARNv M de la souche WSN dans la condition NP trimérique est reportée sur la structure de l'ARN en condition PK trimérique



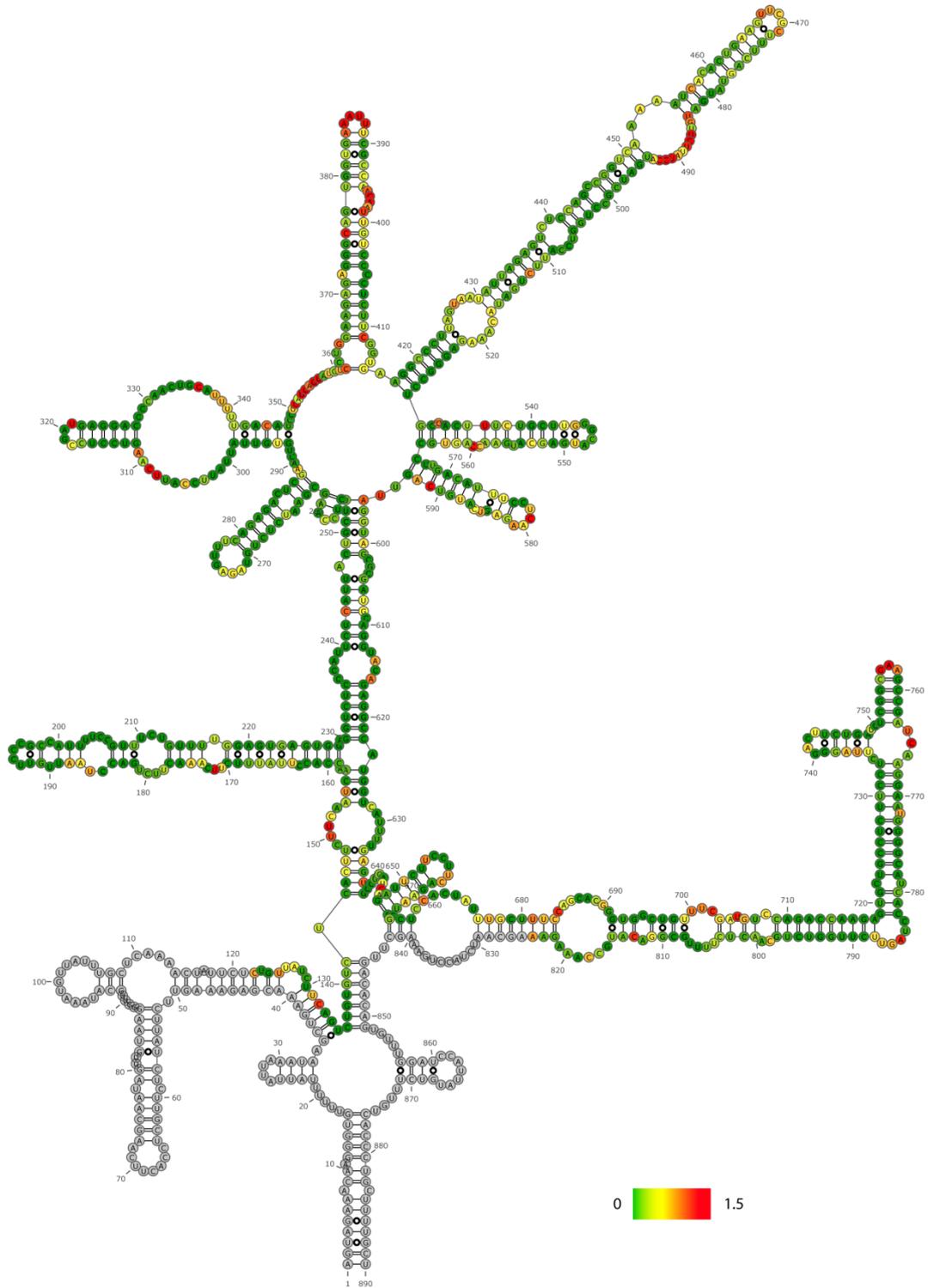
Annexe 12

La réactivité des nucléotides de l'ARNv NS de la souche WSN dans la condition NP monomérique est reportée sur la structure de l'ARN en condition PK monomérique



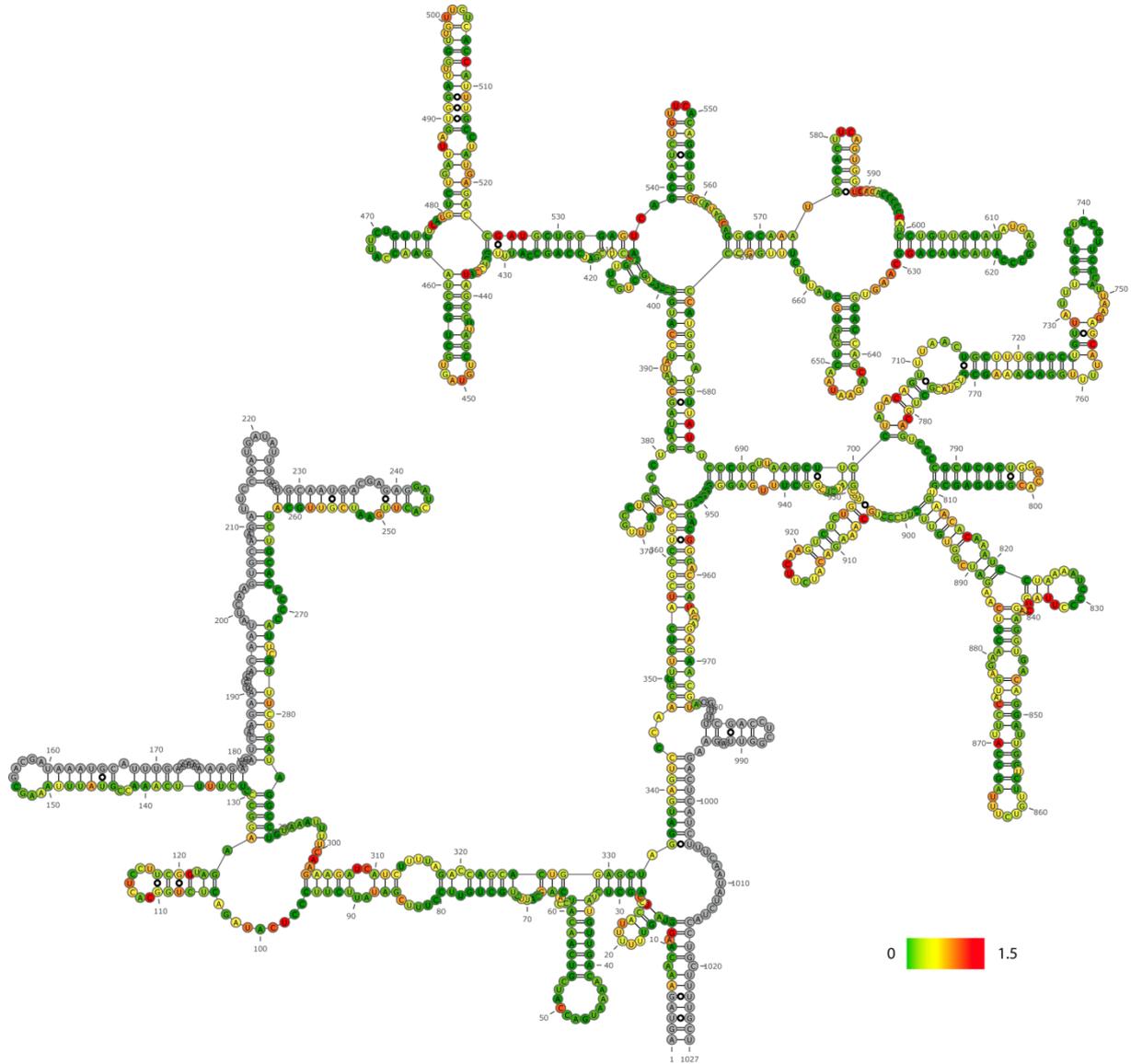
Annexe 13

La réactivité des nucléotides de l'ARNv NS de la souche WSN dans la condition NP trimérique est reportée sur la structure de l'ARN en condition PK trimérique



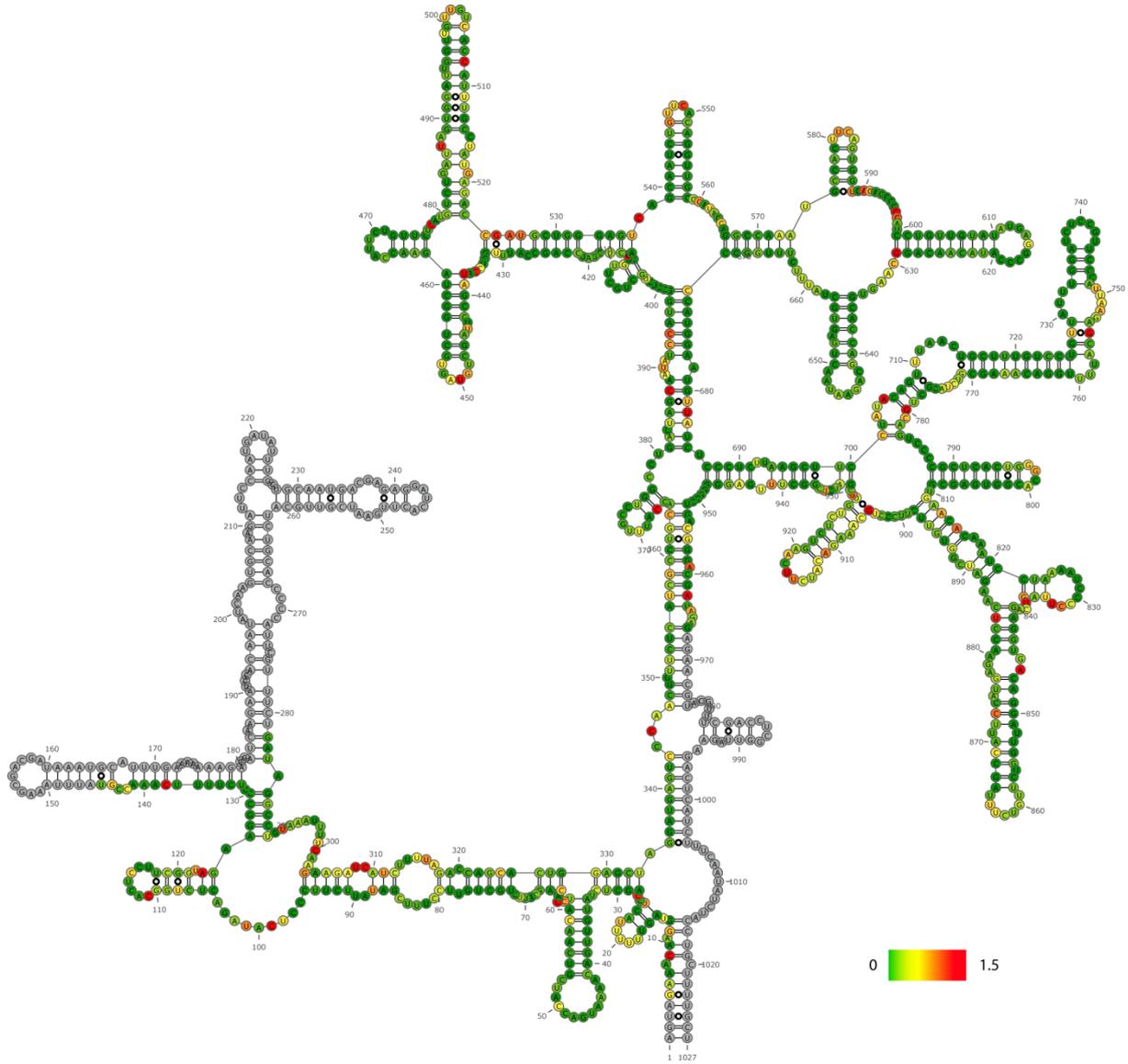
Annexe 14

La réactivité des nucléotides de l'ARNv M de la souche WSN dans la condition NP monomérique est reportée sur la structure de l'ARN en condition NoNP



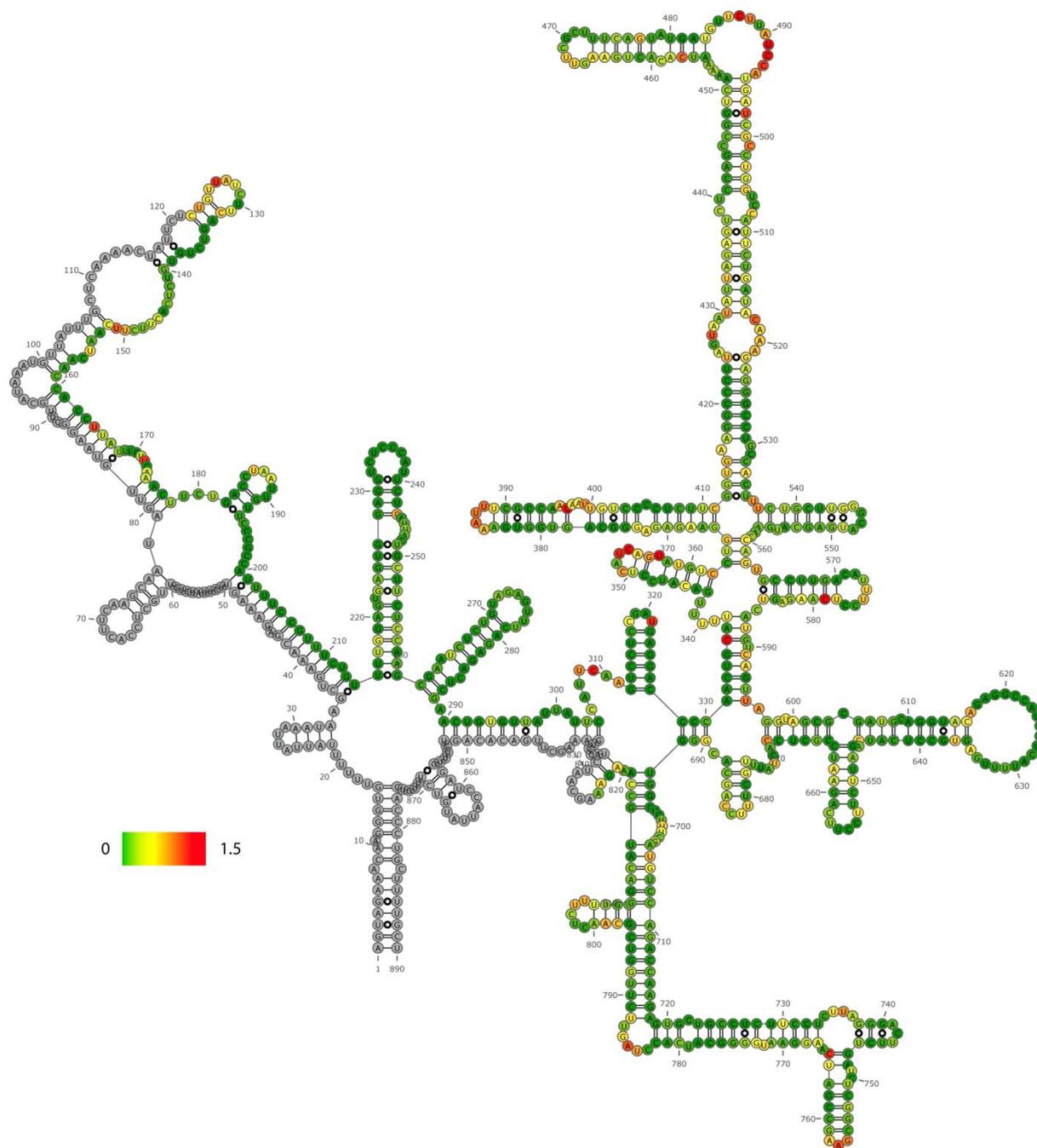
Annexe 15

La réactivité des nucléotides de l'ARNv M de la souche WSN dans la condition NP trimérique est reportée sur la structure de l'ARN en condition NoNP



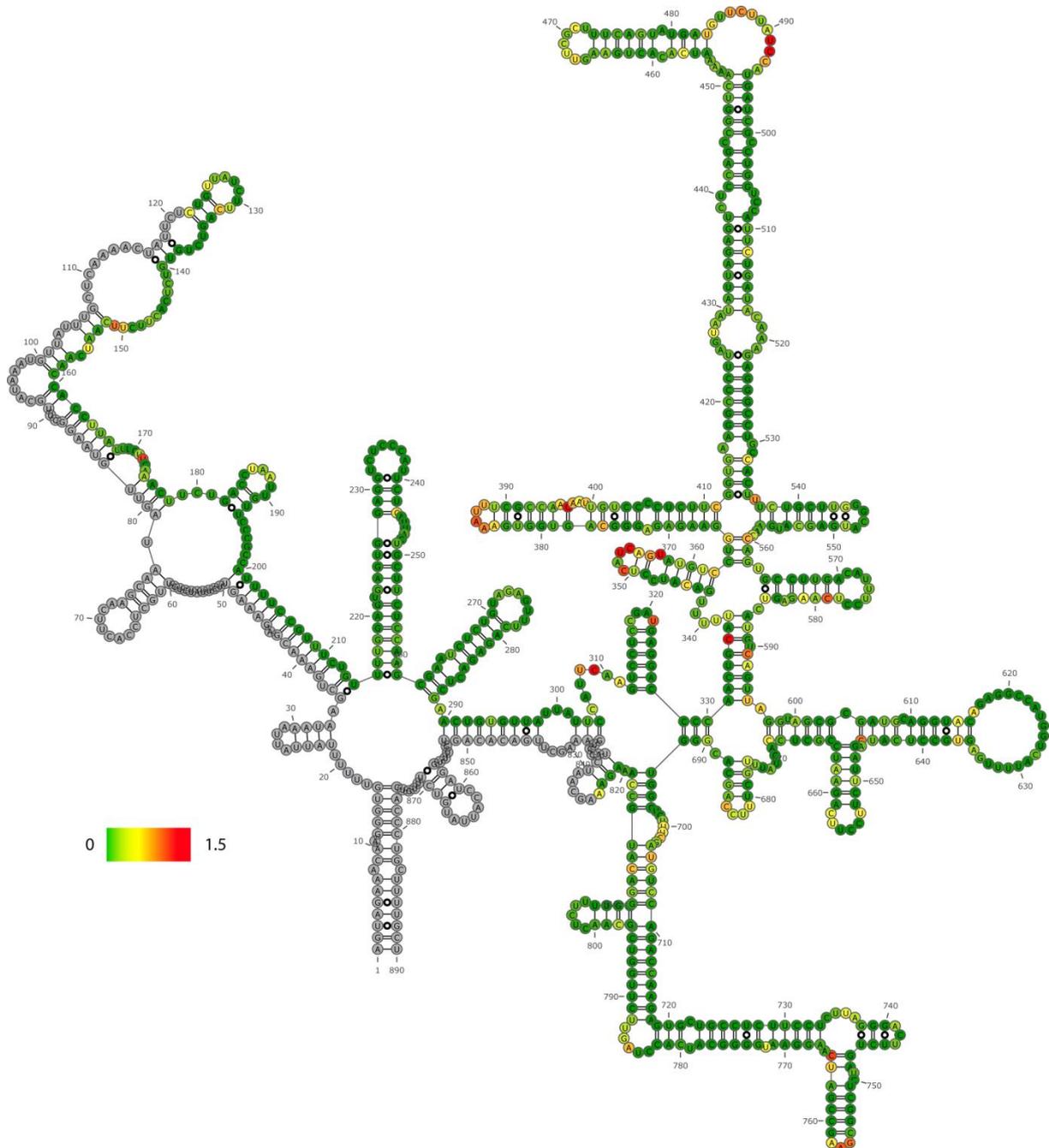
Annexe 16

La réactivité des nucléotides de l'ARNv NS de la souche WSN dans la condition NP monomérique est reportée sur la structure de l'ARN en condition NoNP



Annexe 17

La réactivité des nucléotides de l'ARNv NS de la souche WSN dans la condition NP trimérique est reportée sur la structure de l'ARN en condition NoNP



Structure de l'ARN au sein des ribonucléoprotéines des Influenzavirus A

Le génome des Influenzavirus A est constitué de huit segments d'ARN de polarité négative formant des ribonucléoprotéines virales (RNPv). La segmentation du génome complique l'empaquetage du génome, puisqu'un jeu complet de segments est nécessaire à l'infectiosité des virus. Il est aujourd'hui admis que le mécanisme d'empaquetage est sélectif et fait intervenir des interactions entre les ARN des différentes RNPv, qui dépendent vraisemblablement de la structure de l'ARN. Notre but a été de caractériser la fixation de la protéine virale NP, composant majeur des RNPv, sur l'ARN in vitro. Nous avons également mis en place une stratégie expérimentale afin de déterminer la structure de l'ARN au sein des particules virales. Ces deux aspects du projet ont été abordés par la cartographie chimique qui permet d'interroger la flexibilité de chaque nucléotide de l'ARN. Nos résultats démontrent une activité chaperone de la protéine NP ainsi que des sites préférentiels de fixation. Notre approche in viro a démontré que l'utilisation de deux sondes chimiques permet de discriminer les interactions ARN-ARN des interactions ARN-NP.

Mots clefs : Influenza, ARN, cartographie chimique, SHAPE, activité chaperonne, structure ARN

The Influenzavirus genome comprises eight segments of single-stranded RNA of negative polarity packaged in viral ribonucleoproteins (vRNP). Genome segmentation complicates packaging as a full set of vRNPs is required needed for virus infectivity. It is now accepted that packaging is selective and involves interactions between the RNA components of vRNPs that likely depend on the RNA structure. Our goal was to characterize the binding of the viral protein NP, the major component of the vRNP, on in vitro transcribed RNA. We also developed an experimental strategy to determine the RNA structure inside the viral particles. Both parts of the project were addressed with chemical mapping experiments, that interrogates the flexibility of each nucleotide in the RNA structure. Our results show that NP possess an RNA chaperone activity and binds preferential sites. Our in viro approach demonstrate that using two chemical probes allow us to discriminate between RNA-RNA and RNA-NP interactions.

Keywords: Influenza, RNA, chemical mapping, SHAPE, chaperone activity, RNA structure