



HAL
open science

Interaction des segments d'ARN génomique du virus Influenza A avec la nucléoprotéine virale NP : impact structural et implications pour l'encapsidation du génome

Erwan Quignon

► **To cite this version:**

Erwan Quignon. Interaction des segments d'ARN génomique du virus Influenza A avec la nucléoprotéine virale NP : impact structural et implications pour l'encapsidation du génome. Génétique. Université de Strasbourg, 2022. Français. NNT : 2022STRAJ039 . tel-04047015

HAL Id: tel-04047015

<https://theses.hal.science/tel-04047015v1>

Submitted on 27 Mar 2023

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 :

Erwan QUIGNON

soutenue le : 02 septembre 2022

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

Discipline/ Spécialité : Aspects moléculaires et cellulaires de la biologie

**Interaction des segments d'ARN génomique du
virus Influenza A avec la nucléoprotéine virale
NP : impact structural et implications pour
l'encapsidation du génome**

THÈSE dirigée par :

Dr Roland Marquet

IBMC, UPR9002, Strasbourg

RAPPORTEURS :

Dr Sandie Munier

Institut Pasteur, Paris

Dr Romain Volmer

ENVT, Toulouse

Examinatrice :

Pr Maria Dimitrova

IBMP, Strasbourg

Remerciements

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

Merci beaucoup à nos collaborateurs de Grenoble, Rob, Thibaut et Amélie qui m'ont beaucoup aidé lors de leur accueil pour apprendre à produire cette fameuse protéine NP.

Je voulais remercier les (très) nombreuses personnes de l'institut que je suis allé régulièrement embêter au cours de ma thèse. Merci à tous ceux qui m'ont aidé pour la production et la purification des mes ARN et de mes protéines : Guillaume, José, Anne, ... Ou encore Isabelle pour assurer les cours de microbiologie. Merci à également à Claude et Laurianne pour la chance que vous m'avez donné de travailler sur le petit projet Python, celui-ci a été une vraie bouffée d'oxygène.

Je souhaitais remercier Roland mon directeur de thèse pour son soutien tout au cours de ma thèse et de ses nombreux conseils. Cela m'a permis d'en apprendre beaucoup au cours de ces 4 années.

Je remercie aussi JC pour son énergie et ses conseils. Désolé de te décevoir mais je n'aurais pas profité de la thèse pour boire de l'alcool!

Un grand merci au reste de l'équipe pour le soutien et l'ambiance créée. Merci à Orian, ton aide scientifique a été précieuse et ton aide (et ta motivation) pour mettre le bordel dans le laboratoire ont beaucoup participés à la bonne ambiance. Merci à Benji pour nos nombreuses discussions, toujours un bon moyen pour décompresser! Egalement merci à Cédric pour tes essais répétés de faire une carrière dans l'humour (j'attends le livre de tes meilleurs blagues). Encore merci à vous trois pour les soirées à rester tard au laboratoire avec de la musique de qualité... variable diffusée sur les enceintes. Merci Valérie pour ton aide inestimable que ce soit pour apprendre le SHAPE ou à purifier des ARN. Et aussi à chercher des explications à des tubes qui gélifient sans raison. Un merci à Anne pour ton support sans faille au cours de ma thèse. Ça a toujours été un réel plaisir de travailler avec toi. Merci Serena pour ta bonne humeur qui pouvait diffuser à travers la paroi entre nos bureaux, promis je ne goûterais pas à la pizza cordon bleu-boursin. Merci aussi à Anne-Caroline et Antoine, cela faisait beaucoup de bien d'avoir enfin de nouvelles têtes qui venaient aussi travailler sur la grippe! J'ai beaucoup apprécié travailler (et aller au bar) avec vous! Merci à Charlotte, surtout pour avoir pesté sur tes analyses, ce qui m'a permis de me lancer dans un projet annexe très intéressant. Merci également à Lisa pour les pas-si-nombreuses fois où l'on a pu

se croiser mais où tu étais de bonne humeur! Merci à vous tous, car sans vous ma thèse n'aurait clairement pas eu la même saveur.

Merci à tous ceux que j'ai pu croiser au laboratoire ces dernières années d'une façon plus temporaire : Akhil, Fathima, Suresh, Anjana, ... Nos échanges ont été enrichissants!

Un merci également à ma famille pour son soutien inconditionnel.

Enfin pour finir, je souhaiterais citer un homme politique français qui a déclaré « *Je ne prendrais pas ce chiffre pour dire que c'est un échec, ça n'a pas marché* », une phrase permettant de relativiser sur les expériences qui n'ont pas marchées au cours de ma thèse.

Table des matières

I. Introduction	1
A. Virus Influenza A - Généralités	1
a. Nomenclature	1
b. Épidémiologie	4
c. ARN viral	8
d. Structure des particules virales	8
B. Cycle viral	11
a. Entrée du virus	11
b. Import des RNPv dans le noyau	14
c. Transcription et réplication du génome	14
d. Export et assemblage	15
e. Bourgeonnement	16
C. Empaquetage	17
D. Interactions ARN-NP	19
a. Structure de la protéine NP	19
b. Interactions protéine NP et ARN viral	21
c. Structure de l'ARN viral et interactions inter-segments	21
d. Activité chaperon à ARN des protéines	24
E. Etude de la structure de l'ARN	26
II. Problématique et sujet de recherche	27
III. Matériel et méthodes	29
A. Production des ARN	29
B. Nouveau protocole de production de la protéine NP	30
a. Préparation des bactéries compétentes	30
b. Transformation des bactéries	30
c. Ancien protocole de production de la protéine NP	30
d. Production de la protéine NP	30
e. Ancien protocole de purification de la protéine NP	31
f. Nouveau protocole de purification de la protéine NP	31
g. Caractérisation de la protéine	32
C. Marquage radioactif des ARN et migration sur gel	32

D.	Modification des ARN	34
a.	Modification des ARN - Ancien protocole	34
b.	Modification des ARN - Nouveau protocole	34
c.	Modification des ARN - Mise en évidence des interactions entre les complexes ARN _v NS-NP et ARN _v M-NP	35
d.	Rétrotranscription	36
E.	Microscopie électronique	37
IV. Résultats		38
A.	Mise au point des modifications de l'ARN et optimisation de la production de la protéine NP	38
a.	Modifications de l'ARN nu	38
b.	Modifications en présence de protéine NP	46
c.	Optimisation de la production de la protéine NP	47
B.	Comparaison des réactifs NMIA avec DMS et EDC	50
C.	Cartographie des ARN NS et M	56
a.	Corrélations des données obtenues	56
b.	Activité chaperon de l'ARN de la protéine NP	60
c.	Analyse de la liaison de la NP à l'ARN	73
d.	Comparaison avec la littérature	85
D.	Mise en évidence des interactions entre les segments NS et M en présence de la protéine NP	88
E.	Microscopie électronique	94
V. Discussion, perspectives et conclusions		97
Bibliographie		101
Annexes		131

Table des figures

1.	Schéma du mécanisme de réassortiment génétique	5
2.	Variété des hôtes des virus Influenza	6
3.	Liaisons $\alpha 2, 3$ et $\alpha 2, 6$	7
4.	Organisation commune des ARNv	8
5.	Génome des virus Influenza	9
6.	Représentation d'une particule virale	10
7.	Structure RNPv et configuration $7+1$	11
8.	Cycle viral	12
9.	Structure 3D de la structure de la protéine NP	20
10.	Signaux d'emballage	23
11.	Cartes des plasmides pUC 2000 NS et pUC 2000 M	29
12.	Schéma production protéine NP	33
13.	Profils de réactivités avant et après optimisation	40
14.	Test d'hybridation des amorces	41
15.	Comparaison de l'analyse du segment NS DMS avant et après optimisation	42
16.	Réactivités obtenues avec 100 et 200 mM d'EDC	44
17.	Profils obtenus avec l'ENU	45
18.	Gel test des étapes de modification	46
19.	Profils de réactivité des tests de concentration en NMIA	48
20.	Corrélation entre les concentrations 1,65 mM et 3,3 mM de NMIA	49
21.	SDS-PAGE et DLS lors de la purification de la protéine NP	49
22.	Corrélation entre les données NMIA et DMS-EDC	50
23.	Comparaison catégories NMIA-DMS-EDC	51
24.	Corrélations entre NMIA-DMS et NMIA-EDC	52
25.	Diagrammes en violon des données NMIA, DMS, EDC et DMS + EDC avant et après normalisation	54
26.	Comparaison des structures NMIA et DMS-EDC	55
27.	Profils de réactivité du segment NS dans les différentes conditions « NoNP », « ProtK » et « Comp »	57
28.	Profils de réactivité du segment M dans les différentes conditions « NoNP », « ProtK » et « Comp »	58
29.	Corrélations des données NoNP vs ProtK pour les segments NS et M	60

30.	Prédiction de la structure du segment NS dans la condition NoNP	61
31.	Prédiction de la structure du segment NS dans la condition ProtK	62
32.	Prédiction de la structure du segment M dans la condition NoNP	63
33.	Prédiction de la structure du segment M dans la condition ProtK	64
34.	Comparaison des structures NS « NoNP » et « ProtK » avec RNAStructViz	66
35.	Changements de réactivité NS NoNP-ProtK	69
36.	Changements de réactivité NS NoNP-ProtK	70
37.	Représentation des valeurs obtenues avec Δ SHAPE pour la comparaison Δ NoNP-ProtK pour les segments NS et M	72
38.	Changements de réactivité NS NoNP-Comp	74
39.	Changements de réactivité M NoNP-Comp	75
40.	Représentation des valeurs obtenues avec Δ SHAPE pour les comparaisons Δ NoNP-Comp et Δ ProtK-Comp pour le segment NS	77
41.	Représentation des valeurs obtenues avec Δ SHAPE pour les comparaison Δ NoNP-Comp et Δ ProtK-Comp pour le segment M	78
42.	Représentation des changements de catégories et des régions identifiées à l'aide de Δ SHAPE pour le segment NS entre les conditions « ProtK » et « Comp »	82
43.	Représentation des changements de catégories et des régions identifiées à l'aide de Δ SHAPE pour le segment M entre les conditions « ProtK » et « Comp »	83
44.	Composition des régions potentiellement protégées par la protéine NP en dinucléotides .	84
45.	Corrélations des données NS et M avec les données de B. Dadonaite	85
46.	Diagramme en violons des réactivités NS et M avec les réactivités de B. Dadonaite	86
47.	Région de l'ARN NS comportant un site d'empaquetage potentiel	87
48.	Test d'interactions NS-NP et M-NP	89
49.	Changements de réactivités des complexes NS-M-NP	90
50.	Interactions potentielles entre les segments NS et M	93
51.	Images capturées par microscopie électronique des complexes NS-NP	95
52.	Images capturées par microscopie électronique des complexes NS-NP dans les surnageants et les culots	96
1.	Corrélations entre les réplicats NS.	153
2.	Corrélations entre les réplicats M.	154
3.	Représentation en arcs des structures NoNP et ProtK du segment NS	155
4.	Représentation en arcs des structures NoNP et ProtK du segment M	156
5.	Comparaison des structures NS NoNP avec celles de D. Ferhadian	157
6.	Comparaison des structures NS ProtK avec celles de D. Ferhadian	158
7.	Comparaison des structures M NoNP avec celles de D. Ferhadian	159
8.	Comparaison des structures M ProtK avec celles de D. Ferhadian	160
9.	Prédiction 3D de la structure de l'ARN NS dans la condition « NoNP »	161
10.	Prédiction 3D de la structure de l'ARN NS dans la condition « ProtK »	162
11.	Prédiction 3D de la structure de l'ARN M dans la condition « NoNP »	163

12.	Prédiction 3D de la structure de l'ARN M dans la condition « ProtK »	164
13.	Changements de réactivité NS ProtK-Comp	165
14.	Changements de réactivité M ProtK-Comp	166
15.	Représentation des changements de catégories et des régions identifiées à l'aide de deltaSHAPE pour le segment NS entre les conditions « NoNP » et « Comp »	172
16.	Représentation des changements de catégories et des régions identifiées à l'aide de deltaSHAPE pour le segment M entre les conditions « NoNP » et « Comp »	173

Abréviations

ADNc = Acide désoxyribonucléique complémentaire
AMV = Virus de la myéloblastose aviaire
ARN = Acide RiboNucléique
ARNc = ARN complémentaire
ARNm = ARN messenger
ARNv = ARN viral
ATP = Adénosine triphosphate
BSA = Albumine de sérum bovin
CPSF = *Cleavage and polyadenylation specificity factor*
CRM₁ = *Chromosomal region maintenance 1*
ddGTP = Didésoxyguanosine triphosphate
DLS = *Dynamic Light Scattering*
DMS = Sulfate de diméthyle
DMSO = Diméthylsulfoxyde
dNTP = Désoxynucléotide triphosphate
DTT = Dithiothréitol
DO = Densité Optique
DTT = Dithiothréitol
EDC = 1-éthyl-3-[3-diméthylaminopropyl]carbodiimide hydrochloride
EDTA = Acide éthylènediaminetétracétique
ENU = N-nitroso-N-éthylurée
HA = Hémagglutinine
HAT = Histone acétyltransférase
HEPES = Acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
HPAI = *Highly Pathogenic Avian Influenza*
ICTV = *International Committee on Taxonomy of Viruses*
IPTG = Isopropyl β-D-1-thiogalactopyranoside
LB = *Lysogeny Broth*
LPAI = *Low Pathogenicity Avian Influenza*
MES = Acide 2-(N-morpholino)éthanosulfonique
NA = Neuraminidase
NDSB = 3-(1-Pyridinio)-1-propanesulfonate
NCR = *Non-coding region*

NLS = *Nuclear Localization Sequence*

NMIA = *N-Methylisatoic Anhydride*

NP = *Nucléoprotéine*

NTP = *Nucléotide triphosphate*

NXF1 = *Nuclear RNA export factor 1*

RNPv = *Ribonucléoprotéine virale*

RNPc = *Ribonucléoprotéine complémentaire*

RT = *Rétrotranscription*

SHAPE = *Selective 2'-hydroxyl acylation analyzed by primer extension*

SHAPE-MaP = *Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling*

SST = *SilicoTungstate de Sodium*

UTR = *Untranslated region*

VLP = *Virus-like particle*

I | Introduction

A. Virus Influenza A - Généralités

a. Nomenclature

Les virus Influenza A sont des virus du genre *Alphainfluenzavirus*, de la famille des *Orthomyxoviridae* selon l'ICTV (*International Committee on Taxonomy of Viruses*). Ces virus rassemblent un grand nombre de souches virales, désignées selon la nomenclature suivante : A/swine/Iowa/15/30 (H₁N₁). La première lettre désigne le genre de virus Influenza, suivi par l'hôte d'origine de l'isolement s'il ne s'agit pas de l'homme, puis du lieu d'isolement, du numéro de l'isolat et de l'année. Pour les virus Influenza A, H_nN_n désigne ensuite le sous-type de la souche, selon le type de l'hémagglutinine (H) et de la neuraminidase (N) [1].

Il existe 18 sous-types de l'hémagglutinine notés H₁ à H₁₈ et 11 sous-types de la neuraminidase notés N₁ à N₁₁. Les sous-types de l'hémagglutinine H₁ à H₁₆ peuvent infecter les oiseaux, tandis que les sous-types H₁-H₃, H₅-H₇, H₉ et H₁₀ peuvent infecter l'Homme, mais seuls les sous-types H₁, H₂ et H₃ se sont répandus dans la population humaine. Les sous-types H₁-H₅ et H₉ peuvent infecter les porcs [2] tandis que les sous-types H₁₇ et H₁₈ sont spécifiques des chauves-souris [3, 4]. Les sous-types de la neuraminidase N₁ à N₉ peuvent infecter les oiseaux, les sous-types N₁, N₂ et N₆-N₉ peuvent infecter l'Homme, seuls les sous-types N₁ et N₂ circulant de façon intensive. Les sous-types N₁ et N₂ peuvent infecter les porcs [2] alors que les sous-types N₁₀ et N₁₁ sont spécifiques des chauves-souris [3, 4].

Ces virus possèdent un génome segmenté en 8 ARN (-) codant chacun pour au moins une protéine essentielle (table 1).

Les virus des genres *Betainfluenzavirus*, *Gammainfluenzavirus*, *Deltainfluenzavirus* (respectivement Influenza B, Influenza C et Influenza D), *Isavirus*, *Mykissvirus*, *Quaranjavirus*, *Sardinovirus* et *Thogotovirus* sont des genres proches des virus Influenza A, partageant plusieurs similarités [7].

Les Influenza B circulent principalement chez l'Homme mais peuvent également infecter les phoques [8]. Ils diffèrent des Influenza A par la protéine M₂ qui est remplacée par les protéines NB et BM₂. La glycoprotéine NB ne possède pas encore de rôle clairement établi et n'est ni nécessaire au cycle viral *in vivo* ni avantageuse sur le plan répliatif [9]. La protéine BM₂ sert de canal ionique, de façon analogue à la protéine M₂ [10].

ARN viral	Taille (nt)	Protéines codées	Fonctions
PB2 (1)	2341	PB2	Sous-unité de la polymérase Reconnaissance de la coiffe
PB1 (2)	2341	PB1	Sous-unité de la polymérase Synthèse de l'ARN
		PB1-N40	Entrée du virus
		PB1-F2	Activité pro-apoptotique Inhibition de la réponse immunitaire
PA (3)	2233	PA	Sous-unité de la polymérase Endonucléase
		PA-N155	Rôle inconnu
		PA-N182	Rôle inconnu
		PA-X	Répression de l'expression des gènes de l'hôte
HA (4)	1778	HA	Liaison au récepteur
NP (5)	1565	NP	Nucléoprotéine
NA (6)	1413	NA	Activité sialidase
M (7)	1027	M1	Protéine de matrice Export nucléaire Bourgeoisement
		M2	Canal ionique
		M42	Rôle inconnu Possiblement une alternative fonctionnelle à M2
NS (8)	890	NS1	Inhibition de la réponse immunitaire Répression de l'expression des gènes de l'hôte Contrôle de la synthèse et de l'épissage de l'ARN viral
		NEP/NS2	Export nucléaire Régulation de la transcription et de la réplication
		NS3	Rôle inconnu

TABEAU 1. – Tailles des segments et protéines codées par les ARN génomiques des virus Influenza A [1, 5, 6]

Les Influenza C circulent chez l'Homme, les porcs, les camélidés, les chiens, les chevaux et les bovidés [11]. Ces virus ne possèdent que 7 segments génomiques. Les segments HA et NA sont remplacés par un segment codant pour une glycoprotéine de surface HEF (*Hemagglutinin-Esterase-Fusion*) qui correspond fonctionnellement aux protéines HA et NA. Le segment 6 (M) code pour une protéine additionnelle : CM2. Celle-ci a une fonction analogue aux protéines M2 des Influenza A et BM2 des Influenza B [12, 13]. Elle jouerait également un rôle dans l'emballage et dans la réplication virale [14, 15]. Enfin la protéine PA est remplacée par la protéine P3.

Les Influenza D ont été isolés en 2011 [16]. Ils infectent principalement le bétail, le porc [17] ainsi que les ruminants, les chameaux et les chevaux [11, 18]. La transmission à l'homme serait possible comme le montre l'existence d'anticorps ciblant les Influenza D dans du sérum humain [18]. Le virus ne serait cependant pas capable de circuler et de persister au sein de la population humaine [18]. Proches des Influenza C, ils comportent 7 segments génomiques. Le jeu de protéines exprimées est similaire à celui des Influenza C [17].

Les *Isavirus* qui infectent principalement le saumon atlantique possèdent, comme les Influenza A, huit segments génomiques. Le segment 6 code pour une protéine hémagglutinine-estérase en charge de la liaison au récepteur ainsi que de la destruction de cette interaction. Le segment 5 code quant à lui une protéine de fusion utilisée lors de la fusion à la membrane. Le segment NP est plus long que le segment PA et le segment NS est plus long que le segment M [19].

Les *Mikissvirus* ont été identifiés en 2017 et infectent la truite arc-en-ciel. Comme pour les *Isavirus*, le segment NP est plus long que le segment PA et le segment NS est plus long que le segment M. Une neuraminidase potentielle est codée par le segment 5 comme pour les *Isavirus* bien que la protéine codée soit plus proche de la NA des virus Influenza A et B [20].

Les *Quaranjavirus* circulent au sein des tiques et sont la cause de fortes mortalités chez les oiseaux, mais leur impact sur la santé humaine n'a pas encore été caractérisé. Ces virus ont un génome composé de six ARN [19, 21] et n'ont qu'une seule glycoprotéine de surface GP, qui aurait une fonction analogue à l'hémagglutinine [22] bien qu'elle partage des similarités avec la glycoprotéine gp64 des baculovirus à l'image des *Thogotovirus* [23]. Les segments 1, 2 et 3 codent pour les sous-unités de la polymérase virale (respectivement PB2, PB1 et PA). La glycoprotéine GP est codée par le segment 5. Les deux derniers segments (4 et 6) codent pour des protéines qui n'ont pas encore été caractérisées.

Les *Sardinovirus* sont d'autres virus pouvant infecter le saumon atlantique. Ils possèdent aussi huit segments génomiques. Ils diffèrent des *Isavirus*, le segment 4 codant pour un homologue de la protéine de fusion mais possédant une seconde ORF (*Open Reading Frame*). Les protéines putatives des segments 7 et 8 n'ont pas de correspondance dans les bases de données du NCBI [24, 25].

Les *Thogotovirus* sont transmis aux vertébrés par les tiques. Les différentes espèces peuvent avoir six ou

sept segments. Ils n'ont qu'une seule glycoprotéine GP, codée par le segment 4, apparentée à la glycoprotéine gp64 [23], comme pour les *Quaranjavirus* et aucune protéine avec une fonction similaire à la protéine NA n'a été observée. Le segment 6 coderait pour la protéine M1 ainsi que la protéine *M-long* qui inhibe la réponse immunitaire de l'hôte [26]. Le segment 5 code pour la protéine NP. Le septième segment qui peut être observé dans l'espèce *Thogotovirus dhorienne* n'a pas de rôle connu. [7, 19].

Le travail de cette thèse a porté exclusivement sur les virus Influenza A.

b. Épidémiologie

Selon l'OMS, les virus Influenza seraient responsables de 3 à 5 millions de cas sévères et de 290 000 à 650 000 morts chaque année. Les Influenza A ne sont pas les seuls *Orthomyxoviridae* à causer des épidémies. Lors des épidémies saisonnières on retrouve notamment les virus Influenza B en co-circulation avec les Influenza A. Bien que les Influenza A représentent généralement la plus grande proportion des souches circulantes lors de ces épidémies, les Influenza B peuvent certaines années être le genre le plus représenté, comme lors de la saison 2017/2018 [27]. Cependant les Influenza B n'ont pas de potentiel pandémique, contrairement aux Influenza A [28].

L'évolution des Influenza A est dirigée par deux grands mécanismes : la dérive antigénique et les cassures génétiques.

La dérive antigénique est due aux erreurs de la polymérase virale. Celle-ci ne dispose pas d'activité de relecture et il est estimé qu'elle introduit environ une erreur par génome répliqué; chaque cellule pourrait donc produire 10 000 nouveaux mutants [29]. Cela permet l'échappement au système immunitaire par la génération de nouveaux variants et peut également rendre la vaccination inefficace bien que ce soient les mêmes souches qui circulent. Ce mécanisme rend nécessaire la mise à jour du contenu du vaccin chaque année [30]. De plus, la dérive antigénique mène à l'apparition de résistances aux antiviraux, tels que l'amantadine et l'oseltamivir [29].

La cassure antigénique est un autre mécanisme, qui se produit lors d'une co-infection d'une cellule par deux souches différentes. Lors de cette co-infection, il peut se produire un réassortiment génétique (fig. 1), qui consiste en un échange de segments génomiques entre plusieurs souches [31]. Il en résulte l'apparition de virus qui, s'ils possèdent de nouveaux gènes HA et NA contre lesquels l'immunité est limitée, vont avoir une transmission plus importante et ainsi être à l'origine d'une pandémie [32]. Les virus Influenza A ont ainsi été la cause de plusieurs pandémies majeures depuis le début du XX^{ème} siècle (table 2).

Les travaux de ces dernières années ont mis en évidence la grande diversité d'hôtes potentiels des virus Influenza A et la complexité des transmissions inter-espèce [33, 34]. Les oiseaux sauvages servent de réservoir naturel aux virus Influenza A d'où un grand nombre de transmissions inter-espèces peut survenir (fig. 2). Ces transmissions inter-espèces posent des problèmes de santé publique car lors de l'infection d'un hôte par deux virus Influenza d'origines différentes, le mécanisme de réassortiment génomique peut générer de nouveaux virus potentiellement hautement pathogènes. L'adaptation à d'autres espèces peut ensuite se produire suite à la dérive antigénique. Un risque important est posé notamment par le porc :

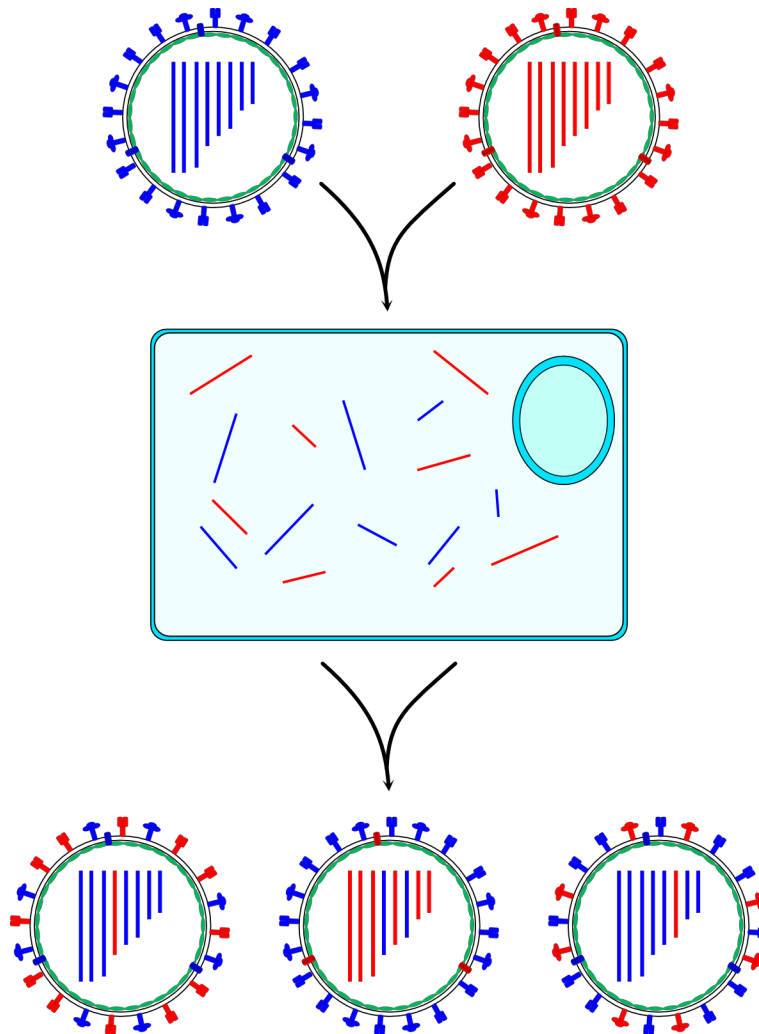


FIGURE 1. – Représentation schématique du mécanisme de réassortiment génétique. Deux virus de souches différentes, représentés par deux couleurs, co-infectent une même cellule, menant à la génération de virus réassortants qui partagent une part plus ou moins grande de similarités avec l'un des virus parentaux.

celui-ci peut être infecté aussi bien par des virus aviaires que par des virus humains; il peut donc s'y créer, par le biais de la dérive et de la cassure antigénique, des virus pouvant infecter l'homme mais possédant des glycoprotéines aviaires, inconnues du système immunitaire et pouvant déclencher des pandémies.

L'adaptation d'un virus aviaire à l'Homme est souvent vue comme l'un des risques principaux en santé publique, les pandémies grippales humaines ayant une origine aviaire [35]. Cette adaptation va dépendre principalement de deux facteurs : la capacité du virus à se lier aux récepteurs humains et sa capacité à se répliquer dans les cellules de l'hôte.

La protéine intervenant principalement dans la reconnaissance des récepteurs est l'hémagglutinine, qui forme un homotrimer possédant un site de liaison reconnaissant spécifiquement les acides sialiques. Les Influenza A humains reconnaissent de manière préférentielle les liaisons $\alpha 2, 6$ alors que les souches aviaires et des équidés ciblent les liaisons $\alpha 2, 3$ [36] (fig. 3). On retrouve chez l'Homme des liaisons $\alpha 2, 6$ au niveau de l'épithélium des voies respiratoires supérieures [37] ainsi que des acides sialiques avec une

Années	Sous-type	Estimation du nombre de décès
1918–1919	H ₁ N ₁	50 000 000
1957–1958	H ₂ N ₂	1 500 000
1968–1969	H ₃ N ₂	1 000 000
2009–2010	H ₁ N ₁	300 000

TABLEAU 2. – Pandémies grippales depuis le début du XX^{ème} siècle

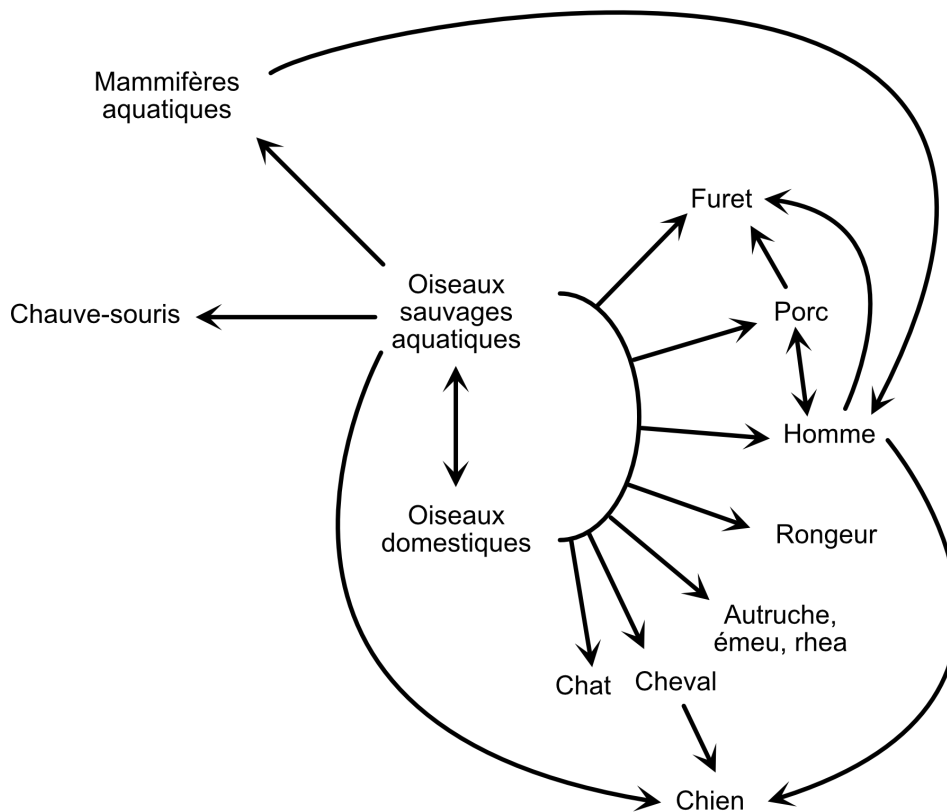


FIGURE 2. – Spectre des hôtes connus pour les virus Influenza A [33, 34], les flèches représentent les transmissions possibles.

liaison $\alpha 2, 3$ au sein de l'ensemble du tractus respiratoire humain [38]. Cependant, les liaisons $\alpha 2, 6$ sont présentes sur des cellules non ciliées et les liaisons $\alpha 2, 3$ sur des cellules ciliées qui ne permettent pas, pour ces dernières, une répllication optimale du virus; l'infection chez l'Homme par des virus aviaires est donc limitée. Le changement de tropisme de l'hémagglutinine est donc nécessaire pour obtenir une transmission efficace d'homme à homme [35]. Les virus d'origine porcine possèdent eux la capacité de se lier aussi bien aux liaisons $\alpha 2, 3$ que $\alpha 2, 6$ [39].

La capacité à se répliquer dans les cellules d'un hôte différent va également avoir un rôle important dans la capacité du virus à franchir les barrières d'espèces. A titre d'exemple, la substitution d'un seul acide aminé ($627\text{Glu} \rightarrow 627\text{Lys}$) de la sous-unité PB2 de l'ARN polymérase suffit à un virus aviaire pour se répliquer en cellules de mammifères [40]. Cette substitution permet l'adaptation du virus à la protéine ANP32A, qui est plus courte chez les mammifères, la protéine ANP32A aviaire possédant 33 acides aminés

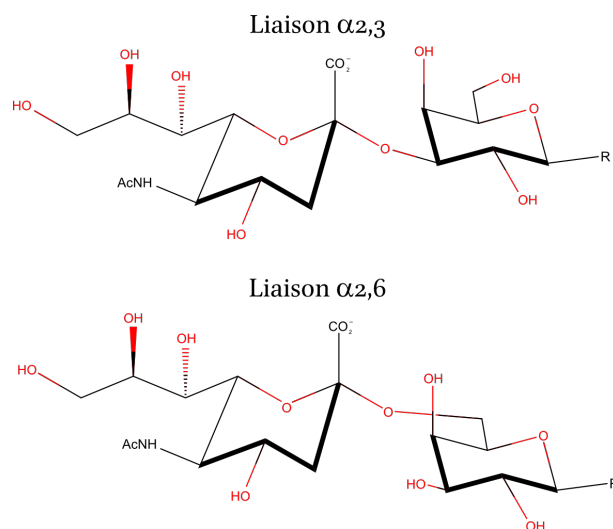


FIGURE 3. – Représentation des liaisons α 2, 3 et α 2, 6.

supplémentaires provenant d'un exon absent chez les mammifères et les autruches [41], *via* une interaction directe avec le domaine C-ter de la sous-unité PB2 de l'ARN polymérase virale, dont la glutamine 627 fait partie, par des interactions hydrophobes et électrostatiques [42, 43].

Les différents virus Influenza A aviaires sont classifiés comme hautement (HPAI : *Highly Pathogenic Avian Influenza*) ou faiblement pathogènes (LPAI : *Low Pathogenicity Avian Influenza*) selon la sensibilité de l'hémagglutinine aux protéases. Celle-ci nécessite un clivage au niveau d'une boucle du précurseur HA₀ qui va définir la sensibilité de la souche aux différentes enzymes de l'hôte. Les virus faiblement pathogènes possèdent un site monobasique pour le clivage alors que les virus hautement pathogènes résultent d'une insertion d'un motif polybasique qui étend la boucle de clivage, rendant celui-ci plus sensible à des protéases exprimées de façon ubiquitaire [44]. Ces sites polybasiques ne sont observés qu'au sein des sous-types H5 et H7. Il a cependant été montré que par modifications génétiques, il était possible d'obtenir des sous-types H2, H4, H8 et H14 comportant un site multibasique, qui exhibent un phénotype léthal. [44, 45].

Interactions avec d'autres pathogènes

Les co-infections entre les virus Influenza et des pathogènes bactériens ou viraux peuvent être fréquentes et représentent un risque supplémentaire pour les personnes atteintes.

Les co-infections bactériennes sont l'un des principaux facteurs de mortalité suite à une infection par un virus Influenza A. Cette infection va diminuer les défenses immunitaires ciblant les bactéries, créant un environnement favorable à leur prolifération dû à la réponse interféron de type I et une déplétion des macrophages alvéolaires [46]. En outre, les cellules infectées par le virus vont exprimer plus de récepteurs et de sites de liaison reconnus par les bactéries et des interactions directes se produiraient entre le virus et certaines espèces de bactéries favorisant l'adhérence des bactéries [47]. On peut retrouver différentes bactéries dans ces co-infections, telles que par exemple *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Pseudomonas aeruginosa* et *Haemophilus influenzae* [46].

Les interactions des virus Influenza avec d'autres virus ont été étudiées afin de chercher si des co-infections pouvaient avoir des pronostics plus sévères ou s'il y avait des risques d'interactions entre les différents virus. A cause de la pandémie de SARS-Cov2, plusieurs études ont examiné les co-infections entre ces deux virus respiratoires. Certaines méta-analyses pointent vers un effet aggravant sur la mortalité en cas de co-infection avec le SARS-Cov2 [48], tandis que d'autres ne montrent pas d'effet [49]. Une étude *in vitro* suggère que des co-infections entre le virus Influenza A et le virus respiratoire syncytial peuvent mener à la génération de particules virales hybrides, portant les glycoprotéines des deux virus, ce qui leur permet d'échapper au système immunitaire et d'infecter des cellules dépourvues des récepteurs reconnus par les Influenza virus [50].

c. ARN viral

Les ARNv des virus Influenza A sont des ARN simples brins de polarité négative qui partagent une organisation similaire : une région codante flanquée d'une région non-codante à chaque extrémité. Aux extrémités 5' et 3' des régions non-codantes se trouvent les régions Uni13 (U13) et Uni12 (U12) de 13 et 12 nucléotides, communes à tous les segments et aux genres *Alphainfluenzavirus*, *Betainfluenzavirus*, *Gammainfluenzavirus* [51–53] (fig. 4).

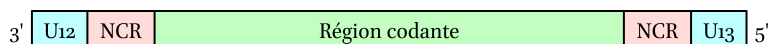


FIGURE 4. – Représentation de l'organisation partagée par les ARNv des virus Influenza. La région codante, en orientation antisens, est flanquée de régions non codantes (NCR) et les régions U12 et U13 sont présentes aux extrémités 3' et 5' respectivement.

Les régions codantes codent pour, selon les segments, une ou plusieurs protéines (fig. 5). Pour la production des protéines virales accessoires, les mécanismes de *leaky scanning*, de changement de cadre de lecture et de ré-initiation sont utilisés par le virus [54].

Les régions non-codantes, à l'exception des régions U12 et U13, sont des régions non-conservées de longueurs variables, spécifiques à chaque segment [55]. Au sein de la région non-codante 5' de tous les ARN génomiques se trouve une région oligo(U) généralement longue de 5 à 6 nucléotides[56]. La variabilité des séquences de ces régions non-codantes dépend des segments, les segments NA et HA possédant la plus grande variation [55].

Les régions U12 et U13, complémentaires, forment une région partiellement double brin. Cette structure secondaire sert de promoteur au complexe polymérase [51, 57–59].

d. Structure des particules virales

Les études menées sur la structure des virions des virus Influenza A ont montré l'existence de particules sphériques et filamenteuses [60–62]. Bien que les deux formes puissent être observées, l'essentiel des

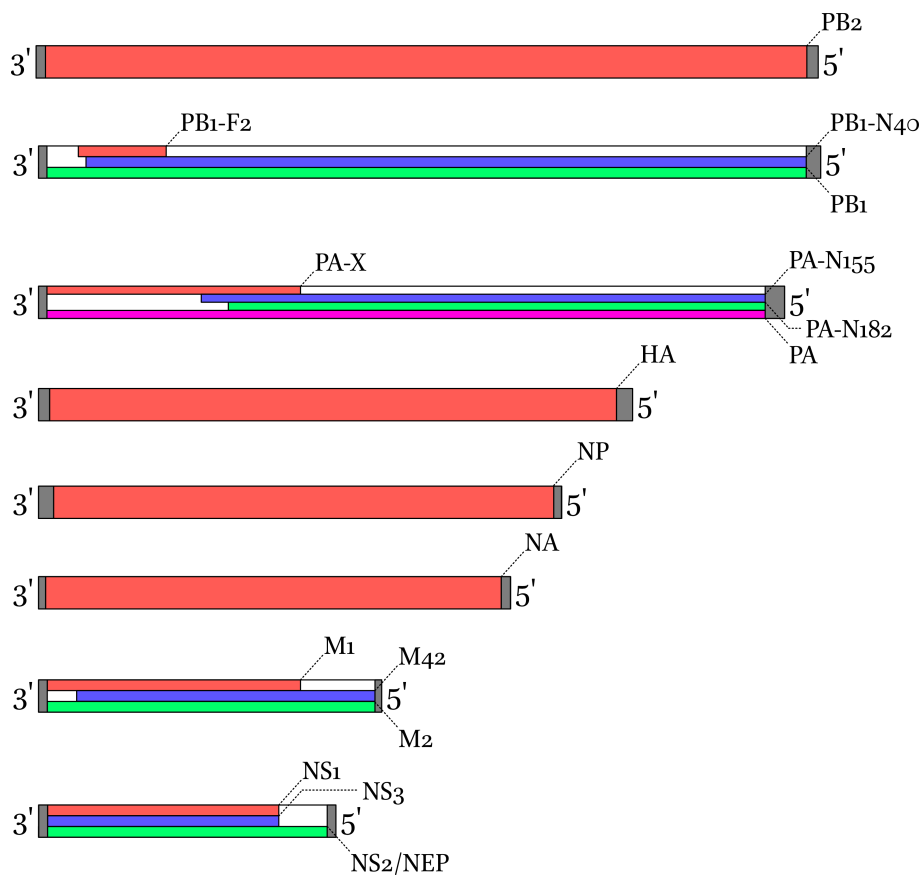


FIGURE 5. – Représentation des ARN génomiques des virus Influenza A. Les ARN sont représentés dans le sens 3' → 5' et les couleurs représentent les différentes régions codantes pour les protéines virales.

recherches a abordé les formes sphériques [63–65], majoritairement observées en laboratoire. Les formes filamenteuses sont plus fréquemment observées *in vivo* [66] ou sur des souches virales qui n'ont ni subi un grand nombre de passages en laboratoire ni été adaptées à des souris [67]. Les études menées par la suite ont montré que le type de cellule infectée impacte la forme des virions, les cellules polarisées (comme celles de l'épithélium respiratoire, ciblé par les virus Influenza) favorisant la formation des virions filamenteux [68]. La forme filamenteuse serait également plus efficace lors de la transmission dans un milieu riche en mucus; ce qui expliquerait que cette forme, qui pourrait améliorer la transmission du virus [69], est principalement retrouvée dans les isolats cliniques [70]. Un facteur important pour l'orientation vers une forme filamenteuse ou sphérique est la génétique du virus, les gènes et leurs origines jouant un rôle dans l'apparition des formes filamenteuses [71]. Ainsi des mutations dans le gène de la protéine M1 ont un rôle important dans l'apparition des formes filamenteuses [67, 72–79]. Ces mutations affecteraient l'interaction entre M1 et NP, changeant la forme de la particule virale [77]. L'efficacité de la liaison de M1 aux RNPv impacte également la morphogenèse des virions [80]. Les souches virales filamenteuses auraient également une activité neuraminidase plus importante [81].

Les particules virales des virus Influenza ont été étudiées notamment par tomographie électronique, montrant que les différentes formes de particules virales partageaient la même organisation [82]. Celles-ci sont composées d'une enveloppe lipidique sur laquelle on retrouve les glycoprotéines hémagglutinine sous forme de trimères et neuraminidase sous forme de tétramère. La protéine M2 est insérée dans

l'enveloppe [83] (fig. 6). La densité en glycoprotéines est identique entre les formes sphériques et filamenteuses. On retrouve environ 340 à 400 trimères d'hémagglutinine par virion, aussi bien dans les formes sphériques que filamenteuses [82, 84].

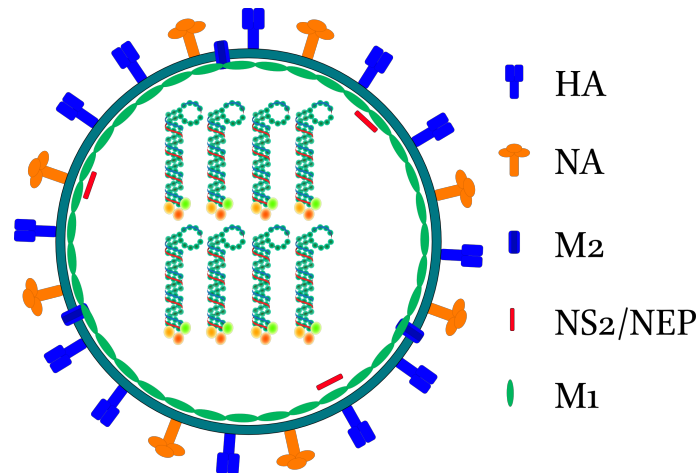


FIGURE 6. – Représentation d'une particule virale. Au centre se trouvent les 8 ribonucléoprotéines virales. Les glycoprotéines HA et NA sont ancrées dans l'enveloppe. La protéine M2 est une protéine transmembranaire.

Le ratio entre les glycoprotéines HA et NA varie fortement entre les différents sous-types mais également entre différentes souches virales puisque l'on observe pour les souches H₃N₂ A/Wisconsin/15/2009 et A/Victoria/210/2009 des ratios HA/NA de 6 :1 et 12 :1 respectivement et un ratio de 3 :1 pour la souche A/California/7/2009 (H₁N₁). Le ratio peut également fortement varier entre les genres de virus, la souche virale Influenza B B/Brisbane/60/2008 ayant un ratio HA/NA de 18 :1 [85]. La protéine HA est cependant dans tous les cas plus abondante que NA. Les particules virales sont polarisées et la répartition des glycoprotéines n'est pas aléatoire. On retrouve les tétramères de neuraminidase en amas, principalement concentrés à une extrémité de la particule virale, tandis que la protéine HA est répartie d'une manière plus uniforme sur le reste de la particule virale [69, 82, 86, 87]. A l'autre extrémité du pôle concentré en NA, on retrouve les RNPv liées au complexe polymérase (composé des protéines PB₂, PB₁ et PA) en interaction avec la protéine de matrice M₁ [65, 86, 88].

L'ARN est enroulé autour d'un squelette formé par la protéine NP [89, 90] et le complexe réplicase (composé des protéines PB₂, PB₁ et PA) interagit avec les extrémités de l'ARN, formant les ribonucléoprotéines virales (RNPv) [91, 92]. Au sein des particules virales on retrouve ces RNPv organisées sous la forme d'un complexe supramoléculaire [93, 94]. Les ribonucléoprotéines adoptent une structure dite en poêle à frire (fig. 7A), les deux extrémités des ARN sont en interaction l'une avec l'autre puis on retrouve une hélice avec une boucle terminale, située au milieu du segment [53]. Les bases de l'ARN sont exposées par la protéine NP, les laissant accessible au milieu environnant [95, 96]. Les RNPv ont une taille pouvant varier entre 30 et 120 nm suivant la taille de l'ARNv [89, 97].

Les RNPv sont réparties dans une configuration dite « 7+1 » avec une RNPv centrale et les 7 autres RNPv

réparties en périphérie [93, 98] (fig. 7B). Cette configuration est très conservée au sein des Influenza, étant retrouvée aussi bien chez les Influenza A que les Influenza B. Une configuration « 7+1 » similaire est retrouvée chez les Influenza C et D bien que ceux-ci n'aient que sept ARNv [99]. L'absence d'un ou plusieurs segments entraîne l'organisation des ARN viraux sous une configuration de l'ordre de « x+1 » avec des configurations allant jusqu'à « 4+1 » [100]. Un seul exemplaire de chaque ARN est empaqueté dans les particules de virus Influenza A [101, 102]. L'arrangement des RNPv dans la particule virale peut varier au sein d'une même souche, mais le segment central pourrait avoir un rôle structural [103]. On retrouve également une variabilité de l'arrangement des complexes entre différentes souches virales [104]. Cette configuration « 7+1 » serait importante pour les virus Influenza A et ceux-ci pourraient empaqueter des ARN non viraux pour conserver cette configuration [105].

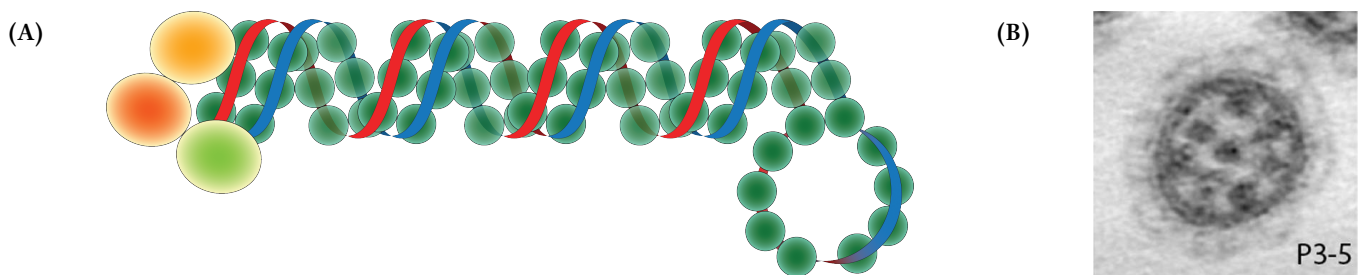


FIGURE 7. – (A) Représentation schématique du modèle dit en poêle à frire des RNPv avec le complexe réplicase hétérotrimérique en interaction avec le complexe ARN-NP. (B) Virion observé en tomographie électronique. Les points plus sombres sont les différentes RNPv virales dans une configuration « 7+1 » avec une RNPv centrale et 7 RNPv réparties en périphérie [94].

B. Cycle viral

a. Entrée du virus

Les glycoprotéines virales HA et NA interviennent aux différentes étapes de l'entrée du virus. Ces deux protéines jouent des rôles antagonistes, dont l'équilibre est important pour la transmission du virus. La protéine HA se lie aux acides sialiques liés par des liaisons covalentes aux glycoprotéines et aux glycolipides présents à la surface des cellules tandis que la protéine NA clive cette liaison covalente.

Lors de l'infection par un virus Influenza, la première barrière rencontrée par le virus est le mucus. Celui-ci est majoritairement composé d'acides sialiques, auxquels la protéine HA peut se fixer. Des virus qui vont s'attacher trop fortement au mucus passeront la barrière trop lentement et seront donc éliminés avant l'infection des cellules. La balance entre les activités de NA et HA joue donc un rôle important dans la capacité du virus à traverser le mucus. Des souches ayant une faible activité de la protéine NA possèdent une transmission plus faible entre furets du fait de la présence de mucus [106]. L'infection virale impacte également la production de mucus ; plus l'infection s'établit tôt, plus rapidement la production de mucus est impactée. Ainsi, une souche moins efficace pour traverser le mucus mettra aussi plus de temps à inhiber la production de ce mucus par les cellules infectées [106]. L'activité de la protéine NA est donc importante

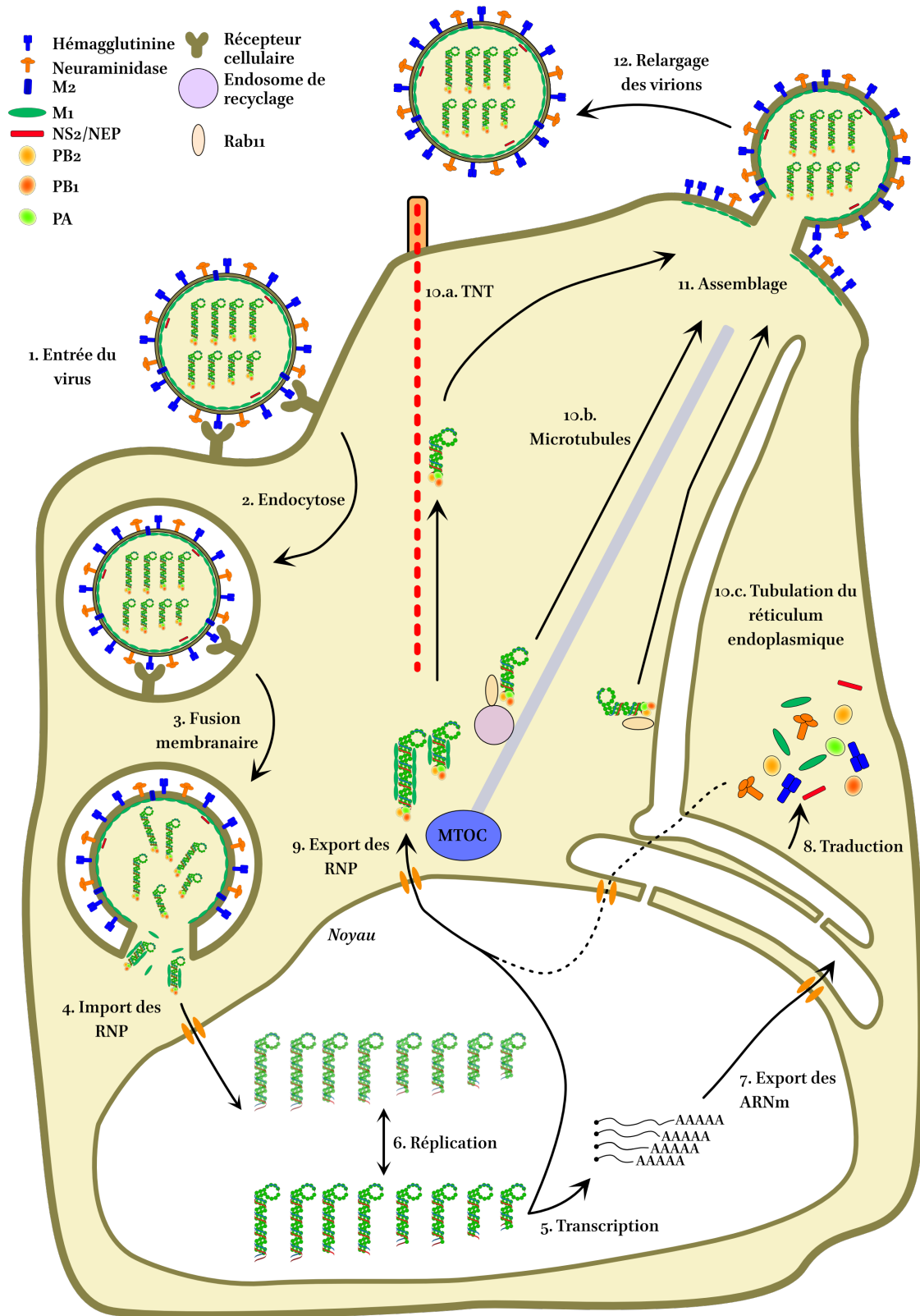


FIGURE 8. – Représentation schématique du cycle viral des Influenza A.

pour les premières étapes de l'infection, afin de lutter contre les effets inhibiteurs du mucus.

Toutefois, l'impact du mucus sur la transmission virale varie entre espèces. Le mucus humain a un effet inhibiteur sur celle-ci, tandis que le mucus des porcs n'a pas cet effet. Une souche porcine peut donc avoir des difficultés à se transmettre à l'Homme [106]. La souche H₁N₁ pdm2009, cause de la pandémie de 2009, est issue de l'acquisition d'une HA avec une avidité plus faible et d'une NA avec une activité faible, acquérant ainsi une balance HA/NA plus compatible avec l'Homme que le virus progéniteur porcine [107].

La polarisation de la neuraminidase permet aux particules virales de se diffuser dans une direction constante opposée au pôle concentré en neuraminidase, en clivant la liaison aux acides sialiques sur la trajectoire [69]. Le clivage des acides sialiques permettrait aux virus Influenza de ne pas repartir en arrière [108].

La protéine HA s'attache aux acides sialiques ou à des glycanes phosphorylés [109]. L'hémagglutinine des virus Influenza de chauves-souris ne s'attache pas aux récepteurs canoniques mais se lie au complexe majeur d'histocompatibilité de classe II [110]. Après fixation sur la cellule, les activités des protéines HA et NA vont permettre, par des cycles de fixation et de clivage de la liaison aux acides sialiques, au virus de « rouler » sur la surface cellulaire afin de trouver les récepteurs cellulaires utilisés pour l'entrée du virus [111, 112] (fig. 8 - 1).

Après fixation sur les récepteurs appropriés pour l'internalisation du virus, l'endocytose est déclenchée (fig. 8 - 2). Plusieurs récepteurs ont été identifiés comme possédant un rôle dans l'entrée du virus comme par exemple la nucléoline, l'EGFR (*Epidermal Growth Factor Receptor*) et CaV_{1.2}, bien que les récepteurs utilisés changent selon les sous-types et que selon les connaissances actuelles, il n'existe pas de récepteur universel pour l'internalisation des particules virales [113]. Ce mécanisme peut être dépendant [114–116] ou indépendant de la clathrine [117, 118], ou s'effectuer par macropinocytose [5, 119].

Une fois les particules virales internalisées au sein des endosomes précoces, ceux-ci commencent leur maturation en étant transportés vers le noyau *via* les microtubules. Pendant la maturation, le pH bas entraîne l'ouverture des canaux ioniques formés par la protéine M₂, ce qui fait entrer des ions potassiums et des protons et provoque l'acidification de la particule virale. A un pH de 6,5, un changement conformationnel de la protéine M₁ permet d'affaiblir les liaisons entre M₁ et les RNPv [120–122]. Quand le pH descend sous 6,0, de nouveaux changements structuraux ont lieu au sein de M₁ et entraînent une perte de stabilité du complexe supramoléculaire formé par les RNPv [122]. Le peptide de fusion est exposé suite au clivage de HA₀ par des protéases permettant la libération des domaines HA₁ et HA₂, lors de la formation des virions [123–125]. Ce peptide, situé en N-ter de HA₂, formé par des acides aminés hydrophobes, s'insère dans la membrane endosomale, ce qui rend le changement conformationnel définitif [126] et permet le rapprochement des deux membranes qui fusionnent ensuite [5] (fig. 8 - 3). Des chaînes d'ubiquitine libres sont exposées, ainsi que la protéine M₁ qui sont toutes deux liées par l'histone désacétylase 6 (HDAC6). Cela permet l'attachement de la couche formée par la protéine M₁ à la dynéine et

à la myosine II et entraîne l'ouverture de la couche de M1 [127]. Les protéines M1 dont la conformation a changé suite à la modification de pH sont reconnues et liées par TNPO1 (importine- β 2), qui permet de retirer la protéine M1 des RNPv et de séparer le complexe supramoléculaire avant l'entrée dans le noyau [128] (fig. 8 - 4).

b. Import des RNPv dans le noyau

Les RNPv sont importées dans le noyau, indépendamment de la protéine M1, par les pores nucléaires [129, 130]. L'ARN viral ne pouvant être importé seul, cet import est réalisé *via* un NLS présent sur les protéines NP, bien que toutes les protéines virales des RNPv des virus Influenza possèdent des NLS [131, 132] (PB2 [133], PB1 [134, 135] et PA [136, 137]). Parmi les NLS présents au sein de la protéine NP, c'est le NLS non conventionnel à l'extrémité N-ter qui est utilisé [132]. Les RNPv se lient à l'importine α , qui sera reconnue par l'importine β [5, 131, 138], permettant l'import dans le noyau. Il existe différentes isoformes d'importine α ; les virus Influenza aviaires reconnaissent les importine α_3 , tandis que les virus Influenza de mammifères ciblent les importine α_7 [139]. Le passage d'un tropisme aviaire à humain requiert une adaptation pour l'import des RNPv. Des isolats de la pandémie de 2009 ont montré une capacité à se lier aux deux types d'importine α , illustrant cette adaptation [139]. Les virus Influenza A semblent pouvoir en majorité reconnaître l'importine α_1 et les importines α_4 et α_5 sont redondantes [139]; cependant l'expression des différentes importine α n'est pas identique au sein des différentes cellules [140] et leurs affinités pour les différentes protéines virales varient [141].

c. Transcription et réplication du génome

Deux mécanismes ont lieu au sein du noyau : la transcription des ARNm (ARN messagers) viraux (fig. 8 - 5) afin de produire les protéines virales et la réplication du génome viral afin de produire de nouvelles RNPv (fig. 8 - 6).

Les ARNm viraux sont transcrits à l'aide du mécanisme de *cap-snatching* [142–144]. La sous-unité PB2 de la polymérase virale se lie aux extrémités 5' des ARNm de l'hôte en cours de synthèse et la sous-unité PA clive l'ARNm dix à treize nucléotides en aval de la coiffe [145, 146]. Une rotation de la sous-unité PB2 permet d'amener l'ARNm cellulaire à l'endonucléase PA puis d'amener l'amorce au site catalytique polymérase porté par la sous-unité PB1 [59, 147, 148]. L'oligonucléotide coiffé sert d'amorce pour la synthèse de l'ARNm viral, qui utilise l'ARNv comme matrice [148–150]. La sous-unité PB1 de l'ARN polymérase virale interagit avec la forme hyperphosphorylée du domaine C-ter de l'ARN polymérase II [151, 152]. L'extrémité 3' de la matrice est séquestrée par la polymérase, le protégeant des exonucléases cellulaires [153]. Les protéines NP liées à la matrice en cours de transcription sont retirées du brin entrant dans la polymérase pour être transférées au brin sortant, afin d'éviter que l'ARN ne soit exposé sans NP au milieu environnant [153]. La séquence oligoU à l'extrémité 5' de l'ARNv permet la polyadénylation des ARNm par un mécanisme de bégayement avant le décrochage de la polymérase [5, 154]. La transcription des ARNm viraux ne nécessite pas l'intervention de protéines virales nouvellement synthétisées, les RNPv seules suffisant à la transcription et à la réplication [155, 156].

L'export des ARNm viraux (fig. 8 - 7) nécessite l'interaction avec l'ARN polymérase II [157]. Les premiers transcrits seraient exportés de façon indépendante des voies NXF1 et CRM1 [158], cependant les transcrits plus tardifs sont exportés *via* la voie NXF1 [158], à l'exception des ARN PA, PB1 et PB2 qui seraient exportés par une voie indépendante de NXF1 et CRM1 [159]. La protéine NS1 se lie au complexe CPSE, ce qui inhibe le clivage et la polyadénylation des ARNm cellulaires et donc leur transport [160, 161].

Le complexe polymérase réplique l'ARN viral en synthétisant l'ARN complémentaire (ARNc), de polarité positive, non coiffé et non polyadénylé, qui est ensuite répliqué en copies d'ARNv. Les ARNc synthétisés s'associent à la protéine NP ainsi qu'à une copie de la polymérase virale pour former des RNPc [5]. La protéine ANP32A permet le recrutement de la NP sur l'ARN en cours de synthèse *via* une région LCAR (*Low-Complexity Acidic Region*) [162]. Le modèle proposé pour la transcription impliquant le transfert de la NP entre les brins entrant et sortant serait également applicable à la répllication [153].

Le complexe polymérase est capable de réaliser aussi bien la transcription des ARNm que la répllication [163], l'équilibre entre les deux mécanismes serait régulé par la protéine virale NS2/NEP [164] bien que ce mécanisme soit mal connu.

d. Export et assemblage

La plupart des protéines virales disposant de NLS, elles peuvent être importées dans le noyau par la voie des importines α et β [165]. Les protéines PB1 et PA sont importées dans le noyau sous la forme d'hétérodimères par RanBP5 [166–168] tandis que NP et PB2 sont importées individuellement [168, 169]. La protéine NS2/NEP entre dans le noyau par diffusion passive [170] (fig. 8 - 8).

Afin d'exporter les RNPv nouvellement produites (fig. 8 - 9), le modèle supposé est que la protéine virale NS2/NEP se lie aux RNPv *via* la protéine M1 qui se lie aux protéines NP composant les RNPv [171–173]. La protéine NS2/NEP permet ainsi l'export *via* la voie CRM1. Une fois dans le cytoplasme, la présence de la protéine de matrice M1 empêche l'exposition des NLS de la protéine NP, ce qui évite que les complexes soient transportés à nouveau dans le noyau [5].

Dans le cytoplasme, les RNPv s'accumulent à proximité du MTOC (*Microtubule Organizing Center*) [174] avant d'être transportées à la membrane cellulaire. La protéine M1 est dissociée des RNPv lors du transport vers la membrane cellulaire [175]. La sous-unité PB2 interagit avec Rab11, ce qui permet le transport des RNPv vers la membrane cellulaire [176–178]. Cette interaction est importante pour l'emballage du virus, l'absence de Rab11 perturbant complètement ce processus et augmentant le nombre de particules virales défectives produites [179].

Plusieurs modèles du transport des RNPv *via* Rab11 ont été proposés. Un premier modèle suggère l'utilisation des TNT (*tunneling nanotubes*) comme voie de transport [180] (fig. 8 - 10.a). Ces derniers sont des tunnels d'échange entre les cellules, utilisés par le virus afin de se propager de cellule à cellule

[181, 182]. La protéine Rab11 pourrait utiliser ces connexions cellulaires afin de transporter les RNPv à la membrane cellulaire. Un second modèle propose que les RNPv se lient à Rab11 sur des endosomes de recyclage *via* le réseau de microtubules [178] (fig. 8 - 10.b). Enfin un troisième modèle suggère que l'infection cause des tubulations du réticulum endoplasmique (fig. 8 - 10.c) et que les RNPv se lient aux Rab11 qui y sont localisées [183]. Ces trois modèles ne sont pas exclusifs et une redondance pourrait exister, ce qui expliquerait que la déstabilisation des réseaux de microtubules puisse selon les cas ne pas avoir d'effet [184] ou impacter la production de particules virales [178]. La proximité des différentes RNPv au cours de leur transport vers la membrane pourrait permettre de faciliter l'établissement des liaisons entre les RNPv et former le complexe supramoléculaire qui sera empaqueté dans les particules virales [185].

Les protéines se fixant dans la membrane de la particule virale (NA, HA et M2) sont traduites au niveau du réticulum endoplasmique. Les protéines NA et HA oligomérisent afin de former des tétramères et des trimères [5]. Après transfert à l'appareil de Golgi, si la HA possède un site de clivage multibasique, elle peut être clivée par la furine, une enzyme exprimée de façon ubiquitaire et au sein de la face *trans* de l'appareil de Golgi [186, 187]. Si la HA possède un site de clivage monobasique, elle nécessitera un clivage extracellulaire par d'autres enzymes (HAT, Tmprss2 ou Tmprss4) qui vont avoir une expression limitée à certains types cellulaires [5].

e. Bourgeonnement

Le bourgeonnement a lieu au niveau des radeaux lipidiques de la membrane plasmique, des domaines riches en cholestérol et en sphingolipides [188]. Les protéines HA, NA et M2 y sont concentrées *via* différents mécanismes. La protéine M1 cytosolique se lie aux queues cytoplasmiques des protéines HA et NA [189, 190].

Les RNPv se décrochent à proximité de la membrane de Rab11, cette protéine n'étant pas retrouvée au sein des particules virales [191]. L'interaction avec la queue cytoplasmique de la protéine M2 permet le recrutement des RNPv et des autres protéines virales [192–194] en interagissant avec la protéine M1 (fig. 8 - 11).

La courbure de la membrane plasmique va être réalisée par plusieurs des protéines virales, mais ce mécanisme est complexe et est encore mal connu. Chacune des protéines virales pourraient induire seule la formation et la libération de VLP (*Virus-like Particles*) : M1 [195] (d'autres études ont montré l'incapacité de M1 à induire seule des VLP [196]), M2 [197], NA [197–199]. Les protéines virales essentielles à la formation de la courbure de la membrane ne sont pas encore déterminées. La protéine M1 a été désignée comme essentielle [200], mais des études plus récentes ont montré la formation de VLP en présence uniquement des protéines HA et NA [197]. Le couple HA/NA n'était d'ailleurs pas considéré comme essentiel selon certaines études [201, 202].

Une fois la particule virale formée, la scission est réalisée par la protéine M2 [203], puis la protéine NA

hydrolyse la liaison glycosidique qui lie les acides sialiques aux sucres. Cela empêche le virus de rester accroché à la cellule productrice (fig. 8 - 12).

C. Empaquetage

La structure du génome des virus Influenza A, séparé en huit segments génomiques, apporte un avantage évolutif au virus en permettant l'échange de matériel génétique entre différentes souches virales [204]. Cependant cela pose plusieurs défis pour l'obtention d'une particule virale infectieuse, qui nécessite la présence des huit ARNv. L'empaquetage doit également être suffisamment efficace pour ne pas générer une proportion trop grande de particules virales non infectieuses. Des particules virales incomplètes sont générées naturellement pendant une infection et l'infection d'une même cellule par plusieurs virions permet une complémentation entre celles-ci [205]. La transmission du virus par aérosol fait que la charge virale au moment de l'infection est faible [206], ainsi la complémentation entre des particules virales incomplètes manquerait d'efficacité et donc l'empaquetage doit permettre de générer un maximum de particules virales infectieuses disposant du jeu complet d'ARNv.

La plupart des virions ne possèdent pas des copies fonctionnelles de l'ensemble des ARNv, une majorité étant en manque d'au moins un segment génomique fonctionnel [207, 208]. Cela est en partie dû à la présence de particules virales défectives interférentes, qui comportent au moins un ARNv interférent qui comporte une délétion empêchant la réplication du génome. Il avait été montré dans un premier temps que les ARNv interférents de ces virus avaient comme origine les différents ARNv [209, 210]. L'étude des ARNv interférents présents au sein des particules virales et ceux présents au sein des cellules a permis de montrer que les ARNv interférents n'étaient pas empaquetés avec la même efficacité, montrant ainsi l'existence d'une discrimination entre les ARNv interférents et les ARNv complets [208].

Deux modèles ont été proposés afin d'expliquer l'empaquetage des virus Influenza A : le premier supposant que les segments étaient incorporés sans sélection, en dépendant uniquement du hasard, le second que des séquences spécifiques à chaque ARN permettait la sélection de chacun des ARNv.

L'hypothèse de l'empaquetage aléatoire suppose que la grande majorité des particules virales retrouvées ne soient pas infectieuses et que l'on retrouverait une grande variabilité dans le nombre de segments d'ARNv et le nombre de chacun des ARNv au sein des virions. Certaines études semblent montrer que l'exclusion des ARN interférents n'était pas retrouvée systématiquement et que les particules virales infectieuses comprendraient plus de 8 ARNv ont étayé cette hypothèse [211–213].

Le modèle de l'empaquetage sélectif était renforcé par le fait que la présence d'ARN défectifs entraîne une diminution de l'empaquetage des ARNv parentaux [204, 214]. De plus, alors que les 8 ARNv sont présents dans des quantités différentes dans les cellules infectées [215, 216], on les retrouve en quantités égales dans les virus [215, 217, 218] avec la présence d'un seul exemplaire de chaque ARN [219, 220], ce qui suppose l'existence d'un mécanisme de sélection. La grande majorité des particules virales comporte une copie de chacun des 8 ARNv et le reste en comporte un nombre inférieur ce qui est également en faveur de l'empaquetage sélectif [100, 102]; forcer l'insertion d'un neuvième segment nécessite une forte

sélection pour le conserver [100, 211]. Selon le modèle d'empaquetage aléatoire, environ 0,2% de particules virales produites seraient infectieuses [220] alors qu'on en observe entre 1 et 10% [204], étayant l'hypothèse d'un modèle sélectif. La découverte et la caractérisation de séquences spécifiques à chaque segment d'ARNv et critiques pour son empaquetage a permis d'argumenter fortement en faveur d'un empaquetage sélectif, qui est aujourd'hui le modèle communément admis.

Les différents ARNv ont été étudiés afin de localiser les signaux d'empaquetage et de déterminer d'éventuelles caractéristiques communes. Dans l'ensemble, les régions situées aux extrémités 5' et 3' des ARNv se sont révélées importantes pour l'empaquetage des ARN sans que la raison ne soit identifiée. Il est important de noter que souvent, ces régions sont aussi importantes pour la réplication du virus.

Des signaux d'empaquetage ont été identifiés au sein de l'extrémité 5' de l'ARNv du segment PB2 de la souche A/WSN/33 (H₁N₁), aux positions 2218-2224 et 2296-2298. Ces signaux couvriraient une région chevauchant la région non-codante et la région codante du segment [221-225].

De même, pour le segment PB1, un signal d'empaquetage a été identifié également en 5' de l'ARN, aux positions 2256-2279 [222, 223, 225].

Dans le cas du segment PA de la souche A/WSN/33 (H₁N₁), des signaux d'empaquetage ont été identifiés aussi bien aux extrémités 5' que 3' de l'ARNv (de respectivement 40 et 66 bases) [222, 223, 225]. Des signaux d'empaquetage ont aussi été identifiés en 5' et 3' (aux positions 2176-2220 à l'extrémité 5' et 13-24 à l'extrémité 3') dans une souche A3/e-3, un clone muté dans le gène NS2 de la souche A/Aichi/2/68 [226].

Les séquences en 3' (9 nucléotides) et 5' (80 nucléotides) du segment HA de la souche A/WSN/33 (H₁N₁) ont été identifiées comme suffisantes pour l'empaquetage du segment, même lorsque les séquences codantes sont remplacées [227].

Sur le segment NP, 60 nucléotides à l'extrémité 3' et les 120 nucléotides à l'extrémité 5' sont nécessaires pour obtenir une incorporation de l'ARNv dans les virions [228]. Plus précisément, les codons 464-466 à proximité de l'extrémité 5' sont nécessaires pour l'empaquetage du segment de la souche A/WSN/33 (H₁N₁). La mutation de cette région entraîne également une réduction de l'empaquetage du segment PA [229].

Pour le segment NA de la souche A/WSN/33 (H₁N₁), les régions non-conservées aux extrémités du segment sont nécessaires pour l'empaquetage [230], de même que la région codante [231]. En étudiant l'effet de mutations dans cette dernière région, celle-ci a été restreinte aux nucléotides 1659-1671 [232]. Une région d'interaction a pu être étudiée plus en détails dans la région terminale 5' du segment NA de la souche A/Puerto Rico/8/1934 (H₁N₁), montrant que des mutations dans la région 22-66 entraîne une baisse de la réplication et de l'efficacité de l'empaquetage [233].

Parmi les codons conservés du segment M de la souche A/Puerto Rico/8/1934 (H₁N₁), certains sont critiques pour l'empaquetage du segment [234]. Les extrémités des régions codantes de la souche WSN sont nécessaires pour avoir un empaquetage efficace [235].

Quant au segment NS de la souche A/WSN/33 (H₁N₁), les nucléotides 16-56 de l'extrémité 3' sont critiques pour l'empaquetage de celle-ci [236]; cette région a été affinée ultérieurement aux nucléotides 27-35 [237]. Chez la souche A/equine/Newmarket/7339/79 (H₃N₈) ce sont les 150 nucléotides de l'extrémité 5' du segment NS qui sont essentiels pour l'empaquetage [238, 239].

Des signaux d'empaquetage ont également été identifiés au sein de virus Influenza B, dans les régions codantes aux extrémités 3' et 5' du segment 4 de la souche B/Yamanashi/98 [240]. Dans l'ensemble, les données ont montré l'importance des séquences non-codantes pour l'empaquetage des différents ARNv, bien que l'existence de signaux d'empaquetage au sein des séquences codantes a pu être montrée.

L'étude des codons conservés au sein des différentes souches montre que certains signaux d'empaquetage sont conservés entre les souches virales [224, 225]. Lorsque des mutations sont introduites dans les régions identifiées comme des signaux d'empaquetage, le virus évolue pour restaurer les séquences sauvages [241]. Bien que les régions comportant des signaux d'empaquetage aient tendance à être conservées du point de vue évolutif, les structures pourraient avoir un rôle plus important que la séquence seule [237, 242].

La formation progressive des complexes supramoléculaires comportant les 8 RNPv [93, 94] se ferait au cours du transport de celles-ci vers la membrane cellulaire, étape durant laquelle les différentes RNPv sont concentrées, favorisant l'établissement de liaisons directes entre les signaux d'empaquetage [243–246]. Les signaux d'empaquetage ont des compatibilités différentes suivant leur degré de conservation, qui influent sur les liaisons pouvant être établies entre RNPv provenant de virus différents, ce qui joue un rôle dans le réassortiment génétique [247]. La longueur des régions non-codantes pourrait également jouer un rôle [248]. L'établissement d'un complexe de 8 ARNv est nécessaire afin qu'un jeu complet d'ARNv soit empaqueté dans une particule virale et avoir une réplication optimale [223, 225, 232, 249]. Toutes les RNPv, ne sont pas empaquetées avec la même efficacité et un modèle a été proposé selon lequel les segments PB2, PA, NP et M commenceraient à former un complexe auquel les segments PB1, HA, NA et NS se joindraient ensuite [250].

Il a été suggéré que les signaux d'empaquetage des ARNv possèderaient deux rôles en agissant en tant que signal d'incorporation du génome et en tant que signal de *genome-bundling*. Le premier permettrait l'incorporation de l'ARNv dans les virions tandis que le second servirait à l'empaquetage du jeu complet des ARNv. Les signaux d'empaquetage présents dans les régions non-codantes serviraient à l'incorporation du génome et ceux présents au sein des régions codantes à l'empaquetage de l'ensemble des ARNv [251].

Il a récemment été montré qu'il existe un code nucléoprotéique qui impacte l'empaquetage des ARNv. Des mutations introduites au sein de la protéine NP entraînent un empaquetage irrégulier des ARNv avec certains ARNv qui sont sur-représentés [252].

D. Interactions ARN-NP

a. Structure de la protéine NP

La protéine NP est une protéine de 56 kDa jouant un rôle dans la formation de la structure hélicoïdale de l'ARN [253] (fig. 9). Cette protéine monomérique tend à oligomériser même en absence d'ARN, pouvant former des trimères, des tétramères, des anneaux et des polymères ressemblant à des RNPv [254], bien

qu'elle soit maintenue principalement sous forme monomérique avant la formation des RNPv [255, 256]. *In vitro*, la concentration en sel joue un rôle crucial sur l'état d'oligomérisation de la protéine NP. A 50 mM en NaCl, la protéine est essentiellement sous forme monomérique, alors qu'à 150 et 300 mM de NaCl elle forme principalement des trimères [255, 257].

La phosphorylation de la protéine NP régule également sa liaison à l'ARN et son oligomérisation, notamment *via* les sites de phosphorylation présent dans le NLS de l'extrémité N-ter [256, 258]. La phosphorylation de ces sites permet d'empêcher l'oligomérisation de la NP et de la conserver sous forme monomérique [256, 259]. La présence d'un ARN d'une taille d'au moins 12 nucléotides permet l'oligomérisation de la NP et la formation de structures ressemblant aux RNPv tandis que l'utilisation d'un ARNv complet permet l'obtention de RNPv similaires à celles observées *in vivo* [257, 260].

La protéine NP a une forme de croissant et se compose d'un domaine « tête » (*head domain*) duquel dépasse un bras flexible et d'un domaine « corps » (*body domain*). Cette organisation est partagée par les différentes souches d'Influenza A [261–264]. Le bras flexible s'insère dans le domaine « corps » d'une autre NP pour former les oligomères, mais il se replie dans le site de liaison à l'ARN au sein de la protéine monomérique [256, 261–263, 265], ce qui expliquerait que la NP monomérique ait une affinité moindre pour l'ARN que les formes trimériques et tétramériques [256, 263, 265, 266].

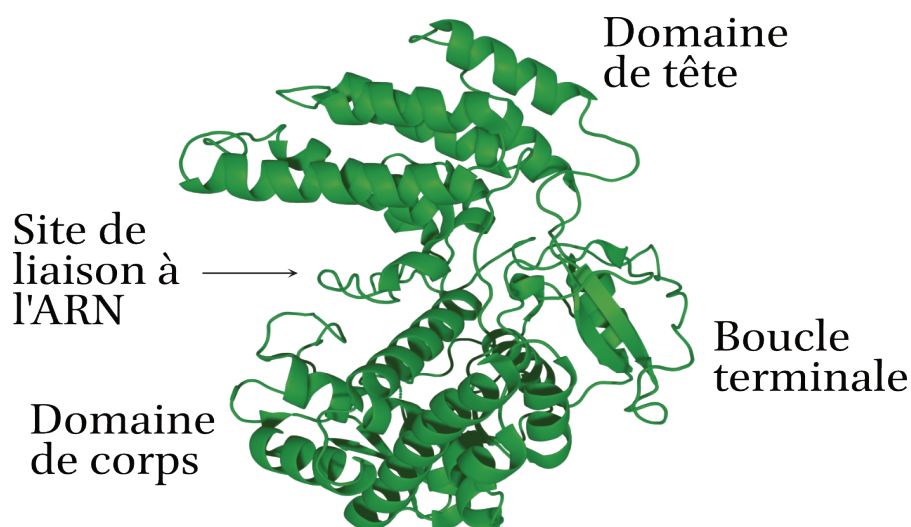


FIGURE 9. – Structure 3D de la protéine NP d'après les données de Chenavas *et al.*, 2013 [256]. La représentation 3D de cette protéine est interactive sous une version récente de Adobe Acrobat DC.

Entre les domaines tête et corps se trouve un sillon comportant de nombreux résidus chargés positivement qui est le site de liaison à l'ARN [261, 262, 267, 268].

La protéine NP possède une structure qui comporte des similarités entre les différents genres d'*Orthomyxoviridae*, que ce soit parmi les Influenza B [269], D [270] ou d'un Isavirus [271].

b. Interactions protéine NP et ARN viral

La longueur de l'ARN influe sur l'affinité de la protéine NP, l'affinité maximale étant atteinte pour des ARN d'une taille d'environ 12 nucléotides au minimum [260]. En absence de protéine NP, l'ARN viral ne semble pas adopter la structure en poêle à frire [96]. La protéine se lie au squelette ribose-phosphate de l'ARN en exposant les bases au milieu environnant [95]. Elle se lie préférentiellement à l'ARN simple brin [95, 272].

Dans les RNPv, on décompte une protéine pour 20 à 24 nucléotides [89, 91, 273], cependant les études de la fixation de la NP sur l'ARNv indiquent que seulement environ 12 nucléotides seraient en interaction avec la NP [271]. Les différents sites de fixation de la NP seraient séparés en moyenne de 25 nucléotides [274].

La répartition des protéines NP n'est pas uniforme le long des ARNv, certaines régions fixant plus densément la protéine NP que d'autres [274–276]. La protéine NP ne semble pas reconnaître de motifs précis et les séquences liées par la protéine NP ne semblent pas être conservées [95, 275, 276]. La protéine se fixerait cependant préférentiellement à des régions riches en G et pauvres en U [275, 276].

c. Structure de l'ARN viral et interactions inter-segments

Le mécanisme d'action des séquences importantes pour l'empaquetage est longtemps resté incertain. L'hypothèse fut émise que les RNPv interagissaient entre elles *via* des interactions ARN-ARN [229, 231]. L'existence de telles interactions ARN-ARN, ainsi que leur rôle a depuis pu être démontrée [243, 244].

Pour la souche Moscou (H_3N_2), la délétion de la région 5' du segment M, désignée comme un signal d'empaquetage, entraîne la perte de l'interaction avec le segment NA *in vitro*. Le même effet est observé pour l'interaction NA-M quand la région 5' du segment NA est délétée. De même, la délétion des régions 5' et 3' du segment M entraînent une perte de l'interaction avec le segment HA [243].

Bien que les signaux d'empaquetage soient principalement considérés comme étant situés dans les 100 nucléotides à chaque extrémité des ARNv, certains signaux d'empaquetage peuvent se situer hors de ces régions, puisque l'on retrouve par exemple une région 514-724 du segment M d'un virus aviaire H_5N_2 impliquée dans l'interaction avec le segment HA *in vitro*. De même la région 613-857 de l'ARNv HA est impliquée dans l'interaction avec M. Les régions 1939-2198 de l'ARNv PB1 et 785-1170 de l'ARNv PB2 sont importantes pour les interactions entre ces segments. Toujours dans la même souche H_5N_2 , la délétion des régions 235-456 et 635-834 de l'ARNv PA diminue fortement l'interaction avec NS tandis que la délétion des nucléotides 435-656 l'abolit complètement [104].

La délétion de régions peut permettre d'identifier grossièrement des régions comportant les signaux d'empaquetage, mais les régions identifiées sont grandes et les délétions peuvent avoir un impact sur la structure de l'ARNv. A l'aide d'oligonucléotides, il a été possible d'affiner les régions identifiées. La région

364-374 du segment NS serait en interaction avec la région 449-459 du segment HA. La région 1216-1226 de l'ARNv NP interagirait avec la région 660-670 de l'ARNv NS. De même, une interaction s'établit entre la région 595-608 du segment M et la région 654-667 du segment HA. On retrouve également une autre interaction entre les régions 626-633 du segment M et 336-343 du segment HA, étayant l'hypothèse d'une redondance des signaux d'empaquetage. Les régions 495-524 et 525-554 du segment NS pourraient interagir avec la région 125-514 du segment PA. Enfin les régions 2059-2118 du segment PB₁ et 961-1020 du segment PB₂ seraient en interaction. L'existence d'une de ces interactions a pu être démontrée *in cellulo* [104, 244].

Les techniques plus récentes basées sur le séquençage à haut débit permettent d'analyser l'intégralité du génome et de rechercher simultanément l'ensemble des interactions possibles *in viro* et *ex viro* sur les souches A/WSN/33 (H₁N₁), A/Puerto Rico/8/1934 (H₁N₁) et A/Udorn/1972 (H₃N₂) [277] et *in vivo* sur une souche A/WSN/33 (H₁N₁) [278]. Dans une étude récente, les réseaux d'interaction des ARNv *in vitro* ont été étudiés par des retards sur gel sur la souche A/WSN/33 (H₁N₁) [279].

En parallèle, des études ciblées ont étudié en détail certaines interactions. Des mutations dans une région de 13 nucléotides de l'ARN HA de la souche A/WSN/33 (H₁N₁) ont montré que cette région est impliquée dans une interaction avec le segment PB₂ [280]. Dans un manuscrit déposé sur BioRxiv, une tige boucle présente dans la région 5' de l'ARNv PB₂ a été identifiée comme un signal d'empaquetage conservé dans plusieurs souches virales [281].

Ces études ont mis en évidence l'existence d'une certaine redondance entre les signaux d'empaquetage, certaines régions pouvant interagir avec plusieurs segments [277, 278]. Éliminer certaines de ces interactions peut avoir pour effet de remodeler le réseau d'interaction entre les RNPv [278] ou bien provoquer l'apparition de mutations qui vont compenser les premières [280].

Les interactions ARN-ARN dirigent l'empaquetage des ARN viraux, comme cela a pu être montré pour la région 1776-2070 du segment PB₁ de la souche Udorn qui interagit avec le segment NA [282]. La tendance de certains ARN viraux à être empaquetés plus fréquemment ensemble proviendrait de ces interactions [247, 277, 283–285], même si certaines associations, observées notamment dans les souches utilisées pour le vaccin, ne sont pas optimales pour la réplication virale [282, 286–288].

Les signaux d'empaquetage ainsi que les réseaux d'interaction identifiés pour la souche A/WSN/33 (H₁N₁) (fig. 10) montrent que la plupart des signaux d'empaquetage identifiés dans la littérature se situent aux extrémités 3' et 5', tandis que les interactions entre les différents segments sont également situées dans la région codante. Les réseaux d'interaction peuvent également varier pour une même souche virale, illustrant la complexité de l'étude des interactions inter-RNP. Certains signaux d'empaquetage identifiés précédemment sont retrouvés dans l'étude de Dadonaite [277], mais pas dans celle de Le Sage [278] et inversement.

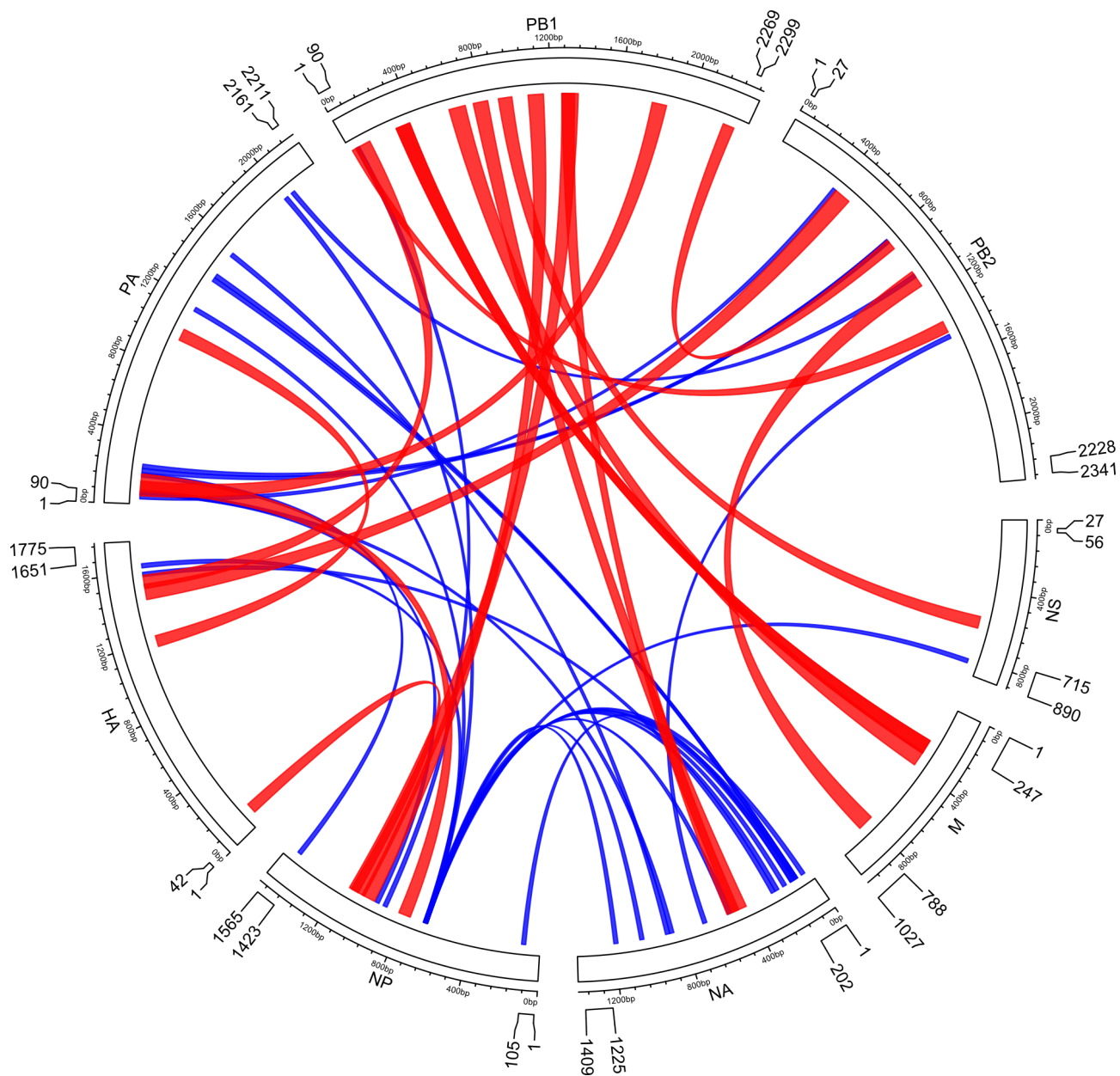


FIGURE 10. – Représentation des signaux d’empaquetage et des interactions inter-RNP identifiées dans la littérature pour la souche A/WSN/33 (H_1N_1). Les signaux d’empaquetage, représentés sur l’extérieur de la figure ont été identifiés au cours de plusieurs études (pour plus d’informations voir [289]). Les arcs rouges représentent les 26 interactions les plus fréquemment retrouvées par l’équipe de Le Sage [278] et en bleu celles déterminées par l’équipe de Dadonaite [277])

Les ARNv des virus Influenza A sont des ARN fortement structurés. Bien que l'on retrouve une diversité au niveau des séquences de ces ARN très importante [290], certaines structures sont plus ou moins largement conservées. Après fixation de la protéine NP sur l'ARN, des structures secondaires sont présentes à la surface des RNPv [260], qui possèdent une flexibilité structurale permettant notamment l'empaquetage du génome [291]. En effet certains signaux d'empaquetage sont localisés au sein de structures secondaires conservées [274, 292]. Plusieurs ont déjà été identifiées dans la littérature, par exemple au sein des ARN M [292], HA [293], NP [294] et PB2 [295]. Les régions moins densément liées à la NP pourraient former des structures secondaires qui exposeraient les signaux d'empaquetage [274, 275, 277, 278].

d. Activité chaperon à ARN des protéines

La structure des ARN est critique pour leurs fonctions et lors du repliement des ARN, des structures secondaires stables se forment. Ces dernières interagissent ensuite afin de former des structures tertiaires qui tendent à minimiser l'énergie libre de la molécule. Dans le cas des virus Influenza A, la structure de l'ARN va être d'une importance particulière pour les interactions entre les RNPv. Il est cependant complexe de diriger le repliement des ARN vers une structure spécifique, ce qui peut aboutir à la formation de nombreux ARN mal repliés, mais dont la structure est stable [296–298].

Le remodelage de l'ARN peut se faire *via* deux types de protéines, les hélicases et les protéines chaperons à ARN. Les hélicases utilisent de l'ATP afin de modifier la structure de l'ARN tandis que les protéines chaperons à ARN sont ATP-indépendantes [299]. L'implication de ces protéines dans le cycle viral des Influenza A pourrait donc avoir un impact majeur sur l'empaquetage des différents ARN.

Deux définitions existent à propos des protéines chaperons de l'ARN. Celle d'origine stipule qu'il s'agit de protéines qui se lient à l'ARN et permettent le repliement vers une structure tertiaire spécifique [297] de façon transitoire. Une fois le repliement de l'ARN réalisé, la protéine n'est plus nécessaire et son élimination n'affecte pas la structure de l'ARN [297, 300–308]. La deuxième définition correspond à la même idée bien que la notion d'une fixation transitoire de la protéine à l'ARN en soit absente. Cela est notamment vrai dans le domaine de la virologie dont des protéines de capsid qui ont un effet de remodelage de l'ARN sans que la protéine ne soit retirée de l'ARN sont catégorisées comme étant des chaperons de l'ARN. Bien qu'initialement hypothétisées uniquement *in vitro* [297], leur existence *in vivo* a pu être démontrée depuis [309]. Ces protéines ont des régions désordonnées et ne semblent pas posséder de motif conservé [310].

Ces protéines chaperons, ici selon la définition la plus large, sont retrouvées au sein de plusieurs virus, les ARN viraux étant fortement structurés afin de réguler et initier les différentes étapes des cycles viraux. On retrouve notamment des structures qui seront importantes au sein des régions 5'- et 3'-UTR et dans les régions codantes qui permettent de former des structures secondaires qui agissent *in cis* [310]. Parmi les virus de la famille *Flaviviridae*, la protéine *core* possède une activité chaperon de l'ARN. Elle jouerait un rôle dans la dimérisation de l'ARN viral [305, 311–313].

L'antigène de l'hépatite D possède une activité chaperon de l'ARN dans son domaine N-ter [314, 315] qui

permet de réguler le clivage ainsi que la ligation de l'ARN viral afin d'agir sur la réplication et la transcription [315–317].

La protéine N des *Coronaviridae* possède également une activité chaperon de l'ARN [318] qui pourrait agir sur la réplication et sur la répression de l'activité d'interférence par ARN de la cellule [319].

Au sein des *Picornaviridae* les protéines 3AB ainsi que la protéine 2C de l'enterovirus 71 possèdent une activité chaperon de l'ARN, bien leur rôle dans le cycle viral n'est pas encore connu [320–322].

Il a pu être montré que la protéine 2C des *Iflaviridae* possède également cette activité chaperon, bien que le rôle celle-ci ne soit pas encore déterminé [323, 324].

La protéine de nucléocapside N des *Bunyaviridae* est une chaperon de l'ARN, qui permet de dissoudre la structure en poêle à frire de l'ARN viral afin d'initier la réplication du génome viral [325].

Au sein des *Reoviridae*, les protéines VP5 du cytovirus 5 et NSP2 des rotavirus sont aussi des chaperons de l'ARN [326, 327].

La protéine de capsid du virus de la dengue possède une activité chaperon de l'ARN qui permet la régulation de la synthèse d'ARN [328–330].

Le virus de l'immunodéficience humaine possède plusieurs protéines possédant une activité chaperon de l'ARN. La protéine NCp7 permet notamment la dimérisation de l'ARN [331–335]. Cette protéine est également impliquée dans la transcription inverse du génome [336]. La protéine Tat possède également une activité chaperon [337] et serait également impliquée dans la transcription inverse [338]. La protéine Vif a aussi une activité chaperon et serait impliquée dans la transcription inverse, la dimérisation de l'ARN et sa maturation [339, 340].

Il a pu être montré récemment que la protéine NS3 des norovirus possède une activité chaperon de l'ARN [341, 342].

La protéine de nucléocapside NC du virus de la leucémie murine de Moloney possède une activité chaperon de l'ARN qui s'exerce par la fixation de la protéine sur des guanosines exposées et la déstabilisation des fortes interactions de ces guanosines, qui permet de remodeler l'ARN afin d'améliorer l'efficacité de la dimérisation de l'ARN viral [343].

Certaines des protéines possédant une activité chaperon de l'ARN, indépendante de l'ATP possèdent également une activité hélicase, dépendante, elle, de l'ATP. C'est le cas de la protéine 2C de l'enterovirus 71 [344] et de la protéine NS3 des norovirus [341].

E. Etude de la structure de l'ARN

Les premières études de la structure des ARN datent des années 1960, et les méthodes utilisées ont depuis lors évoluées, afin de tirer avantage des progrès réalisés dans le séquençage. L'étude des ARN s'est effectuée à l'aide d'enzymes de restriction dans un premier temps puis des techniques utilisant des réactifs chimiques ont fait leur apparition.

Ainsi, au fil des années, plusieurs réactifs chimiques ont été mis au point afin d'obtenir différentes informations sur la structure des ARN, dans des conditions différentes, certains réactifs n'étant, par exemple, pas utilisables dans des conditions *in vivo*.

L'arrivée du premier réactif SHAPE, le NMIA, au début des années 2000 a permis, en ciblant le 2'OH du sucre, d'obtenir des informations sur l'ensemble des quatre nucléotides avec un seul réactif. Ce nouveau réactif a été couplé peu après son apparition avec les techniques d'électrophorèse capillaire, afin d'obtenir des informations sur des ARN plus grands et de quantifier le signal obtenu, aboutissant à la technique SHAPE (*Selective 2'-hydroxyl acylation analyzed by primer extension*).

Les avancées scientifiques de ces dernières années sur le séquençage à haut débit ont également été prises en compte afin d'améliorer la technique SHAPE. Au lieu d'induire des arrêts de la transcription, des conditions spécifiques pour la rétro-transcription sont utilisées, afin d'induire des mutations aux positions modifiées. Ces avancées permettent une quantification plus fine, ainsi que mener des approches sur des génomes complets.

L'historique des méthodes d'étude des ARN a fait l'objet d'une revue rédigée au cours de la thèse [345], qui est présentée au sein des annexes.

II | Problématique et sujet de recherche

Les virus Influenza A sont les agents étiologiques de la grippe saisonnière et peuvent causer des pandémies grippales. L'absence d'activité correctrice de la polymérase virale autorise des mutations fréquentes et la segmentation du génome en 8 ARN génomiques de sens négatif permet l'échange de segments génomiques par réassortiment génétique, ce qui favorise l'apparition et l'évolution de nouvelles souches virales. La dérive antigénique du virus est à l'origine des épidémies saisonnières. Le réassortiment génétique est un mécanisme consistant en l'échange de segments génomiques entre différentes souches de virus coinfectant une même cellule. Ce mécanisme peut être à l'origine de l'apparition de nouvelles souches hautement pathogènes.

Une particule virale n'étant infectieuse que si les 8 ARN viraux (ARNv) sont empaquetés, un mécanisme d'empaquetage sélectif est utilisé par le virus. L'empaquetage du génome des virus Influenza A est régi par un ensemble de règles encore peu connues qui dirigent les possibilités de réassortiment entre souches virales et ont donc un impact majeur sur l'apparition de nouvelles souches par ce biais. La connaissance des règles d'empaquetage aiderait à évaluer les risques d'apparition de nouvelles souches grippales et serait donc utile pour la préparation à de nouvelles pandémies.

De nombreuses études ont démontré que des signaux d'empaquetage sont présents au sein des ARN génomiques, tant dans les régions codantes que non codantes. Ces signaux peuvent être spécifiques d'une souche ou conservés entre plusieurs souches. Le modèle actuellement admis est que des interactions entre segments d'ARNv s'établissent au niveau des signaux d'empaquetage; ces interactions formeraient un réseau redondant qui stabiliserait un complexe supramoléculaire comprenant les 8 ARNv. A la fois la séquence et la structure des signaux d'empaquetage semblent jouer un rôle majeur dans leur fonction.

Les ARNv s'enroulent sur un squelette formé par la protéine NP, hormis leurs extrémités qui recrutent un complexe polymérase formé de trois sous-unités, formant ainsi des ribonucléoprotéines virales. Il est dès lors certain que la protéine NP doit jouer un rôle important dans l'établissement d'interactions entre les ARNv en affectant leur structure et en modulant l'accessibilité des séquences impliquées.

En étudiant l'impact de la protéine NP sur la structure des ARNv il devrait être possible de déterminer les structures et les séquences préférentiellement fixées par la NP, d'identifier les régions de l'ARN en simple brin en présence de NP, et donc disponibles pour établir des interactions avec d'autres ARN génomiques, et mettre en évidence un éventuel effet chaperon de l'ARN de cette protéine.

L'objectif global de mon travail de thèse était de compléter les travaux préliminaires de Damien

Ferhadian afin d'étudier la fixation de la protéine NP sur deux ARNv en utilisant une panoplie de sondes chimiques réagissant soit avec le squelette ribose-phosphate, soit avec les bases des ARNv. Le but était de comparer la réactivité des ARNv soit en absence de protéine NP, soit en complexe avec cette protéine, soit après formation des complexes ARN-NP puis enlèvement de la protéine. Cette dernière condition visait à mettre en évidence d'éventuels remaniements structuraux permanents induits par la NP. Faute de temps, la totalité de ce programme n'a pas pu être mise en œuvre, cependant la technique de SHAPE nous a permis de mettre en évidence des remaniements structuraux transitoires induits par la NP. Nos travaux montrent que la NP a un impact profond sur la structure des ARNv et devrait donc influencer de manière importante les interactions entre ARNv. Enfin, nous avons montré que la technique de SHAPE permet d'identifier des contacts entre ARNv en présence de NP.

III | Matériel et méthodes

A. Production des ARN

Des plasmides pUC₂₀₀₀ contenant les différents segments (NS ou M d'une souche A/WSN/33 (H₁N₁)) sous la dépendance d'un promoteur de l'ARN polymérase du phage T7 sont utilisés (fig. 11). Les plasmides étaient déjà disponibles au laboratoire à mon arrivée.

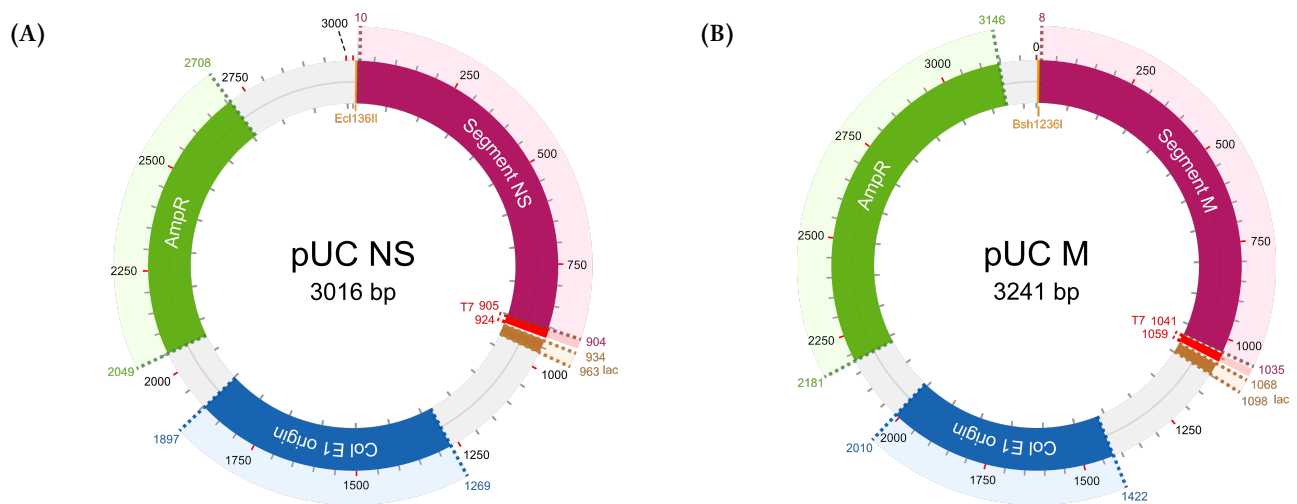


FIGURE 11. – Cartes des plasmides pUC 2000 NS (A) et pUC 2000 M (B).

Un μg de plasmide est linéarisé avec l'enzyme adéquate (Ecl136II pour pUC₂₀₀₀ NS, Bsh1236I pour pUC₂₀₀₀ M). Après vérification de la linéarisation sur gel d'agarose 0,8 %, une transcription *in vitro* est réalisée. A 25 μg de plasmide linéarisé sont ajoutés les NTP à une concentration finale de 4 mM chacun, 1 mM de spermidine, 5 mM de DTT, du triton X-100 à une concentration finale de 0,01 %, 160 U de RNasin (Promega), 0,05 mg/mL final de BSA, 0,5 μL de pyrophosphatase (Roche) dans un tampon T7 1X (0,4 M Tris-HCl pH 8, 0,15 M MgCl₂, 0,5 M NaCl) et un volume final de 300 μL pendant 3 h dans un bain-marie à 37°C. Trente-cinq μL de tampon DNaseI 10X, 3,5 μL de DNaseI (Roche) et 11,5 μL d'H₂O milliQ sont ajoutés avant une incubation d'une heure au bain-marie à 37°C. La réaction est arrêtée par l'ajout de 150 μL d'EDTA 250 mM. Une extraction phénol-chloroforme (v/v) est réalisée, puis après addition de 3 volumes d'éthanol et 0,1 volume d'acétate de sodium 3M à pH 5 une précipitation est effectuée durant la nuit à -20°C. Après centrifugation 30 min à 21 000 g à 4°C, les culots sont lavés 2 fois à l'éthanol 70 % froid (-20°C) et séchés, puis les ARN sont repris dans 250 μL d'H₂O milliQ et purifiés sur colonne TSK Gel G4000SW (Tosoh Bioscience) à un débit de 1 mL/min avec un tampon 0,2 M AcNa, 1% éthanol. Les fractions contenant des ARN sont rassemblées et précipitées à l'éthanol comme décrit ci-dessus et reprises dans 100 μL d'H₂O milliQ. Les différentes fractions sont quantifiées puis l'intégrité de l'ARN est analysée

par électrophorèse sur un gel d'acrylamide 8 %, urée 8 M de 8 cm. Après 2 h de migration à 200 V, les bandes sont révélées à l'aide de *Stains-All*.

B. Nouveau protocole de production de la protéine NP

Un plasmide pet22b fournit par Thibaut Crépin (IBS, Grenoble) contenant la séquence de la protéine NP de la souche A/WSN/1933 (H₁N₁) est utilisé.

a. Préparation des bactéries compétentes

Deux cents mL de milieu LB sans antibiotique sontensemencés avec 50 µL de bactéries BL21 DE3 pLysS et incubés à 37°C. Lorsque la DO atteint une valeur comprise entre 0,4 et 0,5, les cultures sont centrifugées 15 min à 4 000 g. Les culots sont repris dans 120 mL de solution A (10 mM MnCl₂, 10 mM MES pH 6,5, 50 mM CaCl₂) pendant 20 min. Après une nouvelle centrifugation, le culot est repris dans 120 mL de solution B (10 mM MnCl₂, 10 mM MES pH 6,5, 50 mM CaCl₂, 15% glycérol). Les tubes sont congelés dans de la carboglace puis conservés à -80°C.

b. Transformation des bactéries

Cinquante µL de bactéries BL21 DE3 pLysS compétentes sont transformées avec 50 ng de plasmide (fig. 12 (1)). Les bactéries sont incubées sur glace pendant 30 min. Un choc thermique à 42°C pendant 45 s est réalisé, puis les tubes sont déposés sur glace pendant 2 min. Les tubes sont centrifugés à dans une centrifugeuse de paillasse à environ 8 000 g puis les culots sont repris dans 50 µL de LB et étalés sur boîte LB + ampicilline (0,1 mg/mL).

c. Ancien protocole de production de la protéine NP

Deux cents mL de milieu LB supplémentés avec 200 µL d'ampicilline (100 mg/mL) et 200 µL de chloramphénicol (30 mg/mL) sontensemencés avec un stock de bactéries congelées dans 50% de glycérol. Une incubation durant la nuit à 37°C avec une agitation à 160 rpm est réalisée. Deux x 2 L de LB supplémentés avec 2 mL d'ampicilline et 2 mL de chloramphénicol sontensemencés à une DO finale d'environ 0,05. Les cultures sont ensuite incubées à 37°C avec une agitation à 160 rpm. Quand la DO est comprise entre 0,7 et 0,9, les cultures sont refroidies dans un bain eau-glace en agitant les erlenmeyers, jusqu'à atteindre une température de 18°C. Les cultures sont ensuite incubées 30 min à 18°C avant induction avec 1 mM d'IPTG (Euromedex). Le jour suivant, les cultures sont centrifugées 15 min à 4 000 g. Si les culots ne sont pas utilisés immédiatement, ils sont congelés à -20°C.

d. Production de la protéine NP

Quatre mL de LB supplémentés avec 4 µL d'ampicilline et 4 de chloramphénicol sontensemencés avec 1 colonie sélectionnée sur boîte LB + ampicilline (fig. 12 (2)). La préculture est incubée durant la nuit à 37°C avec une agitation à 160 rpm. Le lendemain, 200 mL de LB supplémentés avec 200 µL d'ampicilline et

200 µL de chloramphénicol sontensemencés avec 2 mL de la préculture. Une incubation durant la nuit à 37°C avec une agitation à 160 rpm est réalisée. Deux x 2 L de LB additionnés de 2 mL d'ampicilline et 2 mL de chloramphénicol sontensemencés à une DO d'environ 0,05 (fig. 12 (3)). Les cultures sont ensuite incubées à 37°C avec une agitation à 160 rpm. Lorsque la DO est comprise entre 0,6 et 0,8, les cultures sont refroidies dans un bain eau-glace en agitant les cultures jusqu'à atteindre une température de 18°C (fig. 12 (4-5)). Les cultures sont ensuite conservées à 4°C pendant 1 h puis 30 min à 18°C avant induction avec 0,3 mM d'IPTG sur la nuit (fig. 12 (6-8)). Le jour suivant, les cultures sont centrifugées 15 min à 4 000 g (fig. 12 (8)). Si les culots ne sont pas utilisés tout de suite, ils sont congelés à -20°C.

e. Ancien protocole de purification de la protéine NP

Les culots bactériens sont repris dans 80 mL de tampon de lyse (50 mM Tris-HCl pH 7,5, 300 mM NaCl, 1 M NDSB, 2 mM β -mercaptoéthanol). Les bactéries sont lysées par sonication avec un sonicateur 250 TS 20K (Ultrasons Annemasse), 10 s on/10 s off, 10 impulsions. Les lysats sont centrifugés 1 h à 11 000g à 7°C. Les surnageants sont rassemblés dans un même bécher. Une colonne HisTrap FF crude 5 mL (GE Healthcare) est branchée sur une station BioLogic DuoFlow (BioRad). La colonne est équilibrée avec un tampon à bas sel (50 mM Tris-HCl pH 7,5, 300 mM NaCl, 20 mM Imidazole BioUltra, 2 mM β -mercaptoéthanol). Le lysat centrifugé est injecté sur la colonne à un débit de 2 mL/min. Un premier lavage est réalisé avec 20 mL de tampon à haute concentration en sels (50 mM Tris-HCl pH 7,5, 1 M NaCl, 20 mM Imidazole BioUltra, 2 mM β -mercaptoéthanol) à un débit de 2 mL/min. Un second lavage avec 30 mL de tampon à bas sel est réalisé à 2 mL/min. L'élution est réalisée avec 15 mL de tampon d'élution (50 mM Tris-HCl pH 7,5, 300 mM NaCl, 800 mM Imidazole BioUltra, 2 mM β -mercaptoéthanol). Les fractions à conserver sont sélectionnées d'après le profil du chromatogramme. Les fractions sélectionnées sont rassemblées puis chargées sur une colonne HiTrap Heparin 5 mL (GE Healthcare) équilibrée avec le tampon A (50 mM Tris-HCl pH 7,5, 300 mM NaCl, 2 mM β -mercaptoéthanol) à 2 mL/min. Une fois les fractions chargées sur la colonne, un lavage est réalisé avec 20 mL de tampon A à 2 mL/min. L'élution est réalisée sur un volume de 30 mL à 2 mL/min par un gradient linéaire du tampon A au tampon B (50 mM Tris-HCl pH 7,5, 2 M NaCl, 2 mM β -mercaptoéthanol). Les fractions les plus concentrées (jusqu'à un maximum de 3 mL) sont rassemblées et dialysées durant la nuit contre un tampon de tamis moléculaire (20 mM Tris-HCl pH 7,5, 150 mM NaCl, 5 mM β -mercaptoéthanol) à l'aide d'une cassette G2 Slide-A-Lyzer 3mL (ThermoFisher). Le lendemain l'échantillon est récupéré puis quantifié au Nanodrop.

f. Nouveau protocole de purification de la protéine NP

Le protocole est en grande partie inchangé pour les purifications sur les colonnes HisTrap FF crude et HiTrap Heparin à l'exception de quelques modifications. La lyse est réalisée par un sonicateur VC-500 (Sonics & Materials, 500 W, 20 kHz) et une sonde V1A (Sonics & Materials), 1 s on/1 s off, 70 % de charge, 3 x 5 min et 2 mL d'imidazole BioUltra 1 M (Sigma) est ajouté au lysat pour une concentration finale de 20 mM (fig. 12 (9-12)). Un lavage d'un volume de 20 mL de tampon à bas sel à un débit de 2 mL/min est ajouté avec les lavages à haut sel et bas sel de la colonne HisTrap FF crude pour éliminer le matériel qui n'est pas lié à la colonne.

Après la purification sur colonne HiTrap Heparin, les fractions les plus concentrées (jusqu'à un maximum de 15 mL) sont rassemblées et dialysées sur la nuit contre un tampon de tamis moléculaire (20 mM Tris-HCl pH 7,5, 50 mM NaCl, 5 mM β -mercaptoéthanol). Le lendemain l'échantillon est récupéré puis concentré à l'aide d'un Amicon Ultra-15 (30K cut-off, Millipore) jusqu'à obtenir un volume d'environ 1,2 mL (fig. 12 (13-14)).

Une colonne Superdex 200 Increase 10/300 GL (GE Healthcare) est équilibrée avec le tampon de tamis moléculaire. L'échantillon est injecté en 3 fois (environ 400 μ L à chaque injection). Entre chaque injection, 30 mL de tampon de tamis moléculaire est passé à un débit de 0,5 mL/min et les fractions éluées entre 10 mL et 22 mL sont récoltées (fig. 12 (14)). Les fractions d'intérêt sont sélectionnées selon leur absorbance. 10 μ L de chaque fraction est déposé sur gel SDS-PAGE 12 %. Le gel est révélé à l'aide de bleu de Coomassie. Les fractions pures sont ensuite rassemblées puis dialysées avec des cassettes de dialyse Slide-A-Lyzer™ G2 (10K cut-off, 15 mL, Thermofisher Scientific) contre un tampon HEPES (50 mM HEPES pH 8,0, 50 mM KCl, 5 mM β -mercaptoéthanol) (fig. 12 (15)). La protéine dialysée est récupérée dans un tube puis centrifugée 10 min à 21 000g à 7°C. La moitié supérieure du volume est ensuite récupérée dans un nouveau tube afin d'éliminer les agrégats.

La protéine est ensuite quantifiée au Nanodrop. Si la concentration est inférieure à 20 μ M, la protéine est de nouveau concentrée à l'aide d'un Amicon.

g. Caractérisation de la protéine

L'état d'oligomérisation de la protéine est vérifié par DLS. La fonctionnalité de la protéine est ensuite vérifiée par retard de migration sur un gel d'agarose 0,8 % à différents rapports entre la protéine NP et l'ARN NS : 1/100, 1/50, 1/40, 1/30, 1/20, 1/10.

C. Marquage radioactif des ARN et migration sur gel

Un μ L d'amorce (table 3) à 100 μ M est incubé dans un tampon kinase T4 PNK avec 10 U de kinase du phage T4 et 1,4 μ L de γ -³²P-ATP pendant 1 h à 37°C. Des colonnes Micro-BioSpin 6 (Bio-Rad) sont utilisées pour purifier les amorces marquées puis la radioactivité est quantifiée.

Une demie pmole d'ARNv NS est incubée avec 0,1 μ M d'amorce non marquée et 10 000 cpm d'amorce radioactive dans un volume final de 8 μ L par réaction. Les échantillons sont dénaturés 2 min à 90°C puis incubés 2 min sur de la glace. Deux μ L de tampon de repliement 5X (60 mM Tris-HCl pH 6,8 1 M, 25% glycérol, 2%SDS, 0,7 M β -mercaptoéthanol, 0,1% bleu de bromophénol) sont ajoutés à chaque tube qui sont ensuite incubés 10 ou 30 min à température ambiante, 37°C, 45°C ou 50°C.

Les amorces et ARN hybridés sont ensuite déposés sur gel d'acrylamide natif et analysés par électrophorèse 45 min à 200 V. Les gels sont incubés dans une solution de fixation 10% EtOH, 6% AcNa 3 M pendant 15 min puis déposés sur papier Whatman et séchés sous vide pendant 1 h à 65°C. Les gels

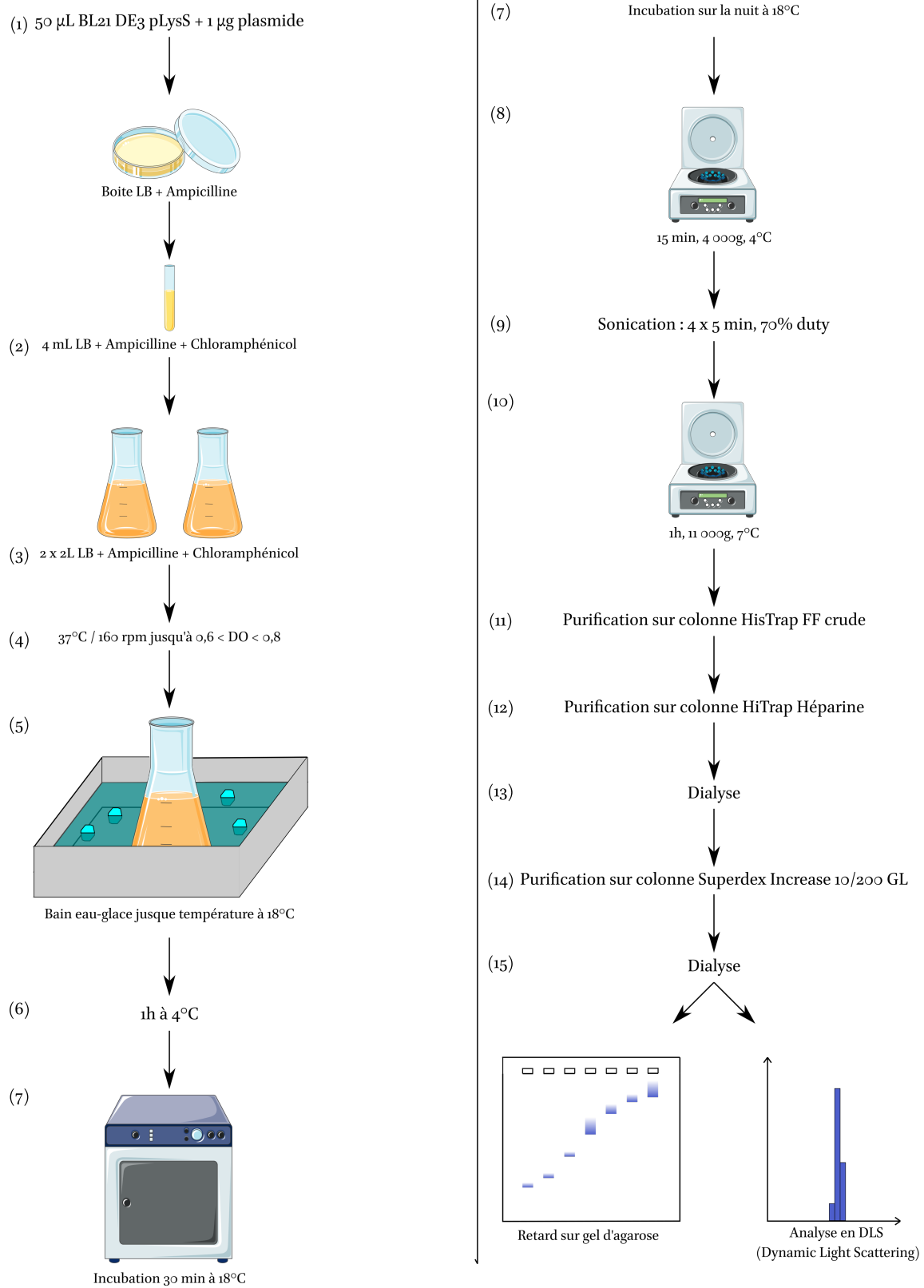


FIGURE 12. – Représentation schématique du protocole de purification de la protéine NP.

Amorce	Séquence
30NS	GAT CCA AAC ACT GTG TCA AGC
341NS	CAT ACC CAA GCA GAA AGT GGC
621NS	CAG AGA TTC GCT TGG TGT TGC
M20	TGA AAG ATG AGT CTT CTA ACC
M312	CAG TTA AAC TGT ATA GGA AGC
M622	AGC AGA GGC CAT GGA TAT TGC
M843	GAT CGT CTT TTT TTC AAA TGC

TABLEAU 3. – Liste des amorces utilisées pour le marquage radioactif.

séchés sont placés dans une cassette avec un écran photostimulable pendant 3 h, puis les bandes correspondant aux espèces radioactives sont détectées avec un Fujifilm FLA-5100 et quantifiées.

D. Modification des ARN

a. Modification des ARN - Ancien protocole

Trois ou 4 pmoles (pour les segments NS et M respectivement) d'ARN sont repliés dans le tampon NMIA (50 mM cacodylate de sodium, 50 ou 150 mM KCl, 5 mM MgCl₂) pendant 30 min à 37°C. Les ARN sont ensuite modifiés avec une concentration de 20 mM ou 50 mM final de NMIA durant 50 min à température ambiante. Une précipitation à l'éthanol avec 1 µL de glycogène (1 µg/µL) est ensuite réalisée pendant 30 min à -80°C. Les tubes eppendorf sont ensuite centrifugés 30 min à 21 000 g. Les culots sont lavés 2 fois avec 300 µL d'EtOH 80% froid et séchés 5 min au SpeedVac puis resuspendus dans 27 µL (NS) ou 36 µL (M) d'eau milliQ.

b. Modification des ARN - Nouveau protocole

Trois conditions sont mises en œuvre en parallèle : les ARN modifiés en absence de protéine (condition NoNP), les ARN modifiés sous la forme de complexes ARN-protéine (condition Comp) et les ARN modifiés après avoir retiré la protéine NP par digestion du complexe à la protéinase K (condition ProtK). Pour chaque condition, un échantillon dit positif est réalisé, ainsi qu'un échantillon dit négatif. L'échantillon positif est traité avec le réactif, qui est remplacé dans les échantillons négatifs par du DMSO pour le NMIA, de l'éthanol 100% pour le DMS et de l'H₂O milliQ pour l'EDC. Huit pmoles ou 10 pmoles (pour les segments NS et M respectivement) d'ARN sont repliés dans le tampon correspondant au réactif : NMIA (50 mM HEPES pH 8,0 à 25°C, 150 mM KCl, 5 mM MgCl₂), DMS (200 mM Bicine pH 8,3 à 25°C, 150 mM KCl, 5 mM MgCl₂) ou EDC (50 mM Borate pH 8,0 à 25°C, 150 mM KCl, 5 mM MgCl₂). Pour les conditions Comp et ProtK, du KCl et du MgCl₂ sont ajoutés à la protéine NP purifiée à des concentrations finales de 150 mM de KCl et 5 mM de MgCl₂ et celle-ci est incubée 20 min à 37°C. Les complexes ARN-NP sont formés pendant 30 min à 37°C dans le tampon de repliement, à un ratio molaire de 1 protéine pour 20 nucléotides.

Les échantillons NoNP et Comp sont modifiés avec les différents réactifs : NMIA (stock à 57,75 mM, concentration finale de 1,65 mM), DMS (stock à 1,05 M, concentration finale de 30 mM) ou EDC (stock à 350 mM, concentration finale de 100 mM). Pour le NMIA, le réactif est neutralisé par un excès d'eau et 10 min d'incubation. Pour le DMS, la réaction est stoppée par l'ajout de 50 mM de β -mercaptoéthanol. L'EDC est quenché par l'ajout de β -mercaptoéthanol à une concentration finale de 20 mM. Les échantillons NoNP et Comp sont ensuite traités avec 1 μ L de protéinase K à 18 mg/mL (Roche) pour 15 μ L de solution pendant 1 h à 37°C avec 10 mM de complexes ribo-vanadyl (Sigma-Aldrich). Trois extractions successives au phénolchloroforme (v/v) sont réalisées, suivie d'une précipitation à l'éthanol. Après formation des complexes ARNv-NP, les échantillons ProtK sont traités avec la protéinase K et 10 mM de complexes ribovanadyl (ribonucléosides complexés à du vanadium). Après incubation 1 h à 37°C, 3 extractions phénol-chloroforme (v/v) successives puis une précipitation à l'éthanol sont réalisées. Enfin les ARN sont modifiés suivant le même protocole que pour les échantillons NoNP et Comp.

Les échantillons sont repris dans 28 μ L (pour les échantillons NS) ou 35 μ L (pour les échantillons M) d'eau milliQ et quantifiés au Nanodrop.

Pour l'ENU, les expériences sont menées en parallèle dans des conditions natives et dénaturantes. 2 pmoles d'ARN sont dénaturées 2 min à 90°C puis refroidis 2 min sur glace. Du tampon 5X dénaturant (1,25 M borate pH 8, 10 mM EDTA) ou 5x natif (1,25 M borate pH 8, 100 mM MgCl₂, 500 mM NaCl, 10 mM EDTA) est ajouté à une concentration finale 1X. Les tubes correspondant à la condition native sont chauffés 30 min à 37°C au bain-marie. De l'ENU à une concentration finale de 75 ou 150 mM est ajouté aux tubes positifs tandis que de l'éthanol 100% est ajouté aux tubes négatifs. La réaction est réalisée pendant 3 h à température ambiante pour la condition native et 2 min à 80°C pour la condition dénaturante. Une précipitation à l'éthanol est réalisée puis les culots sont repris dans une solution 0,3 M AcNa, 10 mM EDTA. Après une seconde précipitation à l'éthanol les tubes sont repris dans 10 μ L de Tris-HCl à pH 9. Les tubes sont incubés 5 min à 50°C puis 7 μ L d'AcNa 3 M et 60 μ L d'éthanol 100% sont ajoutés. Une dernière précipitation à l'éthanol est réalisée puis les culots sont repris dans 7 μ L d'eau milliQ.

c. Modification des ARN - Mise en évidence des interactions entre les complexes ARNv NS-NP et ARNv M-NP

Pour les expériences de cartographie sur le segment NS, 8 pmoles des ARN NS et M produits *in vitro* sont dénaturées 2 min à 90°C puis conservés sur glace pendant 2 min dans deux tubes indépendants. Pour celles sur le segment M, 10 pmoles de chaque ARN sont dénaturées. Le repliement est effectué dans un tampon 50 mM HEPES pH 8,0 à 20°C, 150 mM KCl, 5 mM MgCl₂ pendant 30 min au bain-marie à 37°C. Du KCl et du MgCl₂ sont ajoutés à la protéine NP purifiée à des concentrations finales de 150 mM de KCl et 5 mM de MgCl₂. Elle est ensuite incubée 20 min à 37°C. La protéine NP est ensuite ajoutée aux tubes contenant les ARN. Un μ L de RNasin est ajouté à chaque tube, puis les complexes ARNv-NP sont formés pendant 30 min au bain-marie à 37°C. Après formation des complexes, les complexes ARNv NS-NP et ARNv

Amorce	Fluorophore 5'	Séquence
30NS	Vic Ned	GAT CCA AAC ACT GTG TCA AGC
341NS	Vic Ned	CAT ACC CAA GCA GAA AGT GGC
621NS	Vic Ned	CAG AGA TTC GCT TGG TGT TGC
M20	Vic Ned	TGA AAG ATG AGT CTT CTA ACC
M312	Vic Ned	CAG TTA AAC TGT ATA GGA AGC
M622	Vic Ned	AGC AGA GGC CAT GGA TAT TGC
M843	Vic Ned	GAT CGT CTT TTT TTC AAA TGC

TABLEAU 4. – Liste des amorces fluorescentes utilisées pour la rétrotranscription.

M-NP sont rassemblés dans un même tube et incubés à nouveau au bain-marie à 37°C pendant 1 h. Du NMIA dilué dans du DMSO est ajouté à chaque tube positif à une concentration finale de 1,65 mM et un volume équivalent de DMSO est ajouté aux tubes négatifs. La réaction est réalisée pendant 50 min à 25°C. Du tampon en excès est ajouté et l'incubation est poursuivie pendant 10 min afin de stopper la réaction. Un traitement à la protéinase K est réalisé (1 µL de protéinase K pour 15 µL de réaction) en présence de 10 mM de RVC pendant 1 h au bain-marie à 37°C. Le volume est complété à 100 µL avec de l'eau et 3 extractions phénol-chloroforme successives sont réalisées. Une précipitation à l'éthanol est ensuite réalisée et les ARN sont repris dans 28 µL d'eau milliQ pour le segment NS et 35 µL d'eau milliQ pour le segment M.

d. Rétrotranscription

A l'aide de la quantification réalisée, la même quantité d'ARN entre chaque « couple » négatif et positif est répartie dans 3 (échantillons NS) ou 4 (échantillons M) tubes.

Une rétrotranscription est réalisée après 30 min d'incubation à 50°C avec les amorces adéquates (30NS, 341NS ou 621NS pour les échantillons NS; M20, M312, M622 ou M843 pour les échantillons M) (table 4). Les rétrotranscriptions sont réalisées dans un volume final de 20 µL contenant le tampon Superscript II à une concentration 1X, 10 mM DTT, 0,75 mM dNTP et 40 U de Superscript II (ThermoFisher Scientific). En parallèle une réaction de séquençage est réalisée avec des ARN non modifiés, dans un volume final de 20 µL contenant 2 pmoles d'ARN, le tampon AMV à une concentration finale de 1X, 2,5 µM ddGTP, 5 µL G10 (1 mM dATP, 1 mM dTTP, 1 mM dCTP, 0,25 mM dGTP) et 4 U d'AMV (Promega).

Après rétrotranscription, une extraction phénol-chloroforme (v/v) est réalisée. Les réactions de

séquençage (ddGTP) sont ajoutées aux tubes des échantillons. Une extraction au chloroforme (v/v) suivie d'une précipitation à l'éthanol sont effectuées. Les culots sont ensuite repris dans 10 μ L de HiDi Formamide (Applied Biosystems). Les tubes sont chauffés 5 min à 90°C, vortexés, chauffés 2 min à 90°C puis gardés sur glace pendant 5 min. Les tubes sont centrifugés 5 min à 6 000g avant d'être chargés sur une plaque 96 puits qui est séquencée sur un séquenceur capillaire 3130xl Genetic Analyzer (Applied Biosystems).

Les résultats de séquençage sont ensuite analysés à l'aide de QuShape [346].

E. Microscopie électronique

Les complexes ARNv NS-NP ont été formés de manière similaire à la méthode décrite dans les protocoles de modification. Les tubes ont ensuite été centrifugés 10 min dans une centrifugeuse de paillasse à vitesse maximale à température ambiante. Ils ont ensuite été confiés à la plateforme de microscopie électronique de l'IBS (Institut de Biologie Structurale, Grenoble). Les échantillons ont été absorbés sur un film de carbone sur feuille de mica, colorés au silicotungstate de sodium (SST) 2%, puis transférés sur une grille de cuivre. Les images ont été capturées à une faible dose d'électrons ($<10 e^-/\text{\AA}^2$) avec des valeurs de défocus comprises entre 1,2 et 2,5 μ m avec un microscope électronique Tecnai 12 LaB6 et une caméra CCD Gatan Orius 1000.

IV | Résultats

A. Mise au point des modifications de l'ARN et optimisation de la production de la protéine NP

a. Modifications de l'ARN nu

La première étape de mon travail a été de mettre au point les modifications de l'ARN nu avec les réactifs NMIA, DMS et EDC. Utilisé seul, le NMIA ne permet pas de discriminer les régions d'ARN double brin des régions d'ARN simple brin liées à la protéine NP, ces deux types d'interaction protégeant le ribose de la modification. En effet, d'une part, l'acylation de la position 2' des riboses dépend de leur flexibilité et l'appariement de deux nucléotides diminue celle-ci et réduit donc la réactivité des nucléotides appariés [347]. D'autre part, les ARN_v se lient dans un sillon chargé positivement de la protéine NP et cette liaison impliquant les groupement phosphates des nucléotides devrait aussi protéger les riboses de l'acylation [96]. Au contraire, les bases des nucléotides fixant la NP restent sensibles à la modification par des agents chimiques [95]. Nous avons choisi d'utiliser une combinaison de deux réactifs pour tester les quatre types de bases : le DMS et l'EDC. Le DMS permet la modification des nucléotides A et C, tandis que l'EDC permet la modification des nucléotides G et U. Ainsi, le DMS et l'EDC modifieront les bases des nucléotides non appariés, que ces nucléotides soient ou non en interaction avec la protéine NP. La comparaison des données obtenues avec le NMIA, d'une part, avec celles obtenues avec le DMS et l'EDC, d'autre part, devrait permettre d'identifier les régions impliquées dans des interactions ARN-ARN et celles impliquées dans des interactions ARN-NP.

Ce travail a été réalisé sur les segments NS et M de la souche A/WSN/1933 (H₁N₁), deux ARN étudiés précédemment au laboratoire. Ces ARN ont été choisis, car peu de travaux ont porté sur les signaux d'emballage de ceux-ci. De plus, ces deux ARN étant les plus courts, cela simplifie la mise au point des protocoles, en limitant le nombre d'amorces (1 amorce pour environ 300 nucléotides).

NMIA

Dans un premier temps, les modifications de l'ARN nu ont été mises au point avec le NMIA. Ce réactif permet de modifier les riboses des quatre nucléotides libres, non impliqués dans des interactions ARN-ARN ou ARN-NP. Un protocole avait été établi précédemment au laboratoire par D. Ferhadian [98] pour lequel deux concentrations de réactif avaient été choisies : 20 mM et 50 mM final de NMIA.

Les profils de réactivité représentent l'intensité du signal du fluorophore présent en 5' des amorces utilisées pour la rétrotranscription des ARN modifiés avec 20 mM (fig. 13A) et 50 mM (fig. 13B) de NMIA. La migration dans les capillaires du séquenceur permet de séparer les ADNc obtenus en fonction de leur taille. Il est ainsi possible de déterminer les arrêts de la rétrotranscription. Plus il y a de décrochages de l'enzyme de rétrotranscription à une position, plus le signal y sera élevé. Le signal est très important au début du profil (vers la position 1000). Cette région correspond aux amorces et aux transcrits abortifs courts. Le signal dans cette région est saturé, empêchant l'analyse des 50-60 nucléotides en amont du site de liaison de l'amorce.

En appliquant ce protocole, nous avons rencontré les mêmes problèmes que D. Ferhadian. Notamment, en comparant les profils obtenus avec les concentrations de 20 mM et 50 mM, nous n'observons pas d'augmentation franche du signal malgré l'augmentation de la concentration du réactif. De plus, les pistes positives (ARN modifié avec le réactif, en bleu) et négatives (ARN non modifié, en rouge) ont été comparées l'une à l'autre (fig. 13C) et les profils sont similaires : les pics sont présents aux mêmes positions et d'intensités proches. Pour être en mesure d'exploiter les données, il faut que l'on obtienne des pics présents uniquement dans la piste positive, ou au moins d'intensité plus forte dans cette piste, qui correspondent aux nucléotides modifiés.

Plusieurs adaptations du protocole ont été apportées afin d'obtenir des différences significatives entre les profils positifs et négatifs. La reverse transcriptase du virus de la myéloblastose aviaire (AMV) initialement utilisée pour la rétrotranscription des ARN modifiés a été remplacée par l'enzyme Superscript II. Cette enzyme a un taux d'erreur plus faible que la RT AMV et une meilleure processivité. Bien que les différences entre profils n'aient pas été grandement améliorées, l'utilisation de cette enzyme a permis une meilleure reproductibilité des expériences.

Le protocole de rétrotranscription original ne comportait pas d'étape d'hybridation spécifique entre les amorces et les ARN. Cette étape étant susceptible d'améliorer grandement l'efficacité de la rétrotranscription, les conditions d'hybridation des amorces avec les ARN ont été testées. A cette fin, les amorces ont été marqués au ^{32}P , puis incubées avec les ARN pendant 10 ou 30 min à 37°C, 45°C ou 50°C. Les ARN et les amorces ont ensuite été déposés sur gel d'acrylamide (fig. 14) puis séparés par électrophorèse pendant 45 min en conditions natives. Pour toutes les amorces, on observe une augmentation de l'hybridation entre l'amorce et l'ARN (bandes les plus retardées sur les gels) lorsqu'on augmente la durée d'incubation et la température. Une étape d'hybridation à 50°C pendant 30 min a donc été ajoutée au protocole de rétrotranscription, ce qui a permis d'obtenir des profils de réactivité de meilleure qualité (fig. 13D).

DMS et EDC

Le DMS étant plus réactif que le NMIA, les problèmes posés par l'absence d'étape d'hybridation des amorces avant rétrotranscription n'avaient pas le même impact, des différences entre les échantillons

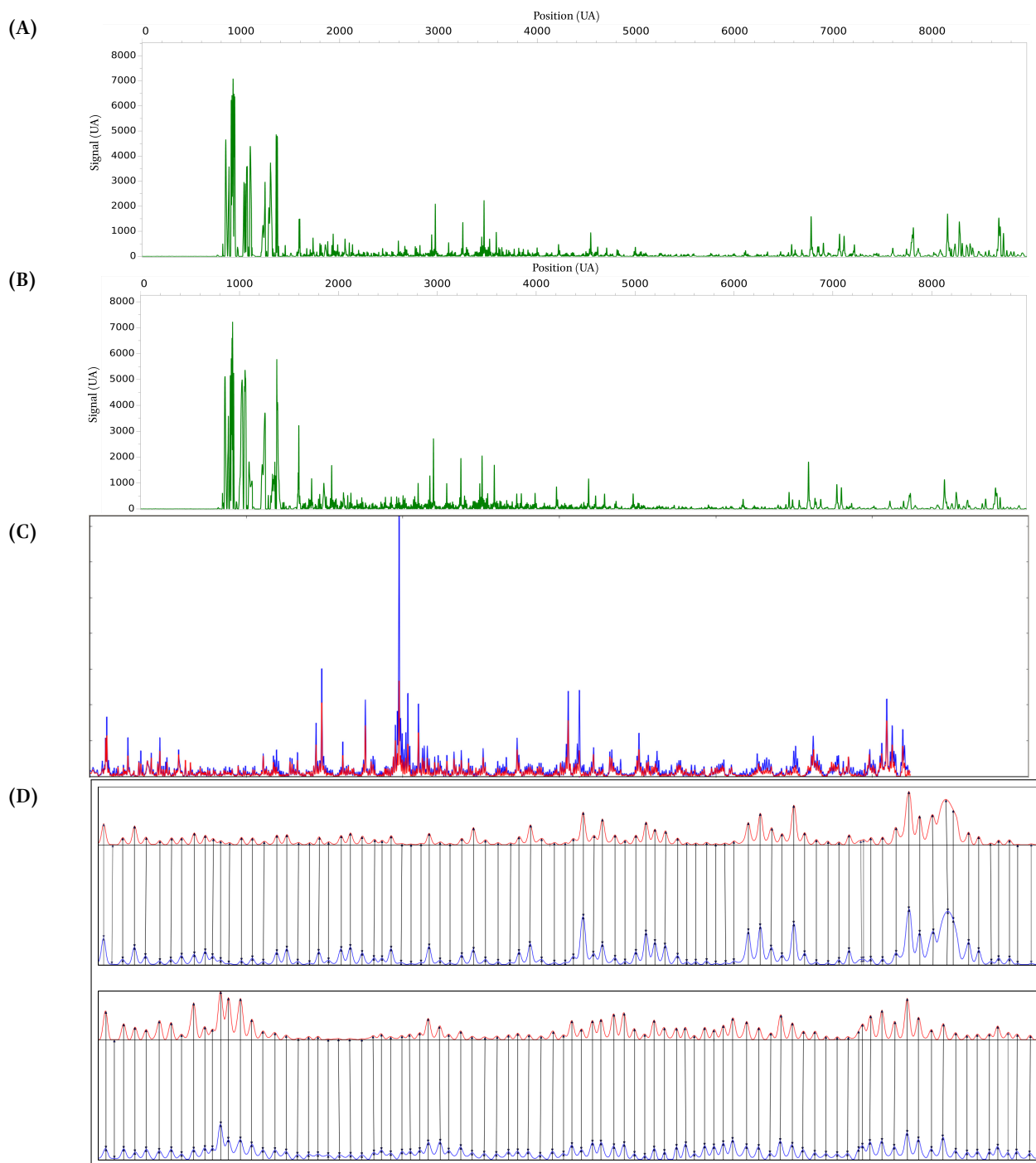


FIGURE 13. – Profils de réactivité obtenus avant et après optimisation du protocole de modification. (A) Profils obtenus avec une concentration finale de 20 mM de NMIA, (B) ceux obtenus avec 50 mM de NMIA, avec le protocole initial. (C) Superposition des traces des échantillons positifs et négatifs obtenus avec 20 mM de NMIA et le protocole initial. (D) Comparaison des traces positives (rouge) et négatives (bleues) entre l'ancien protocole (partie supérieure) et le nouveau protocole (partie inférieure).

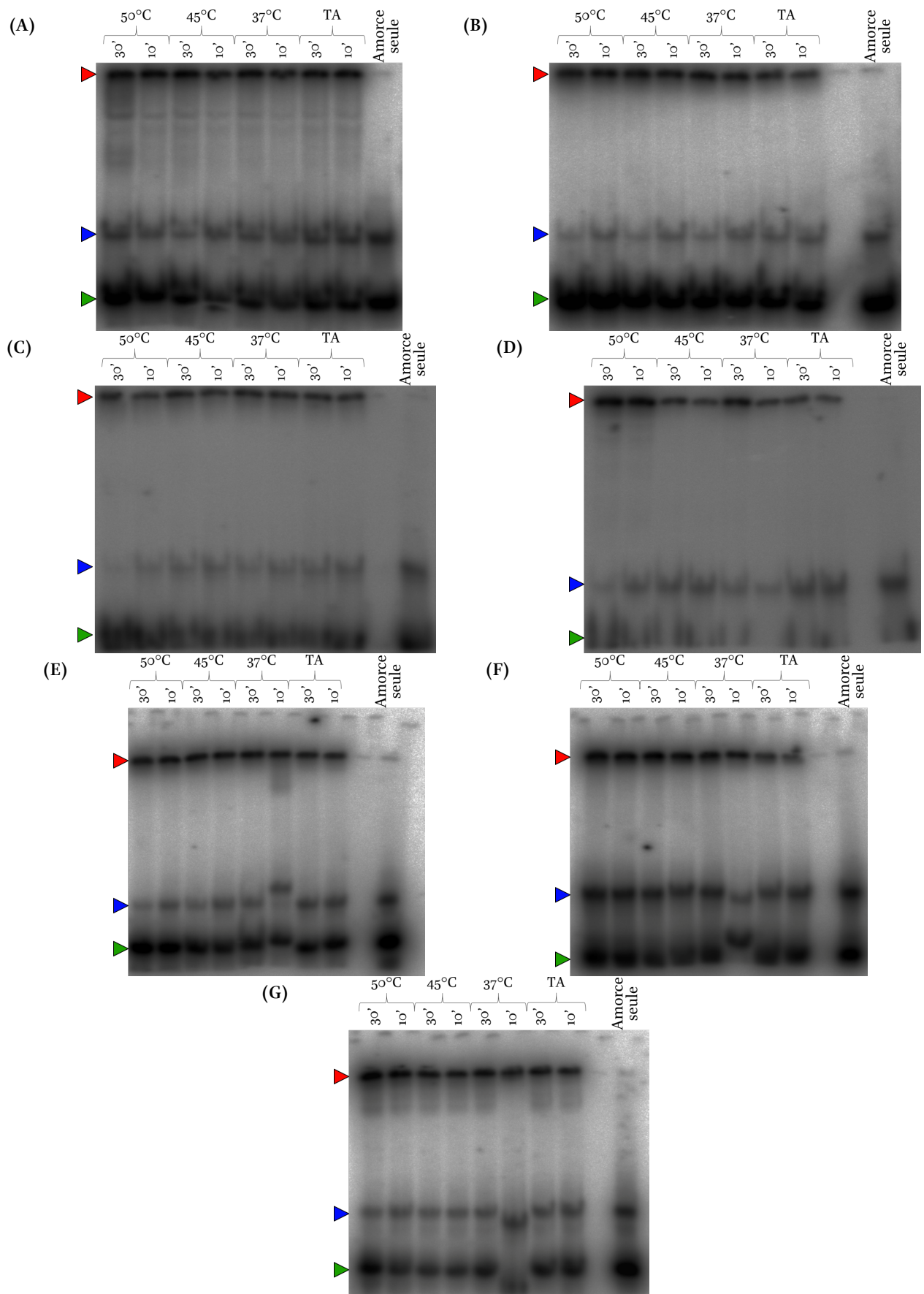


FIGURE 14. – Analyse de l'hybridation des amorces 30NS (A), 341NS (B), 621NS (C), M20 (D), M312 (E), M620 (F) et M843 (G) sur gels d'acrylamide. La température et le temps d'hybridation sont indiqués au-dessus des gels. Les flèches rouges représentent les ARN complexés aux amorces, les flèches bleues les amorces et les flèches vertes le P³² libre.

positifs et négatifs étant toujours visibles. Après la modification du protocole, il s'est cependant avéré que la quantité de DMS utilisée était trop importante, ce qui entraînait une sur-modification des ARN : plus d'une modification par ARN dans la fenêtre étudiée. Il en résulte trop d'arrêts au début de la rétrotranscription qui empêchent l'analyse des régions plus éloignées de l'amorce, voir de la totalité des données car le signal de la seule région accessible peut être saturé. Initialement, la modification au DMS était réalisée pendant 4 min à 52,5 mM final. En diminuant la concentration de DMS à 30 mM et le temps de modification à 2 min, il a été possible d'obtenir des ADNc complets à partir des ARN modifiés, tout en obtenant des profils avec des différences notables entre les profils modifiés et non modifiés (fig. 15). Comme attendu, les pics correspondant aux modifications sur l'ARN sont situés sur des A et des C.

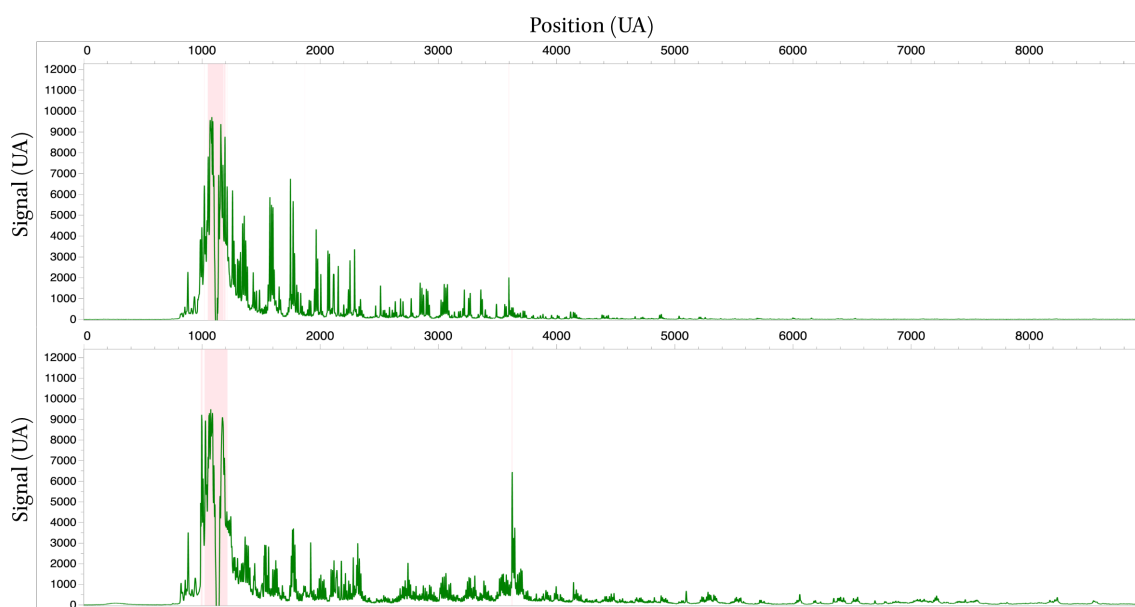


FIGURE 15. – Signal obtenu après séquençage capillaire avec l'amorce 30NS sur le segment NS avant (haut) et après (bas) ajout de l'étape d'incubation des amorces. Les zones rosées représentent les régions où le signal atteint la saturation.

L'EDC n'avait pas été utilisé auparavant au laboratoire, il n'existait donc pas de protocoles préétabli pour la modification des ARNv des virus influenza. La mise au point s'est donc basée sur la littérature existante [348, 349]. Les concentrations testées précédemment dans ces études variaient de 28 mM à 100 mM. Des concentrations de 100 mM et 200 mM final ont été testées, parmi lesquelles la concentration de 100 mM s'est révélée la plus adaptée (fig. 16).

L'ENU est un réactif ciblant les liaisons phosphodiester lorsque les groupements phosphates ne sont pas impliqués dans des interactions tertiaires ou en contact avec une protéine. Il réagit donc aussi bien avec les phosphates des nucléotides en simple brin que ceux impliqués dans des hélices. L'ENU pourrait donc être un réactif de choix pour mettre en évidence la fixation de la protéine NP aux régions hélicoïdales de l'ARN, ce qui n'est pas possible avec le NMIA. La réaction de l'ENU avec les groupements phosphates forme des phosphotriesters qui sont peu stables à 50°C en conditions légèrement alcalines. Des comparaisons sont réalisées entre des échantillons dans des conditions natives et dénaturantes. Deux concentrations de réactif

ont été testées, 75 mM et 150 mM. Les deux ont donné des résultats préliminaires encourageants sur l'ARN nu (fig. 17), mais par manque de temps, l'approche par l'ENU n'a pu être appliquée plus en détails.

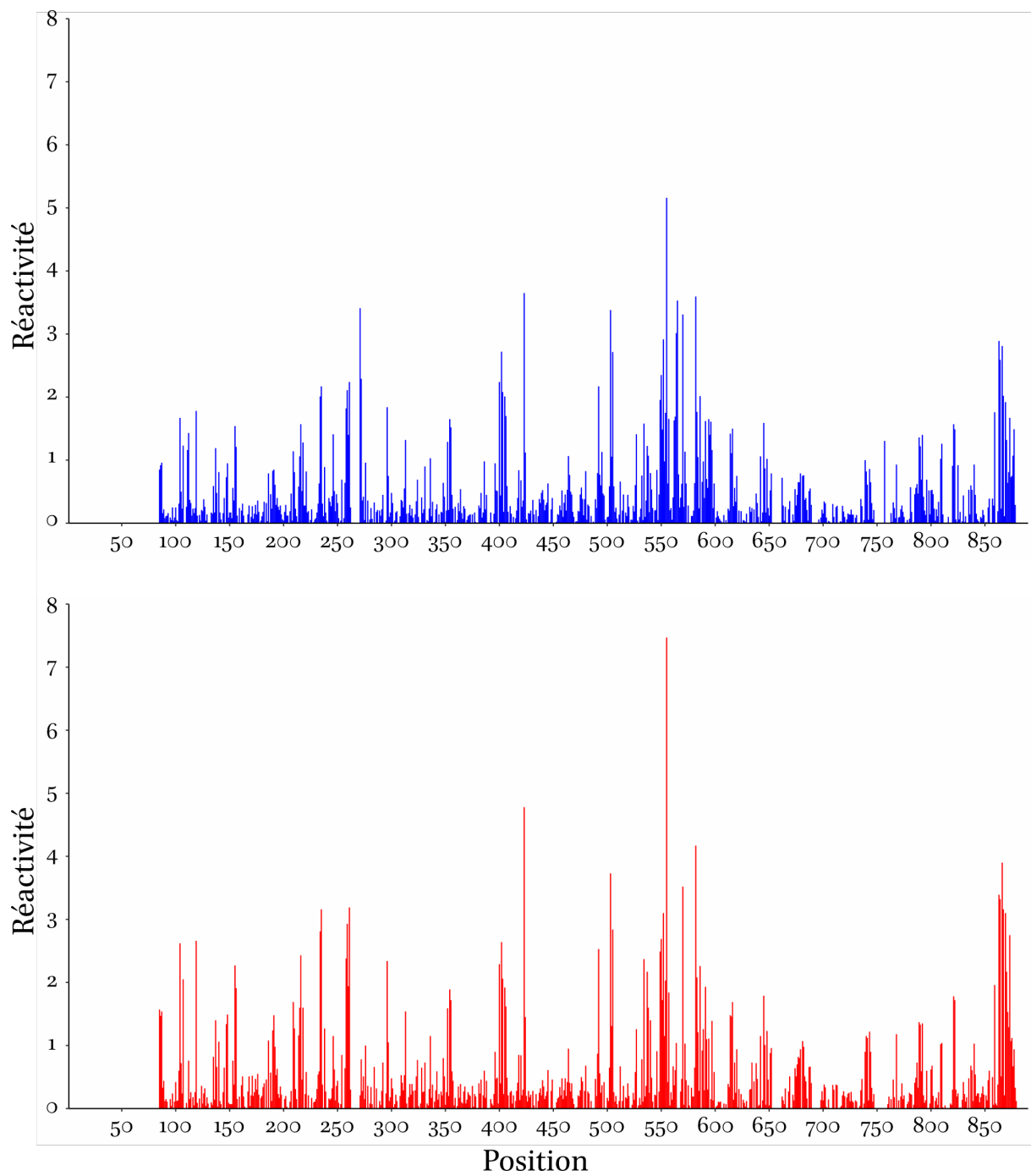


FIGURE 16. – Réactivités obtenues avec des concentrations de 100 mM d'EDC (haut) et 200 mM (bas). La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

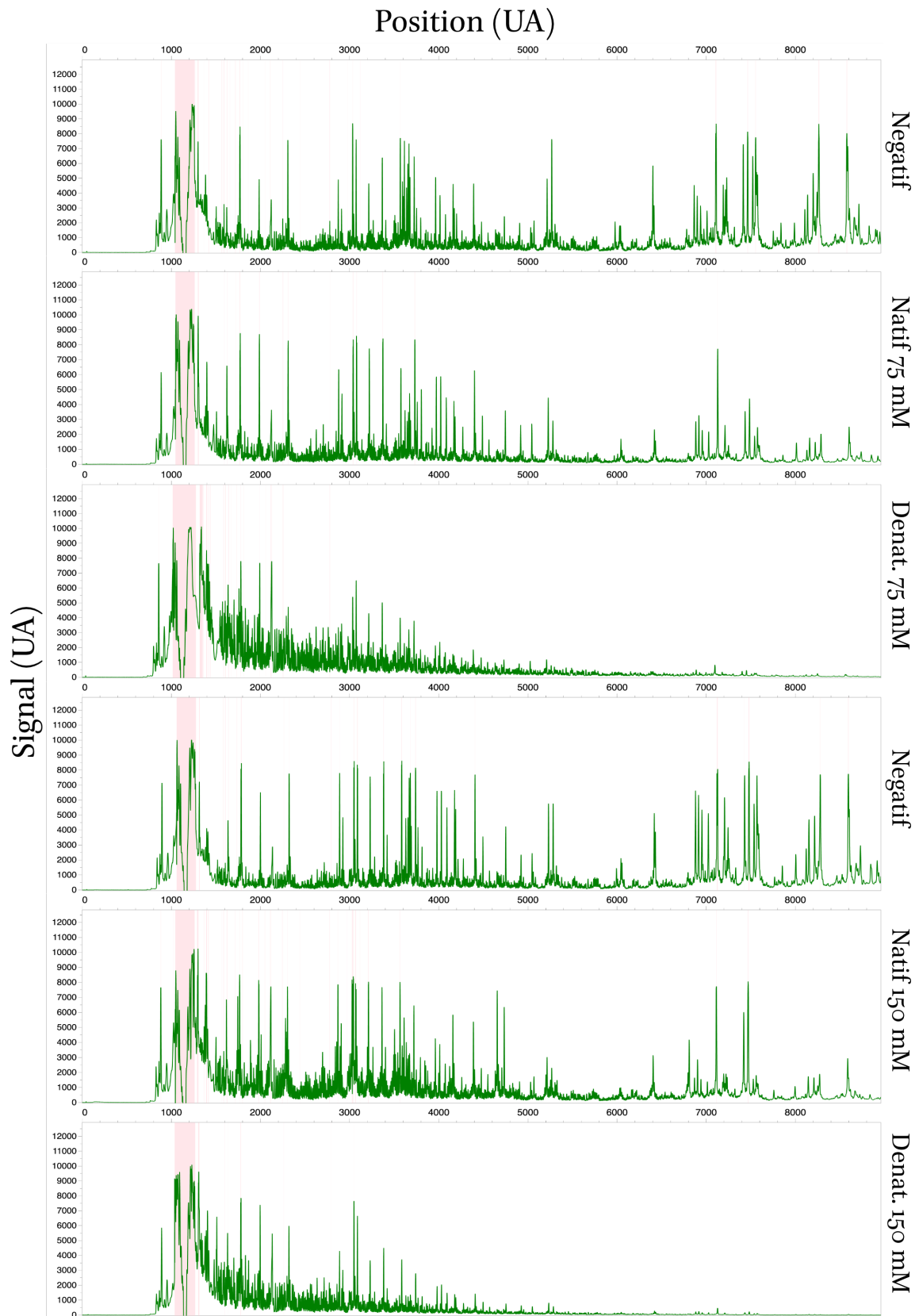


FIGURE 17. – Profils obtenus avec l'ENU dans les conditions natives et dénaturées en présence de 75 ou 150 mM d'ENU. Les zones rosées représentent les régions où le signal atteint la saturation.

b. Modifications en présence de protéine NP

Après la mise au point de la modification de l'ARN « nu » par les différents réactifs *in vitro*, nous avons donc entrepris la mise au point des protocoles de modification en présence de la protéine NP. Les protocoles de modification en absence et en présence de la protéine NP disponibles au laboratoire étaient sensiblement les mêmes à l'exception d'un temps d'incubation pour la formation des complexes ARN-NP et d'un traitement à la protéinase K après modification pour enlever la protéine qui, sinon, affecte la rétrotranscription.

Lors des premiers essais, le signal des échantillons « Comp » (complexe ARNv NS ou M- protéine NP) était beaucoup plus faible que celui de l'ARN nu (condition NoNP). Il semblait donc qu'une partie des ARN était perdue lors des expériences réalisées en présence de NP. Des fractions aliquotes ont été prélevées à deux étapes du protocole et analysées sur gel (fig. 18). Par comparaison à l'échantillon « NoNP (-) », on observe une perte d'ARN assez importante dans toutes les autres conditions. Après modification et précipitation la perte d'ARN est totale dans les conditions Comp (-) et Comp (+). Cependant, nous sommes parvenus à récupérer, dans les mêmes tubes, une fraction significative des ARN en ajoutant après la première précipitation un traitement à la protéinase K et une dernière précipitation. Nous avons émis l'hypothèse que lors de la précipitation après modification, les complexes ARNv-NP adhèrent aux tubes eppendorf et ne sont remis en solution qu'une fois le traitement protéinase K réalisé.

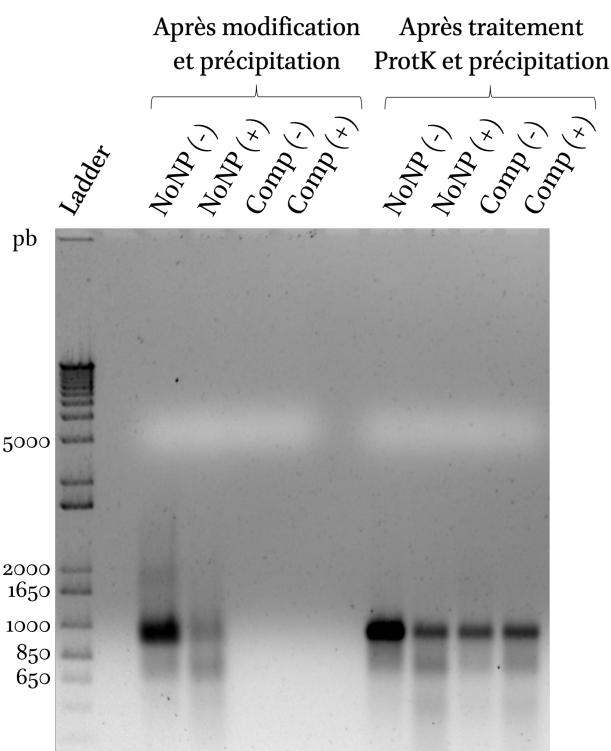


FIGURE 18. – Analyse sur gel polyacrylamide après électrophorèse pendant 45 min à 200 V dans du tampon TBE des ARN du segment NS au cours de différentes étapes de la modification.

Nous avons donc éliminé l'étape de centrifugation et de précipitation des ARN complexés à la protéine pour éviter tout impact sur la récupération des ARN. Cette étape servait à éliminer le réactif du milieu

après le temps de modification choisi et nous avons donc dû trouver des solutions pour neutraliser celui-ci, afin d'éviter une modification excessive des ARN.

Pour le NMIA, le temps de réaction de 50 min suffit à ce que la quasi-totalité du réactif soit consommé [350]. Un temps d'incubation de 10 min après un ajout d'eau en excès a été ajouté au protocole afin d'assurer la neutralisation totale du réactif. Pour le DMS et l'EDC, deux protocoles ont été envisagés. Le premier consistait à utiliser des colonnes Bio-Spin P6 (BioRad) afin de retirer le réactif, cependant les tests réalisés ont montré une perte de la quasi-totalité des complexes sur celles-ci. Le second consistait en l'utilisation de β -mercaptoéthanol pour neutraliser les réactifs, comme déjà rapporté dans la littérature [351, 352]. Ce second protocole s'est montré efficace et a donc été sélectionné.

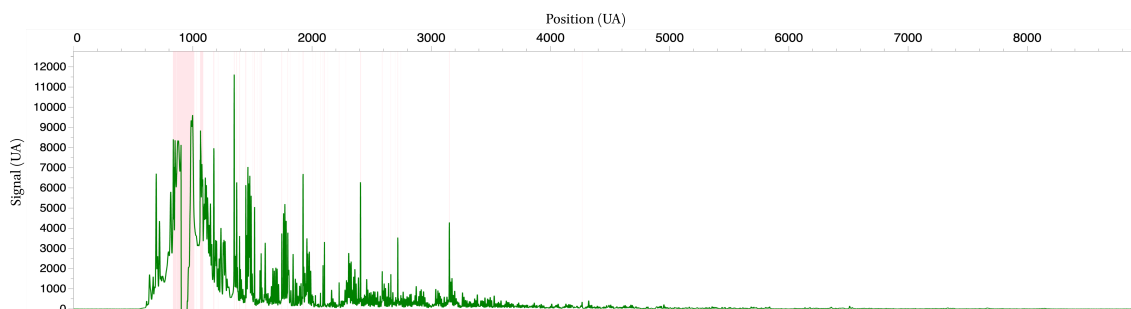
De plus, après rétrotranscription, nous avons observé une perte de signal bien plus importante et rapide pour les échantillons « Comp » (modifications des complexes ARNv-NP) que les échantillons « NoNP » (modification des ARN nus) et « ProtK » (modification des ARN après avoir retiré la protéine NP par traitement à la protéinase K) (fig. 19A). Nous travaillons avec un rapport d'une protéine NP pour 20 nucléotides, légèrement supérieur à la stœchiométrie moyenne d'une NP pour 24 nucléotides observée dans les virus [91]. Comme une protéine NP est susceptible de protéger environ 12 nucléotides [271], la moitié de l'ARN est protégée du réactif NMIA. Ainsi dans la condition « Comp », il y a environ deux fois moins de nucléotides potentiellement modifiables que dans les deux autres conditions. Il pourrait donc y avoir une sur-modification du fait d'une modification du rapport entre le nombre de nucléotides modifiables et la quantité de réactif. Différentes concentrations (0,33 mM, 1 mM, 1,65 mM, 2 mM et 3,3 mM final) en NMIA ont donc été testées, parmi lesquelles la concentration de 1,65 mM a donné des résultats (fig. 19B) extrêmement proches de ceux obtenus avec une concentration de 3,3 mM (fig. 20), indiquant que modifier la concentration en NMIA d'un facteur deux n'entraîne pas des changements significatifs de réactivité.

Au cours des tests effectués, la concentration des ARN à la fin du protocole de modification ont été régulièrement mesurée. Une certaine variabilité a pu être observée de façon occasionnelle et aléatoire, certains échantillons pouvant varier d'environ 40 ng/ μ L pour le tube positif ou négatif à 70 ng/ μ L pour l'autre tube. En suivant le protocole initial, un volume constant d'ARN de 7 μ L était prélevé, pouvant mener à l'utilisation d'une quantité d'ARN variant d'un facteur deux pour la rétrotranscription. Comme cette variation peut induire des déséquilibres dans la quantité d'ADNc obtenus dans chaque couple négatif/positif, les ARN obtenus à la fin du protocole de modification ont été systématiquement quantifiés afin de prélever une quantité de 1 pmole d'ARN de chaque échantillon pour la rétrotranscription.

c. Optimisation de la production de la protéine NP

La protéine NP de la souche A/WSN/33 (H₁N₁) a été produite dans un premier temps en suivant le protocole établi par D. Ferhadian au laboratoire [98]. Les productions ainsi obtenues avaient cependant un rendement faible, aussi bien en termes de volume (environ 500 μ L) que de concentration de la protéine NP (environ 11 μ M) pour 4 L de cultures (fig. 21A). Conformément au protocole établi

(A)



(B)

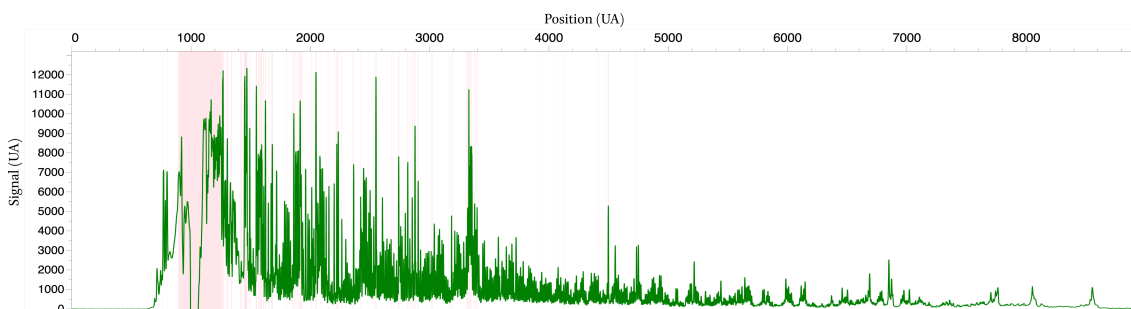


FIGURE 19. – Profils obtenus après séquençage capillaire pour la condition « Comp » avec une concentration de 3,3 mM (A) ou 1,65 mM (B) de NMIA. Les zones rosées représentent les régions où le signal atteint la saturation.

précédemment et comme le travail de cette thèse se focalisait sur la protéine trimérique, la protéine était stockée dans un tampon Tris-HCl à 150 mM de NaCl. On retrouvait donc une population hétérogène, composée majoritairement de NP trimérique et monomérique (fig. 21C). La protéine stockée sous cet état avait l'inconvénient majeur d'agréger progressivement, ayant donc une stabilité limitée dans le temps (environ 7-10 jours).

Afin d'améliorer la production de la protéine, j'ai effectué un séjour chez nos collaborateurs de l'équipe du Pr. Rob Ruigrok à l'Institut de Biologie Structurale de Grenoble. Plusieurs modifications ont été apportées au protocole utilisé jusqu'alors. A la sortie de la colonne Héparine, au lieu de dialyser les 3 échantillons les plus concentrés (1 mL/fraction), les 15 échantillons les plus concentrés sont dialysés puis concentrés à l'aide d'un centricon. Les échantillons sont ensuite passés sur une colonne Superdex 200 Increase 10/300 GL en plusieurs fois et les six fractions (0,5 mL/fraction) les plus concentrées sont dialysées contre un tampon de stockage à 50 mM en KCl. La diminution de la concentration en sel permet d'obtenir une population homogène monomérique (fig. 21D), plus stable et moins sujette à l'agrégation. Les rendements sont également plus élevés : 1,5 mL à environ 30 μ M (fig. 21B).

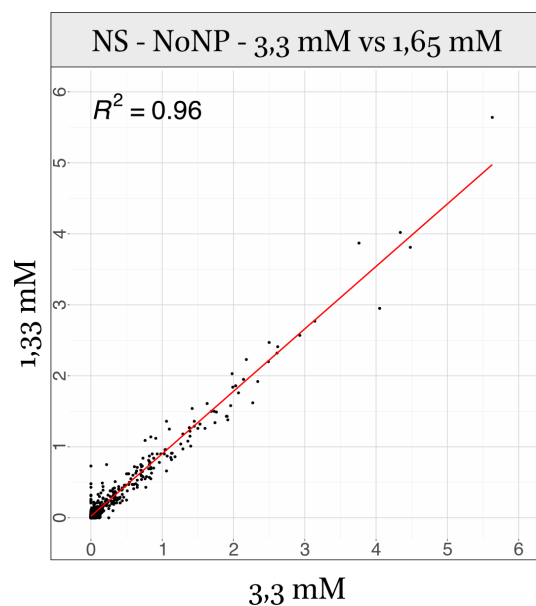


FIGURE 20. – Corrélation des réactivités de l'ARNv NS en condition « NoNP » obtenues avec des concentrations 1,65 mM et 3,3 mM de NMIA.

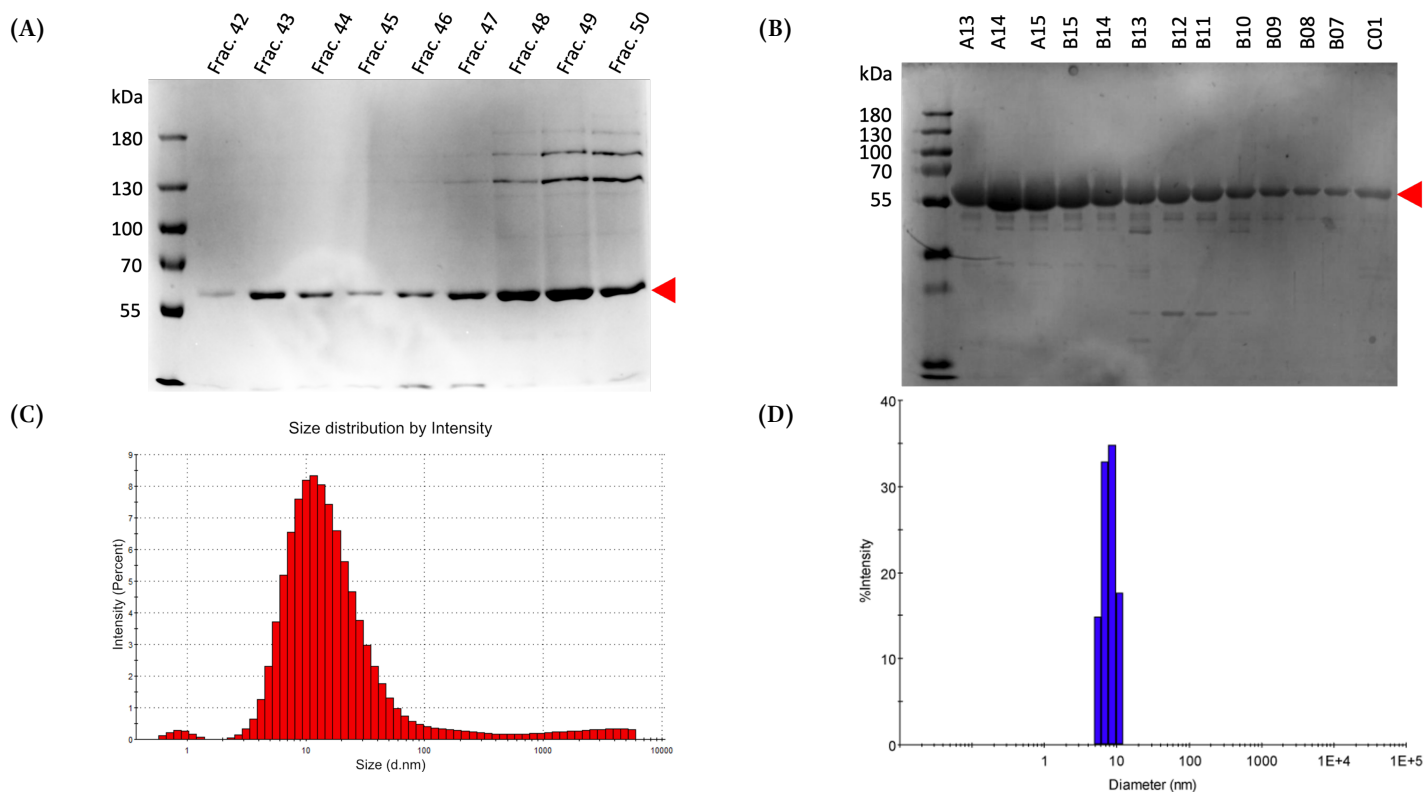


FIGURE 21. – Gels SDS-PAGE 12% présentant les fractions obtenues après la dernière élution de la protéine NP avec l'ancien (A) et le nouveau (B) protocole. Les triangles rouges indiquent la bande correspondant à la protéine NP. Les distributions de tailles des particules en fonction de l'intensité diffusée obtenues par DLS montrent une population non homogène avec l'ancien protocole (C) mais une population bien plus homogène avec le nouveau protocole (D).

B. Comparaison des réactifs NMIA avec DMS et EDC

L'utilisation du NMIA comme sonde pour l'étude de la structure de l'ARN ne permet pas de faire la discrimination entre les régions d'ARN double brin et les régions liées à la protéine, les deux protégeant le ribose. Ainsi, il est nécessaire d'utiliser d'autres sondes, le DMS et l'EDC afin de réaliser des modifications ciblant la base nucléique, qui ne sera protégée du réactif que par les interactions ARN-ARN. Afin de pouvoir comparer les données de ces différentes sondes, il est nécessaire que deux points soient vérifiés : tout d'abord que des régions similaires soient retrouvées comme réactives avec le NMIA et le DMS-EDC et ensuite que les réactivités globales soient comparables.

Pour les réactifs DMS et EDC, par manque de temps, seules des données partielles ont pu être obtenues pour le segment NS. Il est cependant possible d'analyser les données obtenues sur la fin du segment NS nu (630-783) avec un seul réplicat par réactif avec l'amorce 30NS. Les réactivités globales obtenues d'une part avec le NMIA et d'autre part avec le DMS et l'EDC ont tout d'abord été comparées. Pour pouvoir comparer le NMIA avec les données du DMS et de l'EDC, les réactivités des nucléotides A et C des expériences menées avec le DMS ont été rassemblées avec les réactivités des nucléotides G et U des expériences menées avec l'EDC. La corrélation entre les données NMIA et DMS-EDC a ensuite été calculée (fig. 22). Le R^2 est faible, ce qui montre qu'il y a une certaine divergence entre les réactifs NMIA et DMS/EDC. De plus, la droite de tendance ne passe pas par zéro, mais par une valeur d'environ 0,3. Il semblerait donc que les réactivités obtenues avec les réactifs DMS et/ou EDC soient décalées par rapport à celles obtenues avec le NMIA. Les réactivités obtenues avec l'EDC semblent également plus faibles qu'avec le DMS. Un F-test a été réalisé entre les valeurs des deux conditions. Les variances sont significativement différentes avec une p-value de 0,0006615. Afin d'établir si les valeurs DMS-EDC sont plus élevées que les valeurs NMIA, un test de Wilcoxon unilatéral a été utilisé qui donne une p-value de 2,107e-05 pour cette hypothèse, confirmant que les données DMS-EDC sont significativement supérieures aux données NMIA.

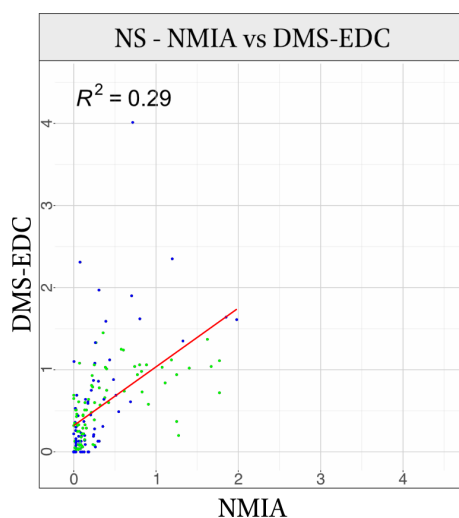


FIGURE 22. – Corrélations des données obtenues avec le NMIA et celles obtenues avec les réactifs DMS et EDC. Les points correspondant aux nucléotides A et C sont colorés en bleu tandis que les points correspondant aux nucléotides G et U sont colorés en vert.

Cependant, des différences au niveau des valeurs de réactivité ne signifient pas forcément que les modifications sont situées à des positions différentes de l'ARN ou qu'elles soient incompatibles. Les réactivités obtenues avec les réactifs SHAPE, dont le NMIA, sont généralement classifiées en trois catégories : les nucléotides non ou peu réactifs (entre 0 et 0,4) qui correspondent, sur de l'ARN nu, aux nucléotides appariés, les nucléotides moyennement réactifs (entre 0,4 et 0,8) qui peuvent se trouver aussi bien dans des boucles empilées que dans une structure secondaire peu stable ou en équilibre avec une structure alternative et les nucléotides fortement réactifs (supérieurs à 0,8) qui sont non appariés et exposés au solvant. Nous avons réparti les réactivités obtenues d'une part avec le NMIA et d'autre part avec le DMS/EDC dans ces trois catégories que nous avons notées respectivement 0, 1 et 2 et nous les avons représentées dans un graphique à barres (fig. 23).

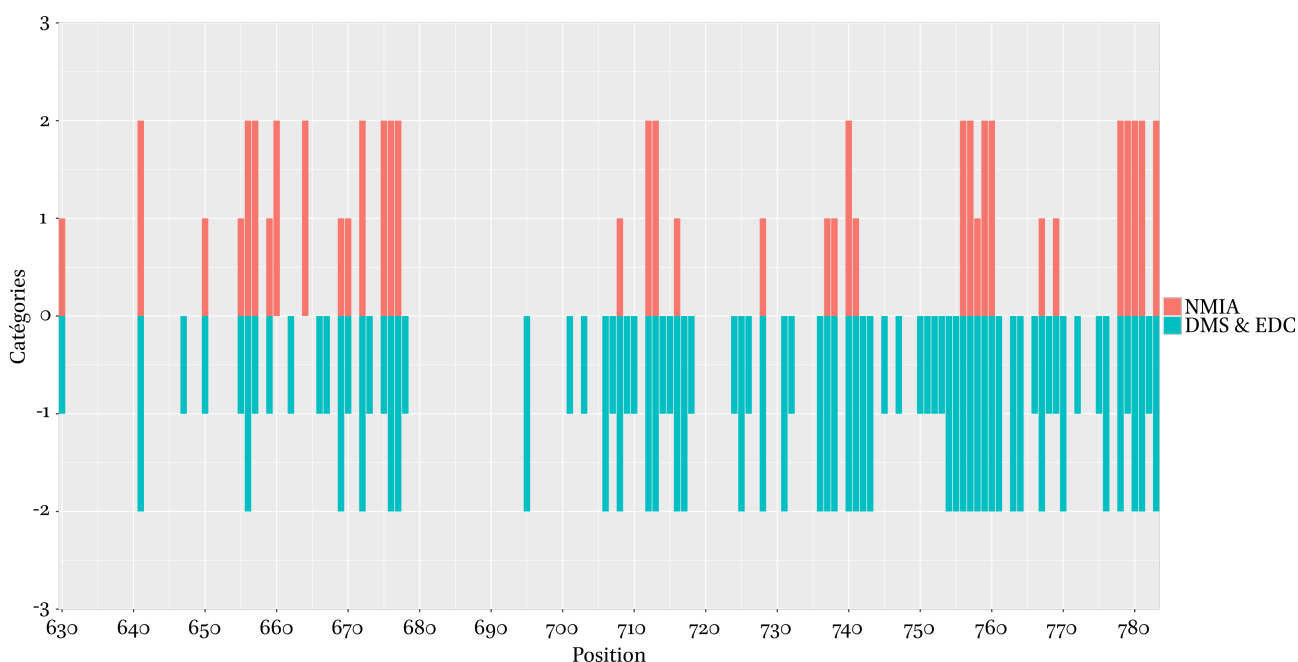


FIGURE 23. – Réactivités par position catégorisées. Une valeur de 0 désigne un nucléotide avec une réactivité comprise entre 0 et 0,4; une valeur de 1 un nucléotide avec une réactivité comprise entre 0,4 et 0,8; une valeur de 2, un nucléotide avec une réactivité supérieure à 0,8. La partie supérieure (rouge) représente les réactivités obtenues avec le NMIA, la partie inférieure (bleue) avec le DMS et l'EDC.

Tous les nucléotides réactifs au NMIA sont réactifs avec le DMS ou l'EDC, à l'exception d'un seul. On retrouve cependant plus de nucléotides réactifs au DMS ou à l'EDC : 36 nucléotides réactifs avec le NMIA contre 77 avec le DMS ou l'EDC. Quinze nucléotides sont moyennement réactifs avec le NMIA contre 37 avec le DMS ou l'EDC. Pour les nucléotides fortement réactifs, on en retrouve 21 avec le NMIA contre 40 avec le DMS et l'EDC. On retrouve un facteur de 2 dans tous les cas, il y a donc globalement plus de nucléotides réactifs au DMS et à l'EDC, répartis d'une façon égale entre les nucléotides moyennement réactifs et les nucléotides fortement réactifs. La région analysée ici se répartit équitablement entre nucléotides modifiables au DMS (47,4% de A et C) et ceux modifiables à l'EDC (52,6% de G et U).

Afin de déterminer si les différences observées avec le NMIA sont spécifiques à l'un des réactifs ou non,

nous avons comparé d'une part les réactivités des nucléotides A et C obtenues avec le NMIA et le DMS et d'autre part les réactivités des nucléotides G et U obtenues avec le NMIA et l'EDC (fig. 24). Pour les deux réactifs, on observe un R^2 assez faible et une ordonnée à l'origine de la droite de tendance supérieure à 0. Cette dernière coupe l'axe des ordonnées à une valeur d'environ 0,25 pour les données DMS et d'environ 0,37 pour les données EDC. Les réactivités globales semblent donc plus élevées, ce qui pourrait expliquer les différences visibles, sans que celles-ci soient dues à un seul des réactifs. Un test de variance F-test entre les données NMIA-DMS et NMIA-EDC pour les nucléotides modifiés donne des p-values de $1,365e-07$ pour la comparaison NMIA-DMS et de $0,1076$ pour la comparaison NMIA-EDC. Pour NMIA-DMS, un test de Wilcoxon unilatéral est donc réalisé et donne une p-value de $0,03059$, montrant que les valeurs obtenues au DMS sont effectivement plus élevées qu'avec le NMIA. Comme ni les données obtenues avec le NMIA, ni celles obtenues avec l'EDC ne suivent une loi normale (p-value de $3,698e-10$ et $0,001233$ en faisant un test de normalité de Shapiro-Wilk), un test de Wilcoxon unilatéral est également réalisé, donnant une p-value de $0,0002581$ et montrant que les valeurs obtenues avec le NMIA sont significativement inférieures à celles obtenues avec l'EDC. Il y a donc bien une différence significative entre les données obtenues avec le NMIA et celles obtenues avec le DMS et l'EDC, ces dernières étant plus élevées qu'avec le NMIA.

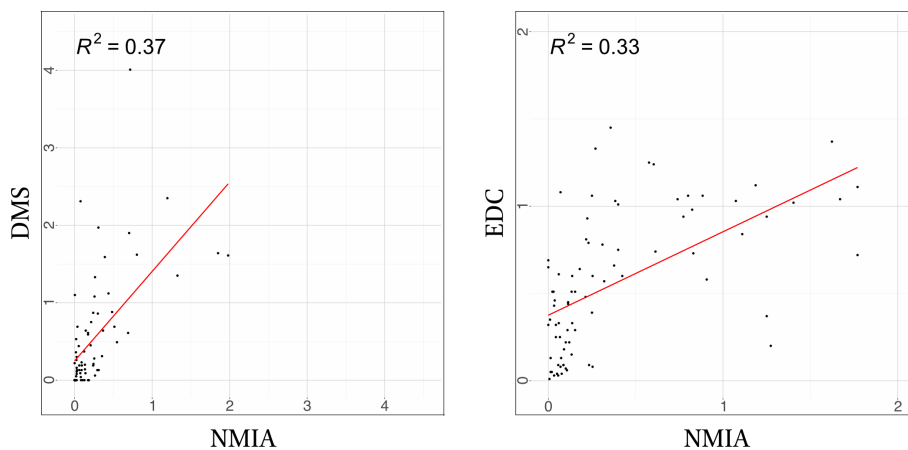


FIGURE 24. – Corrélations entre les données obtenues avec NMIA et DMS pour les nucléotides A et C (à gauche) et entre les données obtenues avec le NMIA et l'EDC pour les nucléotides G et U (à droite).

Bien que des différences subsistent, elles sont donc principalement dues à une réactivité globale plus élevée avec les réactifs DMS et EDC. Les nucléotides qui sont réactifs avec le DMS ou l'EDC sans être réactifs au NMIA sont majoritairement adjacents aux nucléotides réactifs avec le NMIA et le DMS/EDC et les profils partagent de nombreuses similarités. Après obtention des données définitives avec le DMS et l'EDC, une étape de normalisation serait souhaitable.

Une normalisation a donc été effectuée en utilisant la formule suivante :

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (IV.1)$$

Avant normalisation, les données DMS et NMIA sont assez proches sur les différents paramètres à l'exception du maximum, plus élevé dans les données DMS. Le premier quartile, la médiane et le

	Avant normalisation				Après normalisation			
	NMIA	DMS	EDC	DMS+EDC	NMIA	DMS	EDC	DMS+EDC
1er quartile	0.045	0.04	0.08	0.115	0.02273	0.009975	0.05517	0.02868
Médiane	0.1367	0.135	0.21	0.41	0.06902	0.033666	0.14483	0.10224
Moyenne	0.3231	0.3843	0.3595	0.5511	0.16318	0.095832	0.24796	0.13743
3ème quartile	0.3742	0.475	0.5775	0.8325	0.18897	0.118454	0.39828	0.20761
Maximum	1.98	4.01	1.45	4.01	1	1	1	1

TABLEAU 5. – Répartition des données avant et après normalisation.

troisième quartile sont plus élevés avec l'EDC qu'avec les deux autres réactifs. Par conséquent, on observe des différences sur ces mêmes paramètres entre les données NMIA et les données DMS et EDC. Après la normalisation, les données DMS + EDC sont plus proches du NMIA, cependant les données DMS s'éloignent des données NMIA (table 5).

En regardant la répartition des données selon un diagramme en violon, on peut voir qu'avant normalisation (fig. 25A), la répartition des données est effectivement très différente entre les données NMIA et DMS + EDC. Le DMS possède une répartition proche de celle du NMIA, et seul l'EDC diffère : plus de nucléotides réactifs, avec une plus grosse proportion de nucléotides ayant des réactivités supérieures à 0,5. Après la normalisation (fig. 25B), les données DMS + EDC ont une répartition bien plus proche des données NMIA. Les données DMS ont un profil différent mais cohérent : la proportion de nucléotides avec une réactivité faible voire nulle augmente et devient supérieure à celle des données NMIA. Ce réactif ne pouvant modifier que les bases A et C qui représentent environ la moitié des nucléotides de la région étudiée, il est donc cohérent qu'on retrouve un plus grand nombre de nucléotides peu ou pas réactif. Les données EDC en revanche adoptent un profil encore plus différent de toutes les autres données. La proportion de nucléotides peu ou pas réactifs est la plus faible et la proportion de nucléotides reste assez similaire peu importe la valeur de la réactivité. Des nucléotides qui ne sont pas censés être modifiés par le réactif (A et C) ont donc une réactivité assez importante. Il en découle donc une différence marquée entre les données EDC et les données obtenues avec le NMIA et le DMS. L'utilisation de l'EDC dans les conditions actuellement établies afin de combiner les jeux de données DMS et EDC semble donc difficile, un travail de mise au point supplémentaire pourrait être requis.

Afin de voir si les données NMIA et DMS-EDC pourraient, même dans nos conditions actuelles, permettre d'obtenir des prédictions de structure qui soient assez proches, les réactivités obtenues avec le NMIA ou le DMS + l'EDC ont été utilisées afin de prédire des modèles de structures secondaires. On retrouve une région très structurée quand on utilise les données NMIA (fig. 26A), tandis que la région ne comporte que très peu de structures secondaires avec les données DMS-EDC (fig. 26B). La comparaison des prédictions à l'aide de RNAStructViz confirme bien la forte divergence de ces deux conditions (fig. 26C), ce qui confirme bien, qu'en l'état, la combinaison des réactifs DMS et EDC n'est pas encore utilisable.

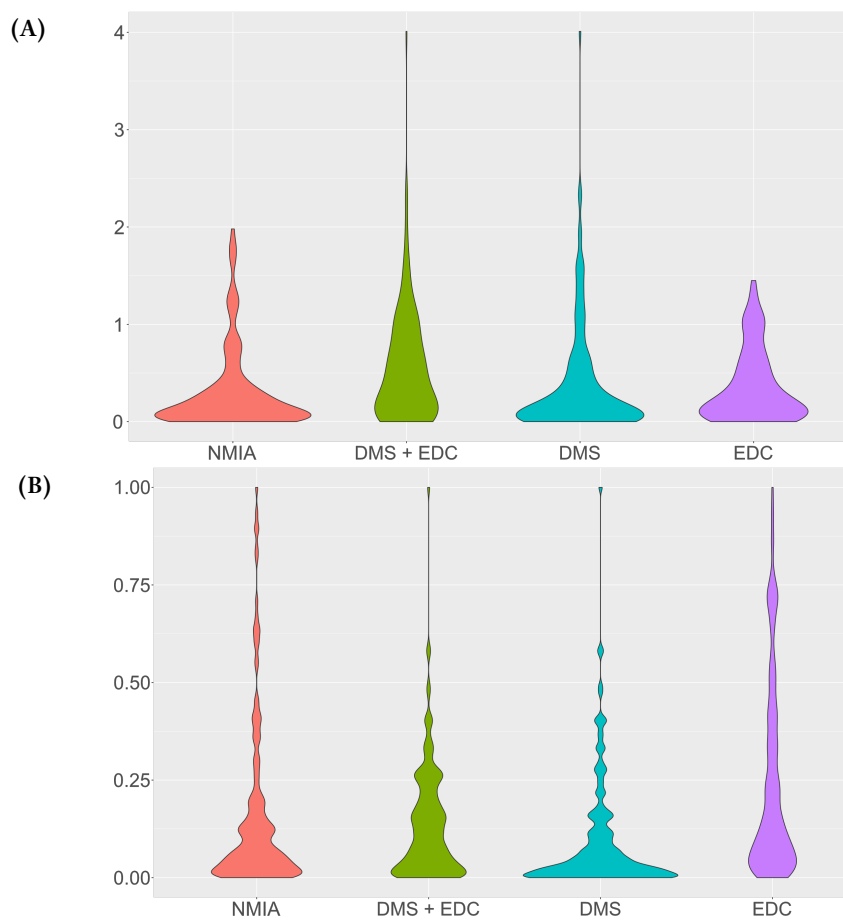


FIGURE 25. – Représentation avec des diagrammes en violon des données obtenues avec les réactifs NMIA, DMS et EDC avant (A) et après (B) normalisation.

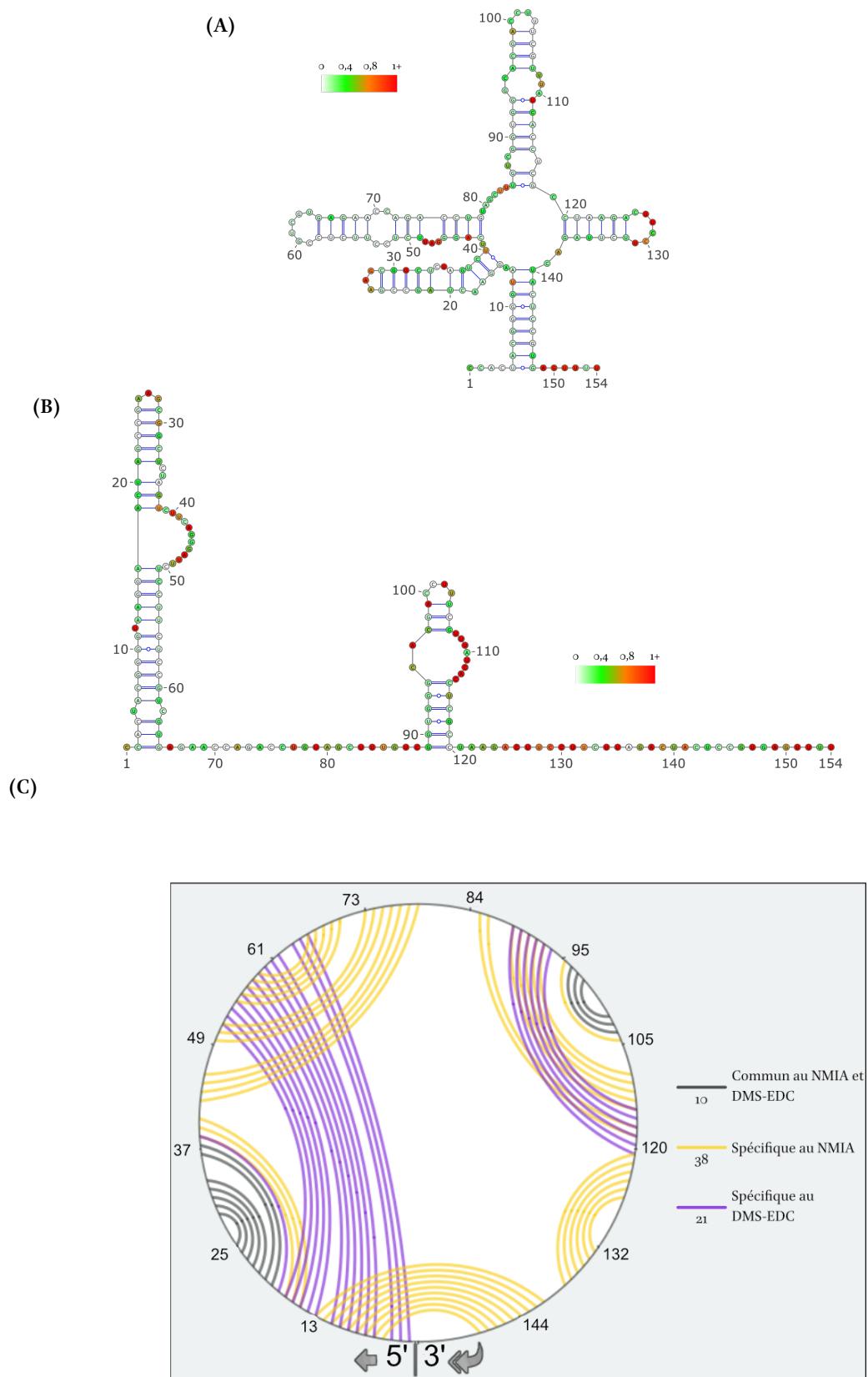


FIGURE 26. – Comparaison des prédictions de structure obtenues avec le NMIA et avec le DMS-EDC. Les structures les plus stables obtenues avec le NMIA (A) et la combinaison DMS-EDC (B) sont représentées, ainsi que la comparaison des prédictions de structure avec RNAstructViz.

C. Cartographie des ARN NS et M

a. Corrélations des données obtenues

Au terme des optimisations réalisées, un triplicat technique a pu être obtenu pour chaque condition étudiée (« NoNP », « Comp » et « ProtK ») avec chacun des ARN NS et M. D. Ferhadian ayant pu obtenir des données préliminaires au cours de sa thèse, celles-ci seront comparées aux résultats obtenus.

Les réactivités obtenues pour chaque position ont été représentées selon un code couleur : noire pour les réactivités comprises entre 0 et 0,4, orange pour celles comprises entre 0,4 et 0,8 et rouge pour celles supérieures à 0,8. Cela a été réalisé pour les conditions « NoNP », « ProtK » et « Comp » des segments NS (fig. 27) et M (fig. 28). Les réactivités obtenues avec le segment NS couvrent une gamme de valeurs plus étendue que celles obtenues avec le segment M : on retrouve dans le segment NS des valeurs très élevées comme le pic aux alentours de la position 500 pour le segment NS dans la condition « NoNP » et les régions aux environs des positions 500 et 870 dans la condition « ProtK » (fig. 27). Pour le segment M, les plus hautes valeurs de réactivité ne sont pas aussi élevées (fig. 28). La situation est cependant plus proche entre les deux segments dans la condition « Comp ». On voit une diminution du nombre de nucléotides fortement réactifs ($> 0,8$) et une diminution globale des réactivités les plus élevées à l'exception de quelques positions (~ 570 et ~ 880 pour le NS, ~ 670 , ~ 900 et ~ 1010 pour le M).

Les corrélations entre les différents réplicats de la condition « NoNP » montrent que globalement, les réplicats du segment NS ont un R^2 plus faible (annexe 1) (0,55 à 0,62) que ceux du segment M (annexe 2) (0,79 à 0,82). Plusieurs hypothèses peuvent expliquer cette différence. Comme les plus fortes réactivités sont atteintes pour le segment NS (maximum de 11,26 pour NS contre 5,765 pour M), il peut exister une plus grande variabilité parmi les valeurs élevées (supérieures à 1) pour le segment NS que pour le segment M. Une seconde hypothèse repose sur l'observation des traces de séquençage. En effet, il existe deux régions au sein de l'ARN NS où l'enzyme de rétrotranscription se décroche fréquemment, aussi bien dans les pistes contrôles que les pistes avec réactif ($\sim 261-271$, $\sim 748-752$). Bien que ces régions aient été exclues pour le calcul du R^2 , la présence de structures induisant des décrochages de l'enzyme dans ces régions pourrait fausser l'analyse avec QuShape, notamment les étapes de normalisation ou de compensation de la perte du signal. Un exemple de structure pourrait être les G-quadruplex, cependant une expérience de SHALiPE [353] a été conduite pour en détecter mais n'a pas donné de résultat probant (résultats non montrés).

La stabilité des structures des ARN NS et M avant et après la fixation et l'enlèvement de la protéine NP a été analysée. A l'aide des contraintes SHAPE, des structures ont été prédites à l'aide de RNAstructure (version 6.4) [354], et la structure la plus stable, c'est-à-dire qui possède l'énergie libre la plus faible, a été choisie pour l'analyse. Pour le segment NS, l'énergie libre passe d'une valeur de -486,4 kJ/mol dans la condition « NoNP » à -500,9 kJ/mol dans la condition « ProtK ». Dans le cas du segment M, on passe d'une valeur de -569,3 kJ/mol à -559,3 kJ/mol entre ces deux conditions. La baisse de l'énergie libre pour le segment NS est marginale et pour le segment M, on observe même une augmentation, marginale elle

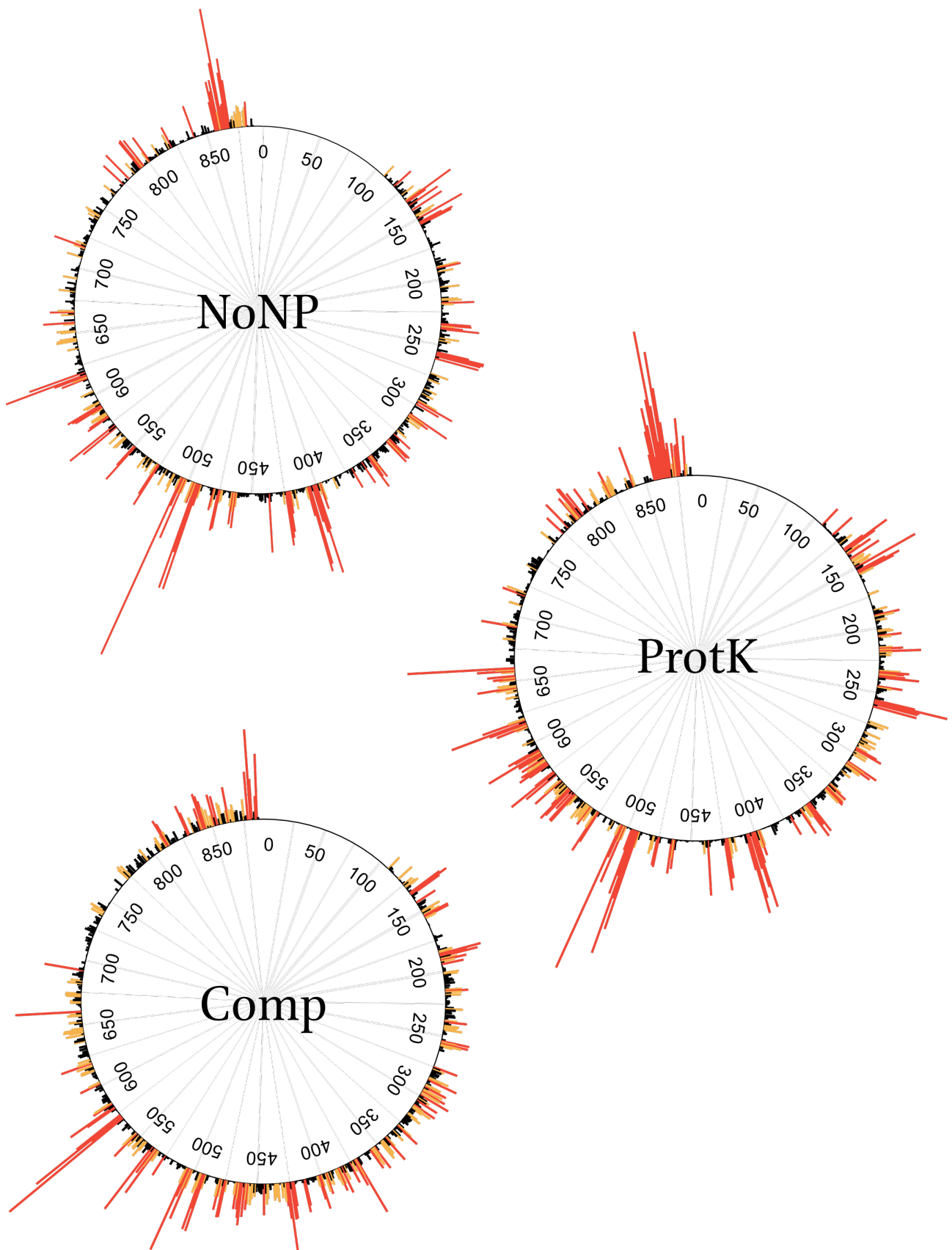


FIGURE 27. – Profils des réactivités pour les conditions « NoNP », « ProtK » et « Comp » sur le segment NS. Les couleurs correspondent aux échelles de réactivité : de 0 à 0,4 en noir, de 0,4 à 0,8 en orange et plus de 0,8 en rouge. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

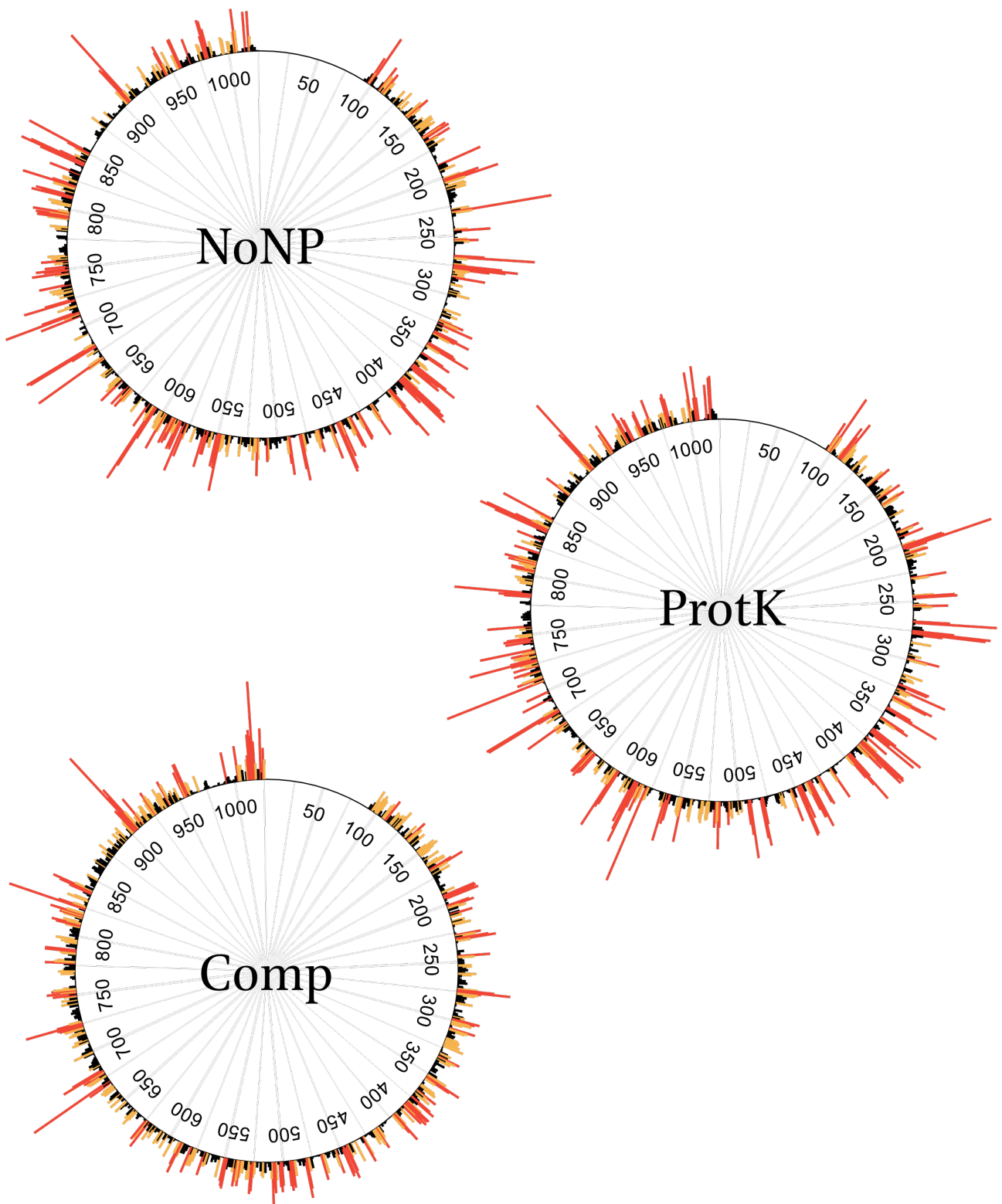


FIGURE 28. – Profils des réactivités pour les conditions « NoNP », « ProtK » et « Comp » sur le segment M. Les couleurs correspondent aux échelles de réactivité : de 0 à 0,4 en noir, de 0,4 à 0,8 en orange et plus de 0,8 en rouge. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

aussi, de l'énergie libre. Il ne semble donc pas que la fixation de la NP permette, après son retrait, d'augmenter la stabilité de la structure de l'ARN, ce qui correspond à ce qu'avait observé D. Ferhadian au cours de sa thèse [98].

La comparaison des coefficients de corrélation entre les différentes conditions apporte d'autres informations. Que ce soit pour l'ARN NS ou l'ARN M, les coefficients de corrélation sont généralement les plus faibles pour la condition « NoNP ». Ceux-ci augmentent un peu dans la condition « ProtK » (pour le segment NS principalement, le réplicat n°2 du segment M comporte quelques valeurs élevées qui diminuent mécaniquement le coefficient de corrélation). Pour les deux segments, les corrélations les plus fortes sont atteintes dans la condition « Comp ». Ces résultats montrent qu'il y a une variation assez importante des réactivités en absence de la protéine NP et que plusieurs structures différentes pourraient co-exister. La présence de la protéine NP, diminue fortement la variation des réactivités, ce qui suggère que la fixation de la protéine diminue la diversité des structures de l'ARN pour en favoriser une seule. Lorsque la NP est retirée, les coefficients de corrélation diminuent, ce qui montre une augmentation de la variation des réactivités et signifierait que quelques structures différentes pourraient co-exister. La diversité structurale resterait cependant plus faible que dans la condition « NoNP ». Certaines interactions ARN-ARN induites par la protéine NP pourraient donc subsister même après l'élimination de la protéine.

b. Activité chaperon de l'ARN de la protéine NP

Afin de déterminer si la protéine NP possède une activité chaperon ou non, nous avons comparé les réactivités des conditions « NoNP » et « ProtK ». Ainsi, il sera possible l'impact structural de la protéine NP sur le maintien de la structure de l'ARN. La NP servant de squelette à l'ARN, il est attendu que des modifications importantes de la structure vis-à-vis de la condition « Comp » soient visibles. En premier lieu, nous avons voulu déterminer si la structure obtenue après avoir retiré la protéine NP serait proche, ou non de la structure de l'ARN qui n'a pas été en contact avec la protéine NP.

L'analyse des R^2 obtenus dans les conditions « NoNP » et « ProtK » (fig. 29), montre que ceux-ci sont très nettement plus élevés que ceux obtenus précédemment par D. Ferhadian [98], 0,8 contre 0,17 pour le segment NS et 0,71 contre 0,16 pour le segment M. Ainsi les diverses mises au point réalisées nous permettent d'obtenir des données nettement plus reproductibles que précédemment. Par ailleurs les courbes de tendances obtenues par D. Ferhadian montraient que la réactivité globale des nucléotides était plus élevée dans la condition « ProtK », ce qui n'est pas le cas ici. Ces données suggèrent que la liaison puis l'enlèvement de la protéine NP a un impact définitif assez limité sur la structure des ARN NS et M.

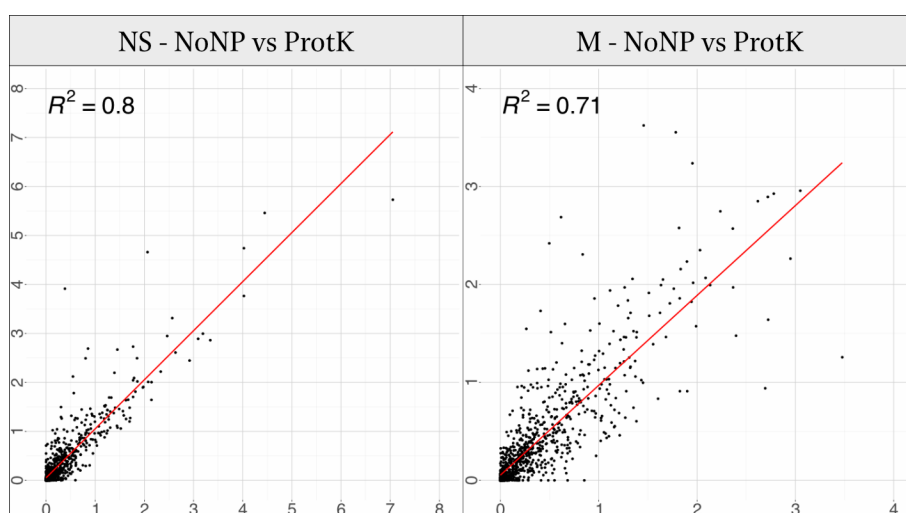


FIGURE 29. – Corrélations entre les données « NoNP » et « ProtK » pour les segments NS et M. Les données de la condition NoNP sont en abscisse, celles de la condition ProtK sont en ordonnée.

Pour compléter cette première analyse, nous avons comparé les structures des ARN NS et M prédites en utilisant les données de SHAPE obtenues dans les conditions « NoNP » et « ProtK » comme contraintes expérimentales (figs. 30 à 33). En effet, les protéines chaperons de l'ARN facilitent son repliement en une structure plus stable qui subsiste lorsque la protéine est enlevée.

Les prédictions de structure des ARN NS et M dans les conditions « NoNP » et « ProtK » ont pu être générées à l'aide de RNAstructure [354] et mises en forme à l'aide de VARNA (version 3.93) [355]. Les structures secondaires représentées sont considérées comme les plus stables. La couleur noire représente les nucléotides pour lesquels la réactivité n'a pas pu être déterminée et correspondent soit à des régions non couvertes en raison de la position des amorces, soit à des régions où l'on observe des arrêts fréquents de la reverse transcriptase aussi bien dans les séquences que dans les ARN modifiés ou non. Les couleurs

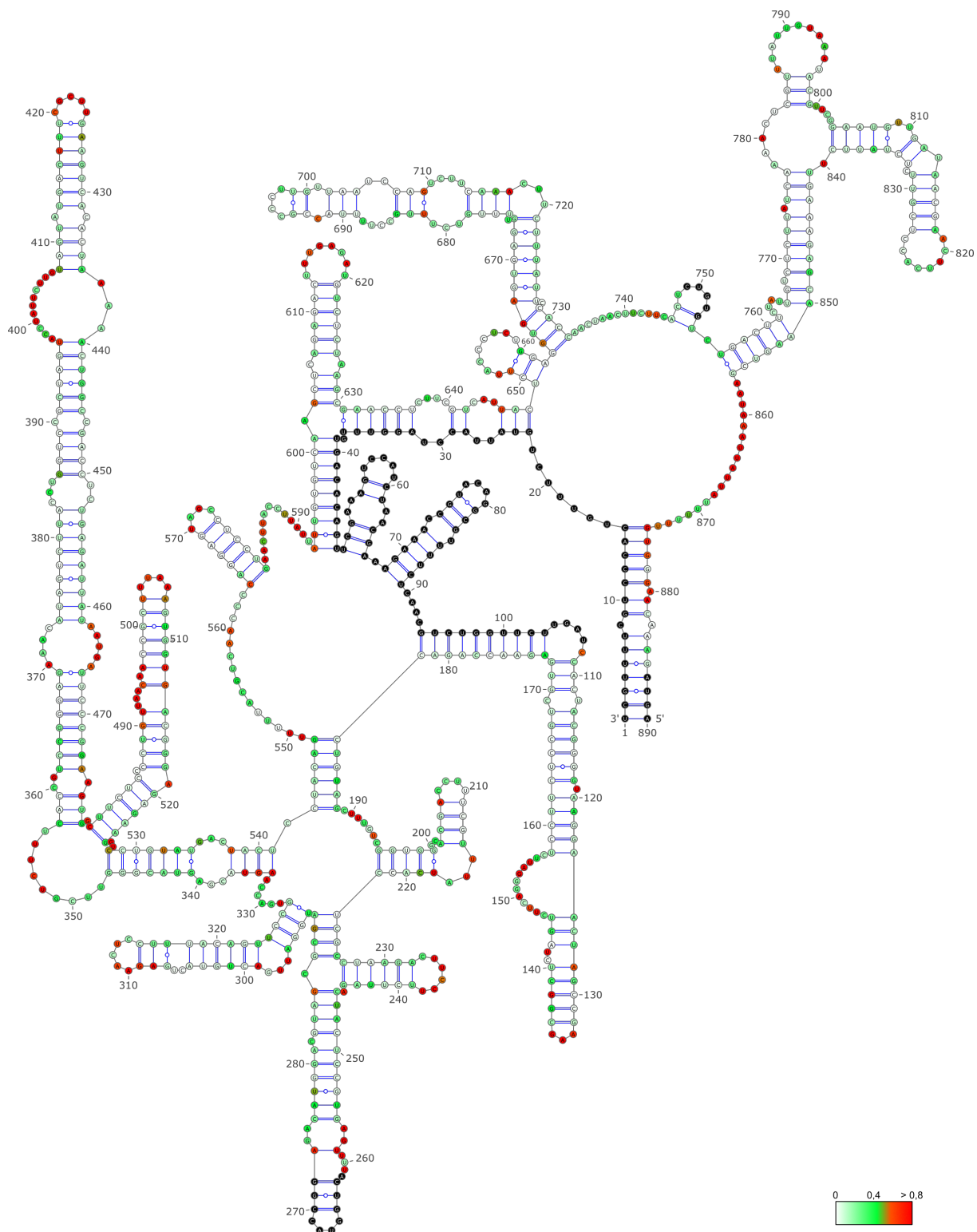


FIGURE 30. – Prédiction de la structure du segment NS dans la condition « NoNP ». Les couleurs représentent les valeurs de réactivité. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

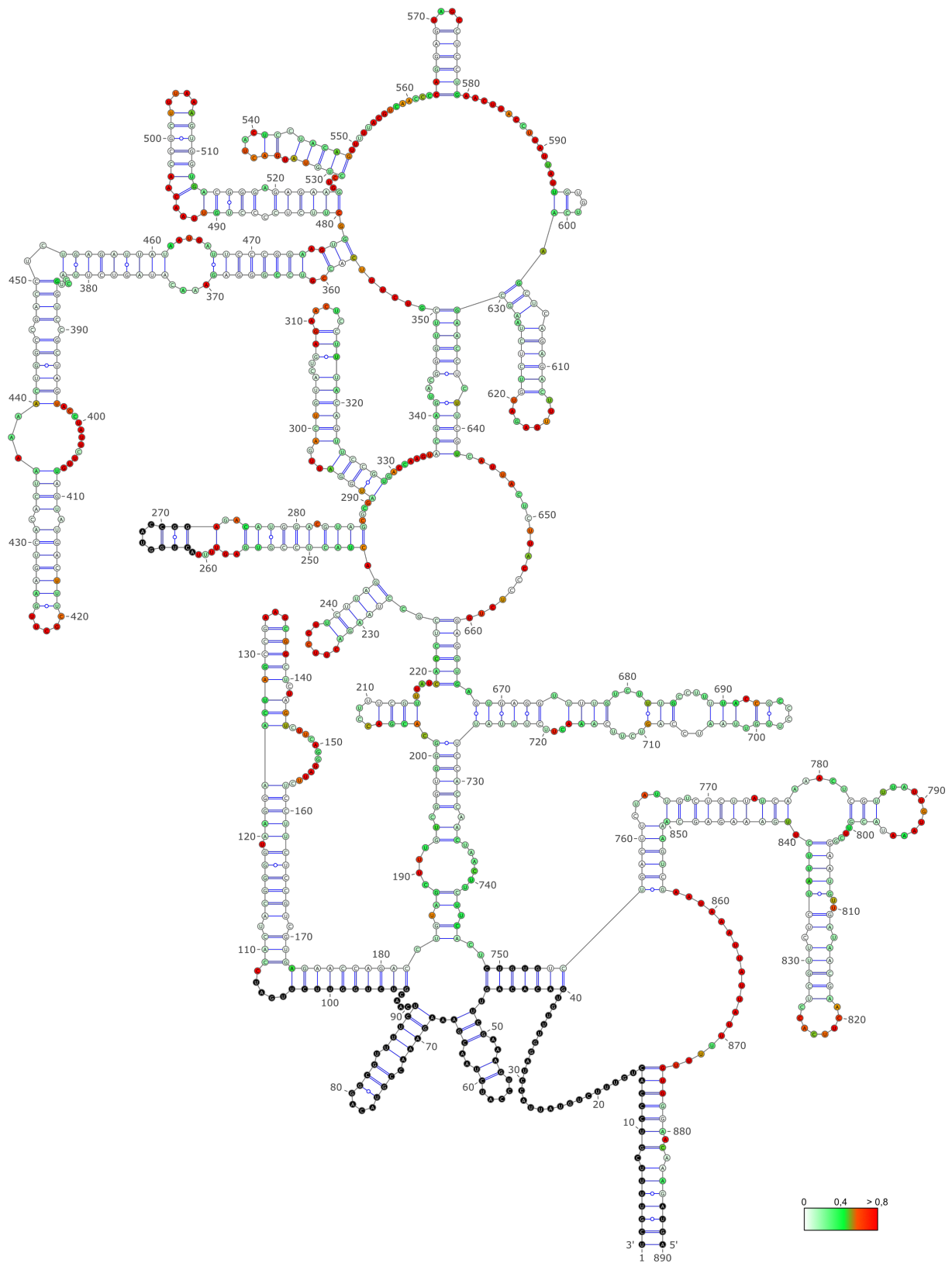


FIGURE 31. – Prédiction de la structure du segment NS dans la condition « ProtK ». Les couleurs représentent les valeurs de réactivité. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

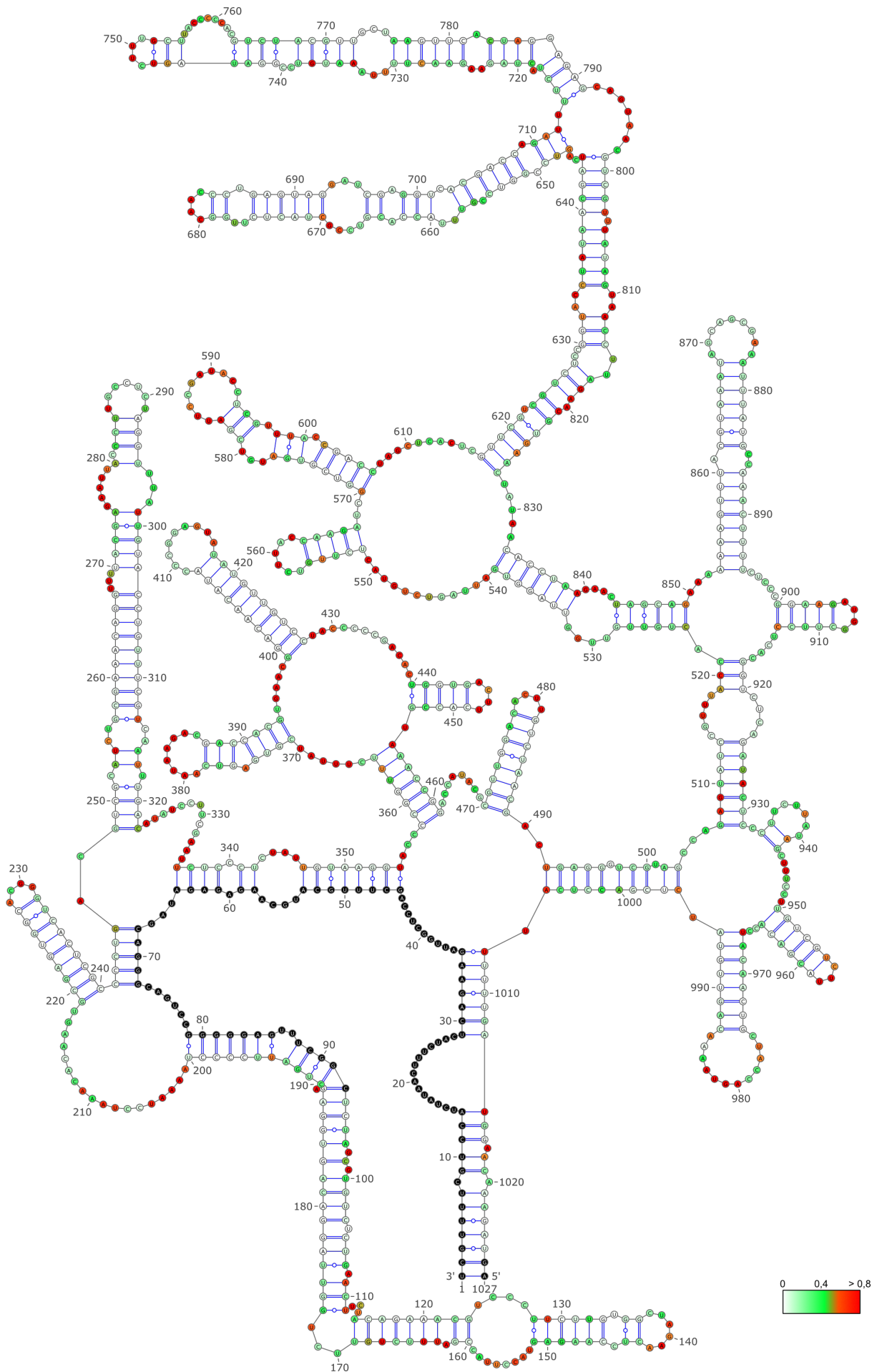


FIGURE 32. – Prédiction de la structure du segment M dans la condition « NoNP ». Les couleurs représentent les valeurs de réactivité. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

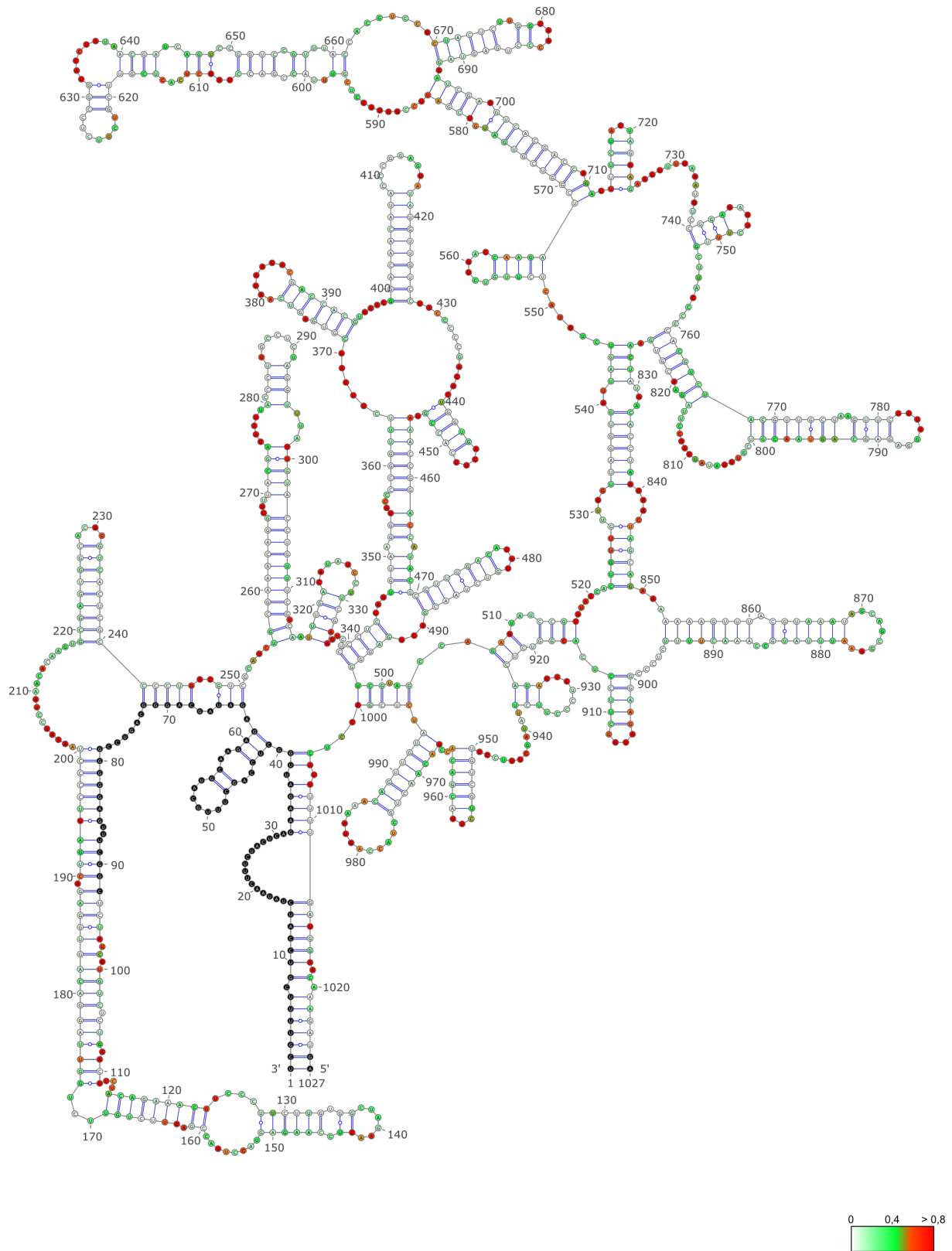


FIGURE 33. – Prédiction de la structure du segment M dans la condition « ProtK ». Les couleurs représentent les valeurs de réactivité. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

représentent les réactivités selon un gradient allant du blanc (0), à vert (0,4), orange (0,5) et rouge (> 0,8).

On observe pour les deux segments des ARN très structurés que ce soit dans les conditions « NoNP » ou « ProtK ». En comparant les conditions « NoNP » (fig. 30) et « ProtK » (fig. 31) pour l'ARN NS, même si les structures secondaires paraissent à première vue très différentes, on observe la conservation de motifs structuraux locaux. On observe également la conservation de certains motifs structuraux pour le segment M, entre les structures obtenues avec les données « NoNP » (fig. 32) et les données « ProtK » (fig. 33). Cela est également visible avec l'aide d'une représentation sous forme d'arc avec le *webservice* R-chie et RNAdbee [356–359], dans laquelle la partie inférieure représente les interactions différentes entre les conditions « NoNP » et « ProtK ». Pour le segment NS (annexe 3) et pour le segment M (annexe 4), on n'observe des changements de structure que dans quelques régions. Afin d'analyser plus en détails les structures prédites par RNAstructure, le logiciel RNAStructViz [360] est utilisé.

Pour le segment NS, on retrouve environ 4 interactions à longue distance uniquement dans la condition « NoNP » et 4 interactions à longue distance propres à la condition « ProtK » (fig. 34A). On retrouve un total de 275 nucléotides impliqués dans des interactions pour la condition « NoNP » et 265 pour la condition « ProtK » pour le segment NS; 309 et 302 pour le segment M. On retrouve donc un nombre proche de nucléotides impliquées dans ces interactions et les changements consistent majoritairement en des relocalisations vers des régions qui étaient également en interaction dans l'autre condition. Au sein des nucléotides impliqués dans ces interactions, on retrouve pour les conditions « NoNP » et « ProtK » du segment NS 123 et 122 nucléotides dans des interactions G-C (environ 45%), 124 et 118 dans des interactions A-U (environ 45%) et 28 et 25 dans des interactions G-U (environ 10%). Pour le segment M, on retrouve 138 et 140 nucléotides dans des interactions G-C (environ 45%), 138 et 132 nucléotides dans des interactions A-U (environ 45%) et 33 et 30 nucléotides dans des interactions G-U (environ 10%). Il n'y a donc pas de biais induit par la fixation de la protéine NP sur les types d'interaction entre nucléotides, que ce soit pour le segment NS ou le segment M. Les quelques différences entre les conditions « NoNP » et « ProtK » sont à priori dues à des réarrangements entre des régions déjà en interaction avant le contact avec la protéine NP.

Pour le segment M, on retrouve 4 interactions à longue distance propres à la condition « NoNP » et 4 propres à la condition « ProtK », ainsi que plusieurs interactions à plus courte distance propres à chacune des conditions (fig. 34B). La plupart des interactions à longue distance se forment soit à proximité d'une interaction commune aux deux conditions, soit à proximité d'une interaction se formant dans l'autre condition (au milieu de la région 475-554 par exemple). On ne trouve ainsi que deux et une interaction(s) à longue distance propres aux conditions « NoNP » et « ProtK ». Pour la condition « NoNP », on retrouve 39% de nucléotides qui sont spécifiques à cette condition, contre 21% pour le segment NS. Dans la condition « ProtK », on en retrouve 37% pour le M contre 18% pour le NS. Le segment M voit donc plus de réarrangements structuraux s'effectuer que le segment NS, mais ils ont un impact moindre sur la structure globale de l'ARN, car il s'agit essentiellement de réarrangements locaux.

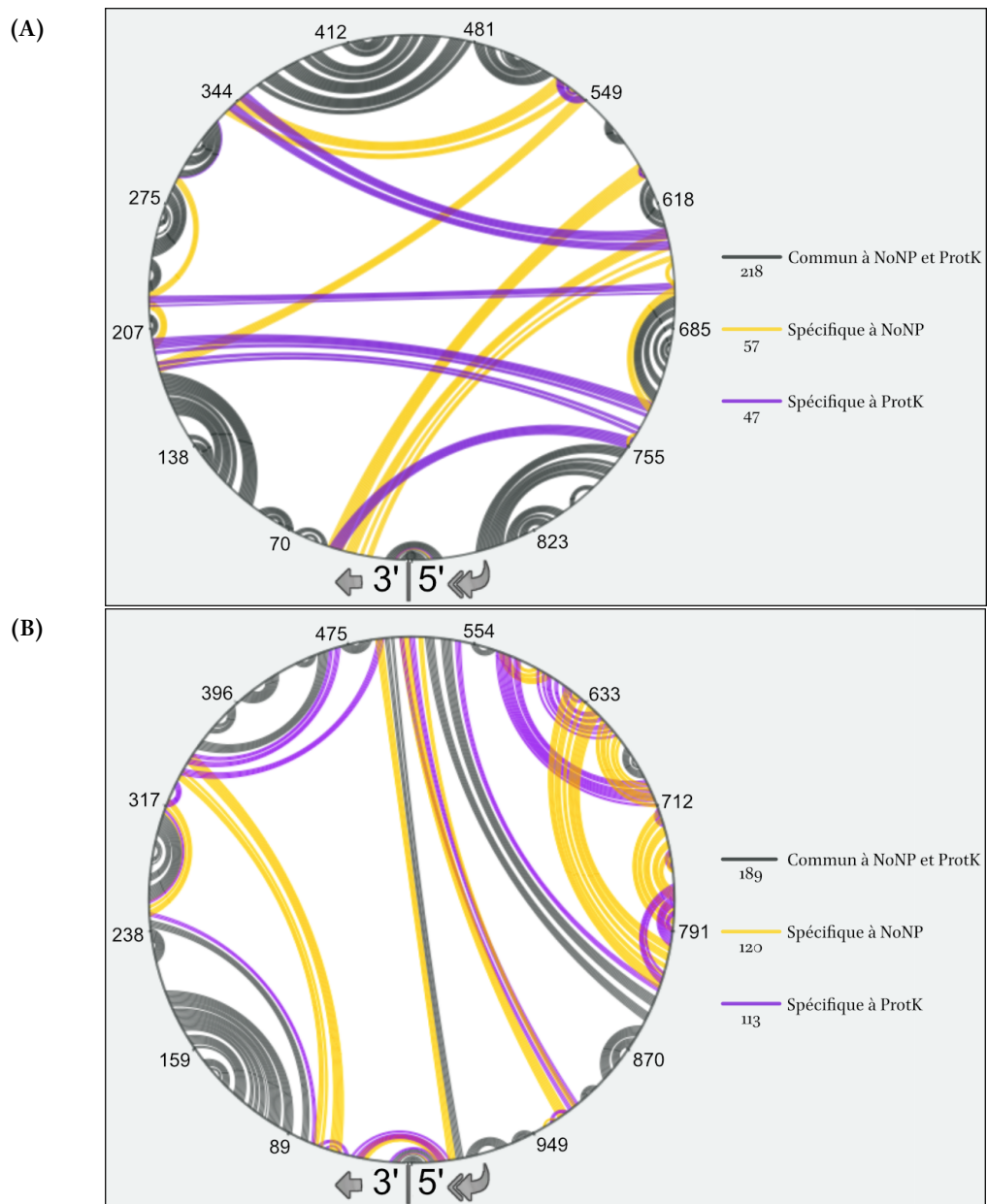


FIGURE 34. – Représentation des différences entre les structures obtenues dans les conditions « NoNP » et « ProtK » pour le segment NS (A) et M (B). Les arcs représentent les interactions entre les différentes régions de l'ARN selon les conditions : en noir pour les interactions que l'on retrouve dans les deux conditions « NoNP » et « ProtK », en jaune pour les interactions que l'on ne retrouve que dans la condition « NoNP » et en violet pour les interactions que l'on ne retrouve que dans la condition « ProtK ». Les chiffres dans la légende correspondent au nombre de nucléotides en interaction dans chaque condition. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

La comparaison des structures prédites avec celles obtenues précédemment par D. Ferhadian [98] montre que la plupart des tiges boucles sont retrouvées. Pour le segment NS dans sa condition « NoNP » (annexe 5), les structures les plus longues (structures n°1, 7, 8) semblent partager le même *pattern* pour les nucléotides fortement réactifs, que l'on retrouve aux mêmes positions. Les tiges boucles plus courtes peuvent également partager beaucoup de similarités (structure n°5), être plus réactives dans les données plus récentes (structures n°3, 4 et 6) ou bien moins réactives (structure n°2). Dans la condition « ProtK » (annexe 6), On observe un schéma un peu différent avec moins de structures similaires en termes de positions réactives (structures n°1, 4), plus de structures plus réactives dans les données récentes (structures n°2, 3, 5, 6 et 7) et autant de structures avec une réactivité moindre dans les données récentes (structure n°8).

On retrouve pour la condition « NoNP » du segment M une majorité de structures avec des réactivités proches dans les données de D. Ferhadian et les plus récentes (annexe 7) (structures n°1, 3, 5 et 7), et des proportions équivalentes de structures plus réactives dans les données récentes (structures n°2 et 4) et de structures moins réactives (structures n°6 et 8). Dans la condition « ProtK » on a, contrairement au segment NS, plus de structures similaires entre les données de D. Ferhadian et les données les plus récentes (annexe 8) (structures n°4, 5, 6, 8, 9, 10), une structure de plus étant plus réactive dans les données récentes (structures n°1, 3, 7) et deux structures plus réactives avec les données les plus récentes (structures n°2, 11).

Bien que l'on puisse observer dans les résultats de D. Ferhadian une différence entre les données « NoNP » et « ProtK » (notamment *via* les coefficients de corrélation), on ne retrouve pas de différences marquées entre les structures prédites avec les données les plus récentes. Ainsi, les différences observées dans les données de D. Ferhadian étaient potentiellement dues à une baisse de la réactivité au niveau global dans la condition « ProtK », diminuant le coefficient de corrélation sans pour autant modifier drastiquement les prédictions de structure.

Les variations des réactivités entre chaque condition considérées comme biologiquement significatives ont été déterminées selon la méthode suivante : la différence de réactivité pour chaque nucléotide entre chaque condition a été calculée. Un *cut-off* de 0,15 a été appliqué afin de retirer les différences trop faibles. Le pourcentage de changement de réactivité a ensuite été calculé (par exemple pour une comparaison NoNP-Comp, (valeur NoNP - valeur Comp)/(valeur Comp)) et un nouveau *cut-off* a été appliqué afin de retirer les pourcentages de moins de 40% d'augmentation ou de diminution. Les différences ont ensuite été illustrées sur les structures secondaires de chaque ARN prédites à l'aide des données « NoNP ».

En analysant les figures obtenues, on n'observe qu'un nombre assez restreint de nucléotides pour lesquels on observe un changement important pour le segment NS (fig. 35). De plus, ces changements sont en majorité des augmentations de réactivité au sein de boucles ou des diminutions de réactivité dans des régions prédites comme étant double brin, ce qui ne semble pas indiquer un grand nombre de changements structuraux. L'analyse du segment M montre des conclusions proches (fig. 36), avec des

	Δ NoNP-Comp				Δ ProtK-Comp				Δ NoNP-ProtK			
	NS		M		NS		M		NS		M	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
Dimin. >50%	169	19,0%	255	24,8%	196	22,0%	253	24,6%	45	5,1%	105	10,2%
40% > Dimin. < 50%	6	0,7%	5	0,5%	9	1,0%	7	0,7%	4	0,4%	8	0,8%
Dimin. ou Augm. <40%	529	59,4%	520	50,6%	504	56,6%	510	49,7%	745	83,7%	810	78,9%
40% > Augm. < 50%	11	1,2%	24	2,3%	16	1,8%	24	2,3%	13	1,5%	21	2,0%
Augm. >50%	175	19,7%	223	21,7%	165	18,5%	233	22,7%	83	9,3%	83	8,1%

TABLEAU 6. – Nombre et pourcentage de nucléotides pour lesquels des changements importants, ou non, de réactivité sont détectés.

diminutions de réactivité dans les régions double brin et des augmentations dans les tiges boucles. Quelques changements de réactivité pourraient montrer des modifications structurales, mais ces régions sont peu nombreuses (région 690-695 du segment NS et régions 726-729 et 782-786 par exemple).

Les pourcentages de réactivité ont également été calculé entre les conditions « NoNP » et « Comp » ainsi que « ProtK » et « Comp » (leur analyse plus détaillée sera réalisée dans la prochaine partie), afin de pouvoir comparer les éventuelles différences avec la comparaison « NoNP-ProtK ».

La quantification des changements de réactivité observés (table 6) montre qu'à l'exception de la comparaison Δ NoNP-ProtK pour le segment NS, on retrouve à peu près le même pourcentage de nucléotides pour lesquels on observe une diminution ou une augmentation de réactivité d'au moins 50%. Quand on compare les conditions « NoNP » ou « ProtK » à la condition « Comp », le pourcentage de nucléotides pour lesquels les variations sont limitées à moins de 40% reste très proche (59,4% et 56,6% pour NS, 50,6% et 49,7% pour M), confirmant qu'il n'y a que peu de différences entre les conditions « NoNP » et « ProtK ». La comparaison de ces deux conditions montre que l'on retrouve une grande majorité de nucléotides pour lesquels les changements de réactivité sont limités (83,7% pour NS, 78,9% pour M). Bien qu'une proportion plus importante de nucléotides observe une augmentation de réactivité dans la condition « ProtK » par rapport à la condition « NoNP » pour le segment NS, cela montre qu'il n'y a qu'une part minoritaire de nucléotides pour lesquels des changements importants de réactivité sont observés.

Afin d'exploiter au mieux les données, une analyse avec le logiciel Δ SHAPE [361] a été réalisée. Ce logiciel calcule les différences de réactivités sur une fenêtre glissante de 5 nucléotides. Les erreurs standards pour chaque position sont utilisées afin de réaliser un test Z-factor modifié ainsi qu'un Standard score afin de déterminer les différences significatives.

La comparaison des conditions « NoNP » et « ProtK » pour les segments NS (fig. 37A) et M (fig. 37B) montre que pour le segment NS on n'observe que quatre régions avec des diminutions significatives (table A1). Pour le segment M (table A2), on observe plus de variations, avec sept régions avec des

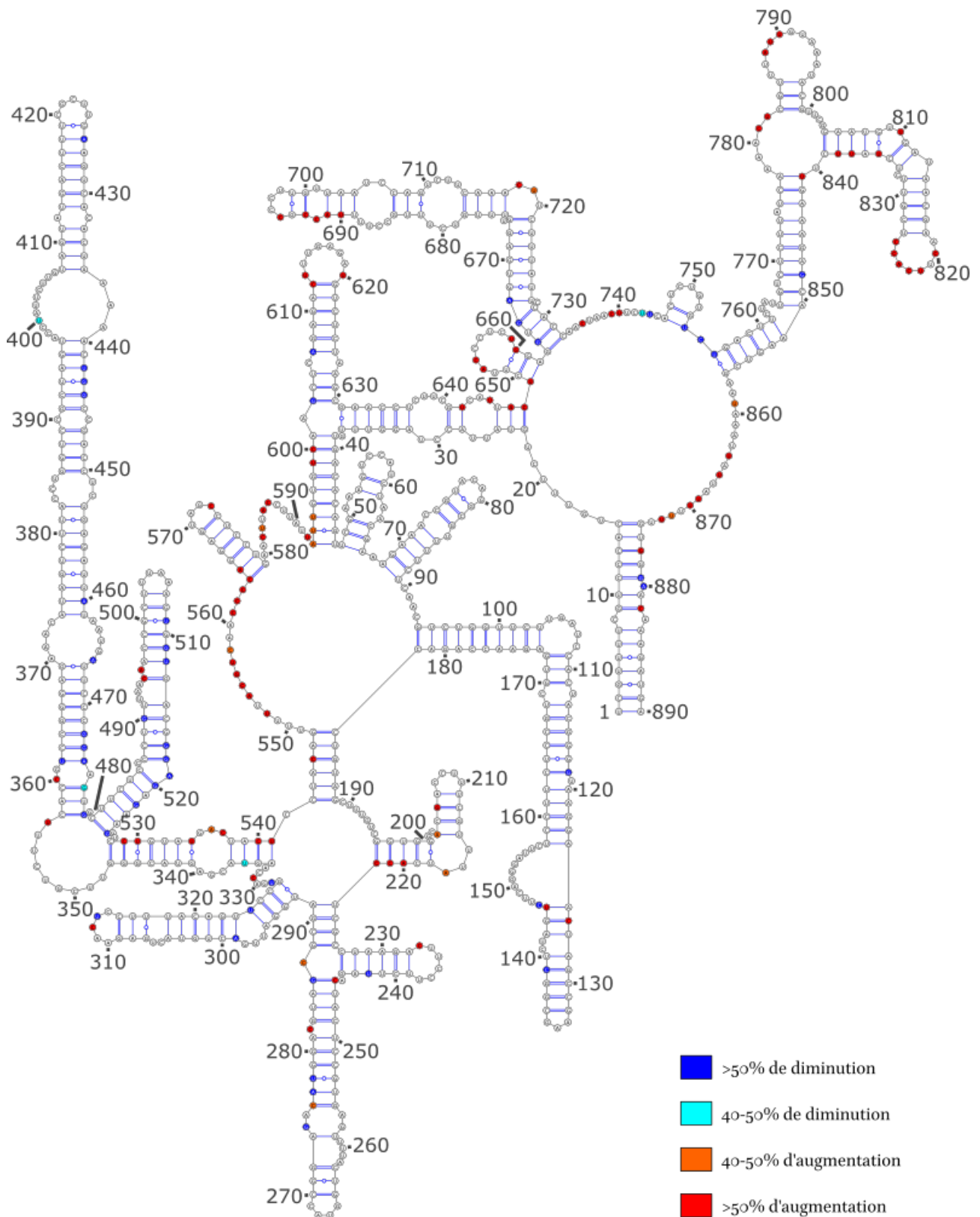


FIGURE 35. – Représentation des changements de réactivités de l'ARN NS entre les conditions « NoNP » et « ProtK » sur la structure de l'ARN prédite avec les données de la condition « NoNP ». La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

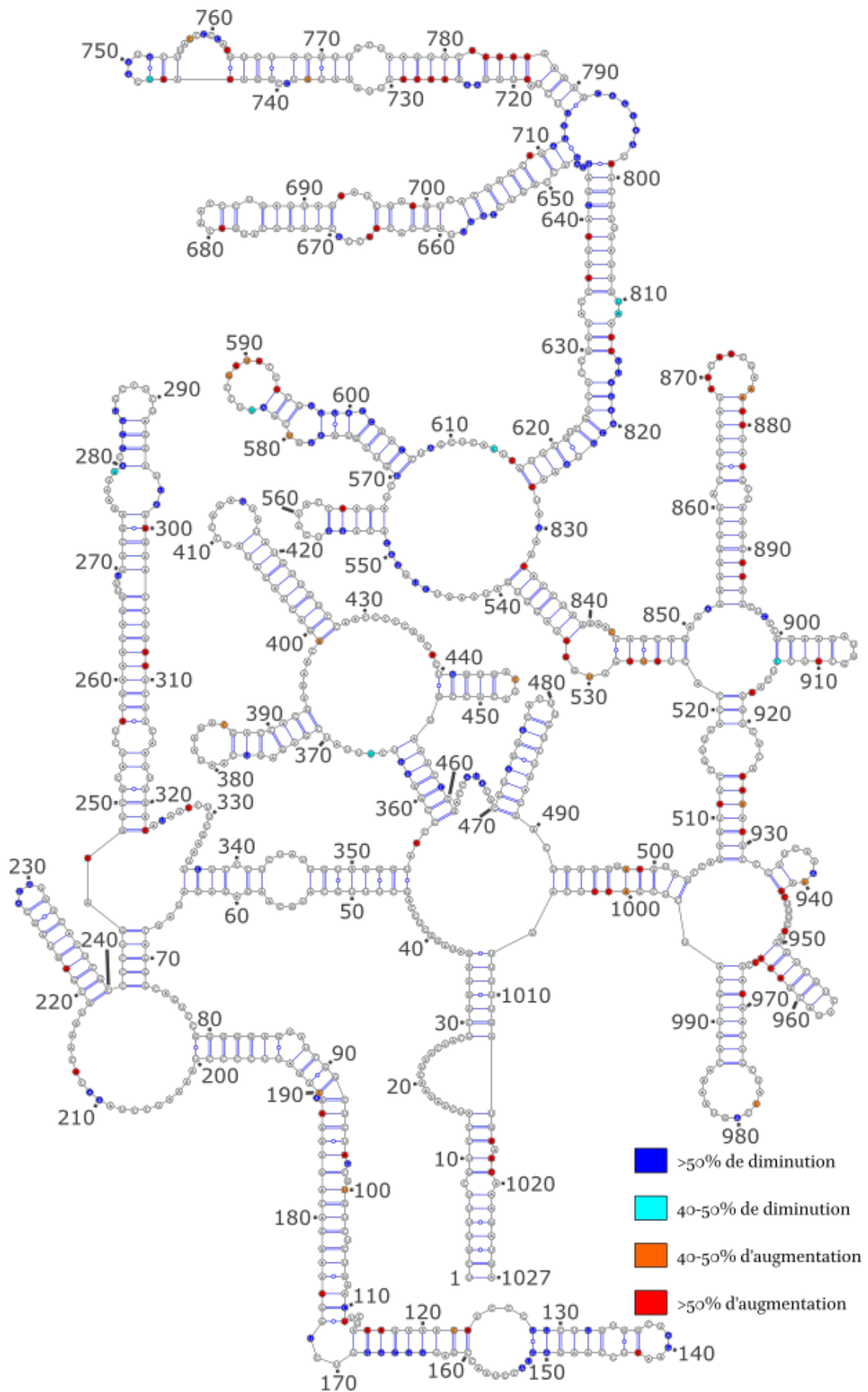


FIGURE 36. – Représentation des changements de réactivités de l'ARN M entre les conditions « NoNP » et « ProtK » sur la structure de l'ARN prédite avec les données de la condition « NoNP ». La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

diminutions significatives et dix-huit régions avec des augmentations significatives. Il faut noter que Δ SHAPE ne conserve que les régions pour lesquelles au moins 3 des 5 nucléotides consécutifs présentent une différence significative de réactivité, ce qui par rapport à notre analyse précédente élimine automatiquement tous les nucléotides ou les paires de nucléotides isolé(e)s. Les différences calculées sont également plus importantes pour le segment NS que pour le segment M. On peut observer plus en détails les 4 régions déterminées comme ayant des différences significatives pour le segment NS par Δ SHAPE : 564-567, 594-597, 653-662 et 700-707. La première région se situe dans une région prédite comme étant simple brin dans la condition « NoNP » et la réactivité augmenterait dans cette région dans la condition « ProtK ». La seconde montre une augmentation d'environ 40% de la réactivité à la base d'une tige boucle qui se situe auprès d'une boucle interne. La troisième correspond à une tige boucle dont seuls 4 nucléotides montrent une augmentation de réactivité d'au moins 50%; cependant 3 de ces nucléotides sont dans la région simple brin de la tige boucle. Enfin la quatrième région ne montre aucune variation de réactivité qui serait biologiquement significative. En conclusion, on ne peut pas conclure sur l'existence d'un remodelage de l'ARN par la protéine NP qui persisterait après l'enlèvement de la protéine pour le segment NS. En effet, il n'y a que peu de différences significatives et celles-ci ne sont pas porteuses d'un sens biologique. Pour le segment M, plus de différences sont visibles (23 régions détectées avec Δ SHAPE), cependant l'un des réplicats de la condition « ProtK » du segment M comporte des valeurs aberrantes qui consistent en des augmentations fortes de réactivité sur un ou deux nucléotides. Quand on regarde les régions détectées à l'aide de Δ SHAPE, la majorité des régions détectées font 3 ou 4 nucléotides de longueur (respectivement 48% et 22%). Concernant les régions plus longues, on observe pour les régions 597-602 et 816-821 une diminution de la réactivité dans des régions prédites comme étant double brin. Pour les régions 708-714, 722-726 et 792-797, on observe des diminutions de réactivité dans des boucles internes. Pour la région 743-751, les réactivités augmentent ou diminuent deux à deux. Enfin dans la région 782-786 les réactivités augmentent dans une région simple brin. Ainsi, contrairement au segment NS, il pourrait y avoir quelques réarrangements structuraux qui seraient toujours présents après l'enlèvement de la protéine NP. On confirme bien ici que les conditions « NoNP » et « ProtK » sont assez proches pour le segment NS et présentent quelques différences pour le segment M. Les quelques différences significatives, certes limitées, indiquent que la protéine NP n'a qu'une action de remodelage pérenne de la structure des ARNv plutôt limitée.

Les données obtenues sur les segments NS et M dans les conditions « NoNP » et « ProtK » ont été utilisées afin de modéliser des structures 3D des ARN à l'aide de 3dRNA [362]. Les structures sont présentées dans les annexes. Il est intéressant de noter que pour le segment NS, la structure modélisée pour la condition « NoNP » (annexe 9) apparaît comme plus allongée et moins condensée que pour la condition « ProtK » (annexe 10). Pour le segment M, on observe des structures modélisées plus proches entre la condition « NoNP » (annexe 11) et la condition « ProtK » (annexe 12), avec une tige boucle qui semble se dégager du reste de la structure dans la condition « ProtK ». Les données du segment NS peuvent permettre d'émettre l'hypothèse que certaines des différences visibles entre les données « NoNP » et « ProtK » pourraient être dues à une compaction de l'ARN suite à l'élimination de la protéine NP, qui forme un squelette au centre de l'ARN.

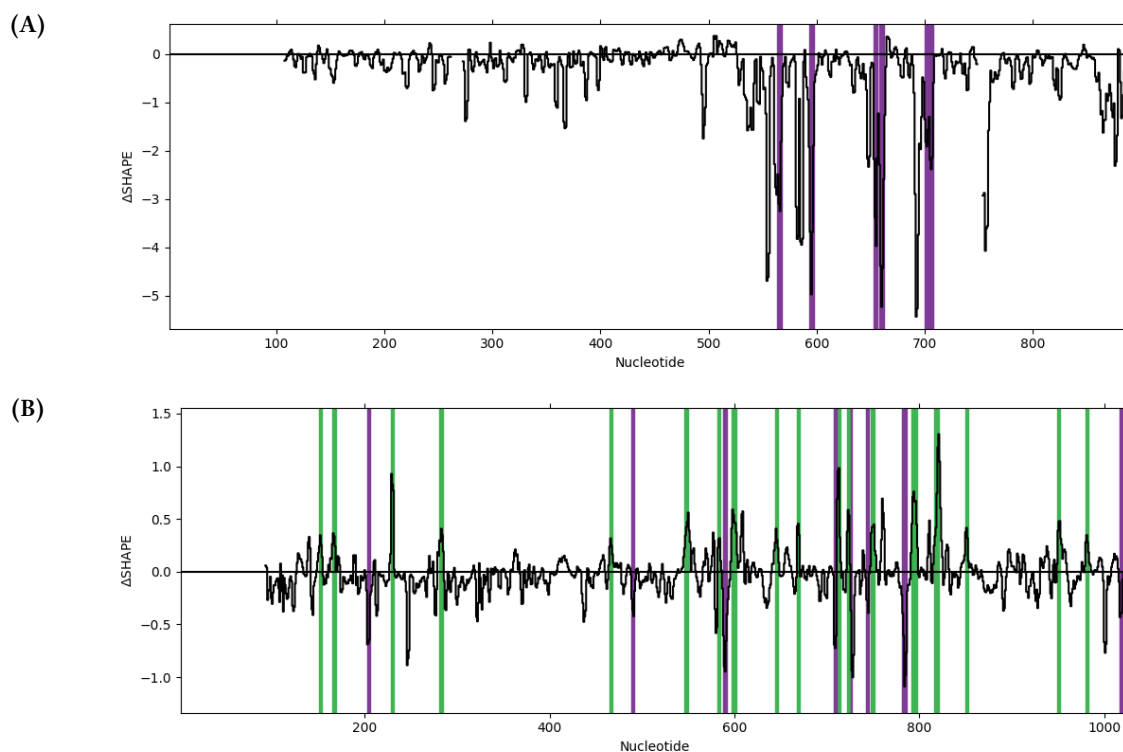


FIGURE 37. – Représentation des différences obtenues avec Δ SHAPE entre les conditions « NoNP » et « ProtK » (Δ NoNP-ProtK) pour les segments NS (A) et M (B). Les barres vertes et violettes représentent respectivement des différences significatives de réactivités positives et négatives. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

Ainsi, mes données ne sont pas en faveur de l'existence d'une activité chaperon au sens strict de la protéine NP, ceci pour plusieurs raisons. Tout d'abord, une protéine possédant un effet chaperon de l'ARN se fixe de façon transitoire à l'ARN, ce qui n'est pas le cas de la protéine NP, qui reste liée en permanence à l'ARN, à l'exception des étapes de transcription et de réplication. Ensuite, cette activité chaperon de l'ARN subsiste même après élimination ou décrochage de la protéine NP. Mes données montrent qu'après avoir retiré la protéine NP, l'ARN retrouve une structure proche des ARNv nus. Quelques différences subsistent, mais contrairement aux ARNv nus qui se replient d'une façon plus « anarchique », les ARNv desquels sont retirés la protéine NP ont tous une structure très proche, ce qui limite la variabilité des structures obtenues lorsque la protéine NP est retirée et que les RNPv « s'effondrent » sur elles-mêmes.

c. Analyse de la liaison de la NP à l'ARN

Afin de déterminer les différentes régions exposées au solvant en présence de la protéine, les complexes entre l'ARN et la NP ont été formés avant de réaliser la modification. Une modification de l'ARN seul a été réalisée en parallèle afin de pouvoir observer les différences de structure entre les deux conditions. La présence de la protéine NP devrait avoir deux effets : d'une part remodeler la structure de l'ARN, changeant les régions simples brin ou liées à l'ARN et d'autre part se fixer dans certaines régions, y laissant une empreinte.

Modifications de la structure de l'ARN

Dans un premier temps, les changements de réactivité ont été analysés comme précédemment pour les comparaisons NoNP-Comp et ProtK-Comp pour les segments NS et M. Lorsque l'on compare les conditions « NoNP » et « Comp » du segment NS (fig. 38), on observe de nombreux changements de réactivité sur plusieurs nucléotides consécutifs. La majorité des diminutions de réactivité sont présentes au sein des boucles terminales, internes et des hernies et plus rarement sur les extrémités d'hélices peu stables. La principale difficulté pour ces analyses en présence/absence de la protéine NP est qu'il n'est pas possible de distinguer, avec uniquement le réactif NMIA, si une région est double-brin ou liée à la protéine NP. Ainsi, il ne sera pas possible d'obtenir des régions potentielles de liaison pour toutes les protéines NP liées à l'ARN. Deux régions sont assez longues (397-408 et 858-877) et pourraient être des sites de liaison à la protéine NP. On pourrait avoir la formation d'une boucle interne au niveau des nucléotides 274-281. La tige boucle 227-244 devient simple brin en présence de NP alors que la boucle terminale pourrait devenir double brin ou fixer la NP. Les hélices 292-327 et 804-839 s'ouvrent également. Des protections sont observées sur les nucléotides 614-619, 791-795 qui sont soit protégés par la NP soit deviennent appariés.

On retrouve pour le segment M des caractéristiques identiques (fig. 39). Les changements de réactivités sont majoritairement observés sur des nucléotides contigus, avec des diminutions de réactivité au niveau des boucles et des augmentations au niveau des régions d'ARN double brin. Ici aussi, les augmentations et diminution sont très majoritairement d'au moins 50% de la valeur initiale. A la différence du segment NS, la région correspondant au milieu du segment (~513-514) ne se situe pas dans une longue région simple brin qui s'ouvre en présence de la protéine NP mais comporte plus de boucles terminales et de boucles internes.

On observe la formation d'une région simple brin aux positions 117-122, ainsi qu'à proximité de l'extrémité apicale de la tige boucle 133-146.

Les régions 123-128, 139-144 seraient soit fixées par la NP, soit appariées à une région de l'ARN. Les extrémités apicales de la tige boucle 371-393 et 440-453 seraient également double brin ou liées à la NP quand la protéine est présente. La tige boucle 552-567 et la boucle terminale 583-594 subiraient les mêmes changements.

Dans la tige boucle 647-712, on observe des diminutions de réactivité au niveau des boucles internes qui seraient donc soit fixées par la NP, soit appariées alors que les régions 687-692 et 698-701 s'ouvrent. On retrouve la même chose pour la tige boucle 714-791 où les régions double brin entre les boucles internes s'ouvrent en présence de NP alors que les boucles internes voient leur réactivité diminuer. Pour la tige boucle

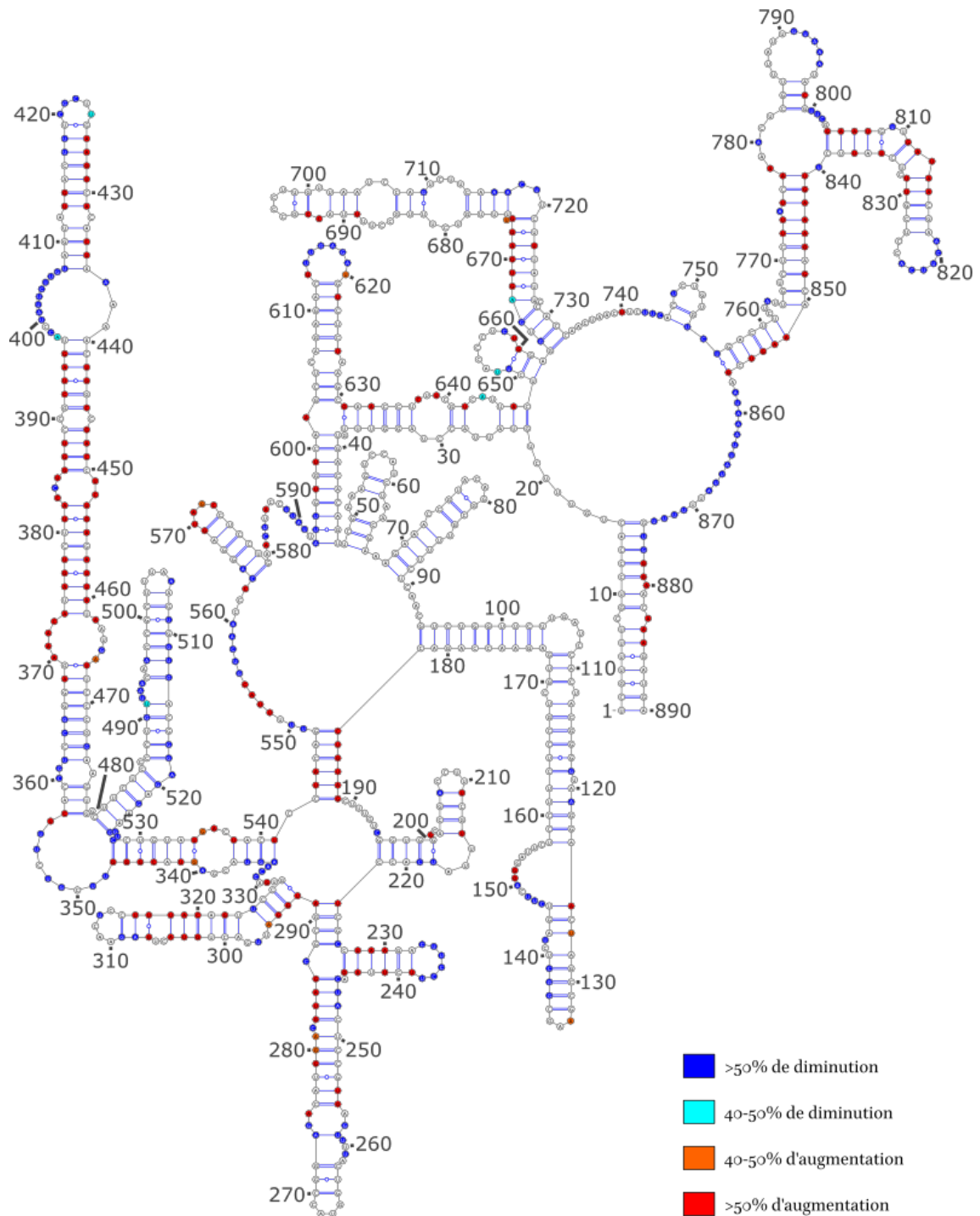


FIGURE 38. – Représentation des changements de réactivités de l'ARN NS entre les conditions « NoNP » et « Comp » sur la structure de l'ARN prédite avec les données de la condition « NoNP ». La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

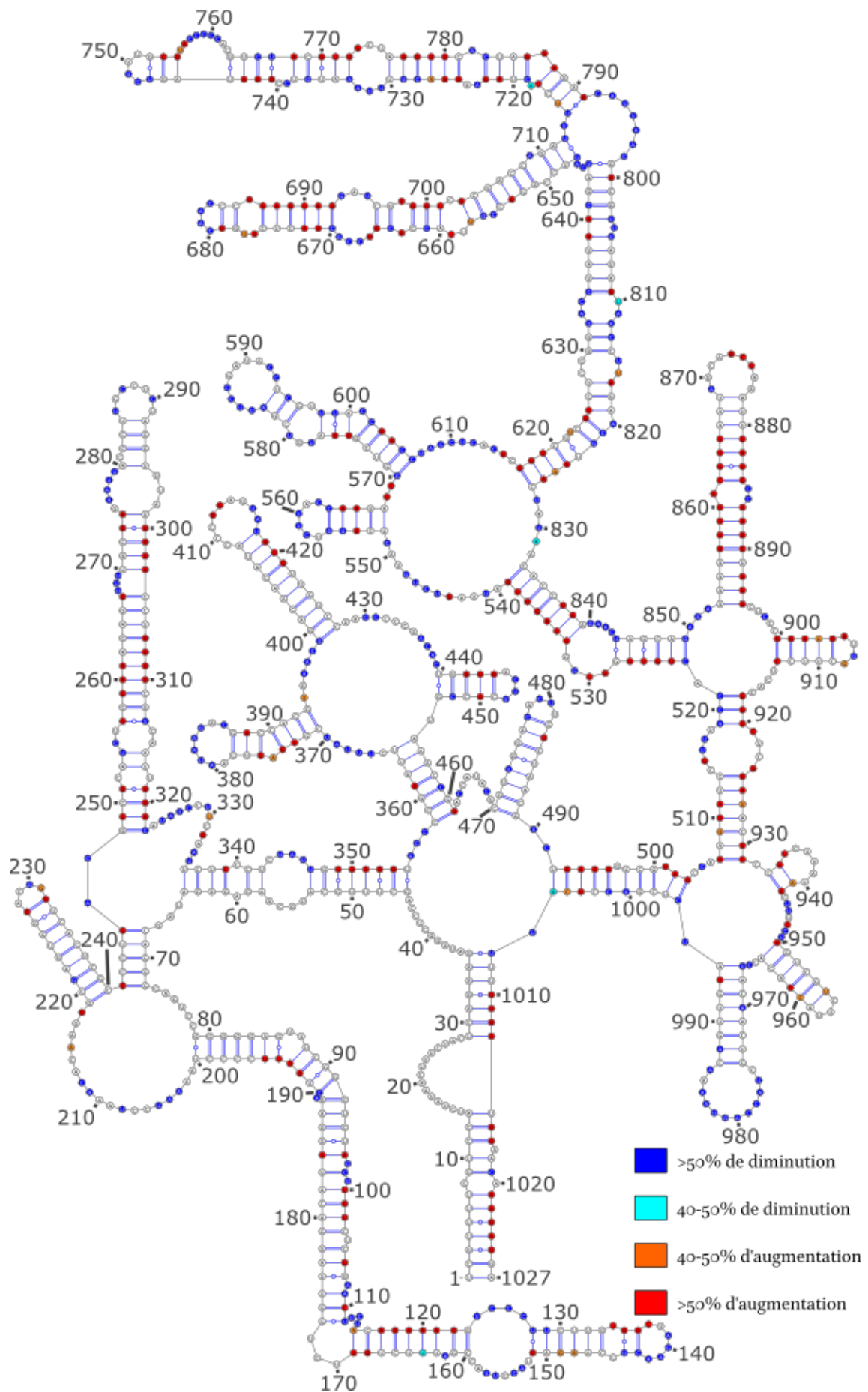


FIGURE 39. – Représentation des changements de réactivités de l'ARN M entre les conditions « NoNP » et « Comp » sur la structure de l'ARN prédite avec les données de la condition « NoNP ». La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

248-322, les deux boucles internes en 266-270 et 274-280 deviendraient double brin ou liées à la NP tandis que les régions 300-303 et 307-310 deviennent simple brin. La boucle interne formée par les nucléotides 644-647, 712-714 et 792-799 serait également modifiée de la même manière. Cela serait aussi le cas pour la boucle terminale 974-987. La tige boucle 853-894, ainsi que la moitié de la tige boucle 900-913 s'ouvrent presque en présence de la protéine.

La région double brin correspondant au milieu du segment (507-513) devient simple brin après fixation de la protéine. Les autres régions double brin à proximité (493-496, 501-503, 524-527 et 533-539) voient également leurs réactivités augmenter fortement en présence de la protéine. Les boucles internes à proximité possédant de nombreux nucléotides dont la réactivité diminue, il y aurait donc des restructurations importantes dans cette région.

On peut noter l'augmentation de réactivité dans des régions proches de l'extrémité 5' du segment (1021-1025), sensées interagir avec le complexe réplicase.

Si l'on compare les conditions « ProtK » et « Comp », on observe que les variations de réactivités sont peu ou prou les mêmes que lors de la comparaison des conditions « NoNP » et « Comp », que ce soit pour le segment NS (annexe 13) ou le segment M (annexe 14). On retrouve quelques régions où les réactivités varient de façon importante entre les deux conditions (par exemple les régions 165-169, 522-525 et 791-799 pour le segment M), mais ces régions sont généralement sans changements significatif dans une des comparaisons avec la condition « Comp » alors que l'autre montre une augmentation ou une diminution de la réactivité. On ne retrouve cependant pas de régions longues pour lesquelles on observerait dans une comparaison une augmentation de réactivité en présence de la NP et dans l'autre comparaison, une diminution de la réactivité. Le milieu du segment ne montre pas de changements importants de réactivités entre les conditions, que ce soit pour le segment NS ou le segment M. On a donc bien des données assez proches entre les deux conditions « NoNP » et « ProtK », bien que quelques différences locales soient visibles, ce qui confirme les observations précédentes.

Bien que nous ayons pu déterminer plusieurs régions potentiellement impliquées dans la liaison à la protéine NP, il reste encore à affiner l'analyse de ces régions, en cherchant lesquelles sont statistiquement et biologiquement les plus significatives et quel est le sens biologique des augmentations et diminutions de réactivité. En effet, ces deux dernières auront un sens biologique clairement défini uniquement si l'augmentation ou la diminution de la réactivité dénote d'un changement de la flexibilité de la base. A titre d'exemple, une diminution d'une valeur de 2 de réactivité à 1, représente une diminution importante cependant, pour les deux valeurs, ce nucléotide serait considéré comme ayant une forte flexibilité. Il faut donc prendre ce paramètre en compte lors des prochaines analyses.

Analyse de la fixation de la protéine NP sur l'ARN

Dans un second temps, nous avons voulu affiner l'analyse de la fixation de la protéine NP en cherchant des régions avec des différences significatives qui pourraient correspondre à des sites de fixation de la protéine NP ainsi que la composition de ces régions protégées.

Comme pour la recherche de l'activité chaperon de l'ARN de la NP, les comparaisons $\Delta\text{NoNP-Comp}$ et $\Delta\text{ProtK-Comp}$ ont été analysées à l'aide du logiciel ΔSHAPE .

Les différences sont assez limitées pour le segment NS lorsqu'on compare les données $\Delta\text{NoNP-Comp}$ (fig. 40A) et $\Delta\text{ProtK-Comp}$ (fig. 40B). On retrouve quelques différences néanmoins. Par exemple, on retrouve la région 235-237 où la différence est significative pour les données $\Delta\text{NoNP-Comp}$ mais pas pour $\Delta\text{ProtK-Comp}$. Cependant le profil des courbes à cet endroit est le même, suggérant qu'une plus grande variabilité au niveau des erreurs standards pourrait expliquer cette différence de signifiante statistique. On observe la même situation pour les régions 426-428 et 458-461, où les profils sont similaires bien que non statistiquement différents dans les données $\Delta\text{ProtK-Comp}$, au contraire des données $\Delta\text{NoNP-Comp}$. Certaines différences sont cependant dues à des différences de profil, comme les régions 590-595, 654-663, 705-707, 754-758 et 842-847.

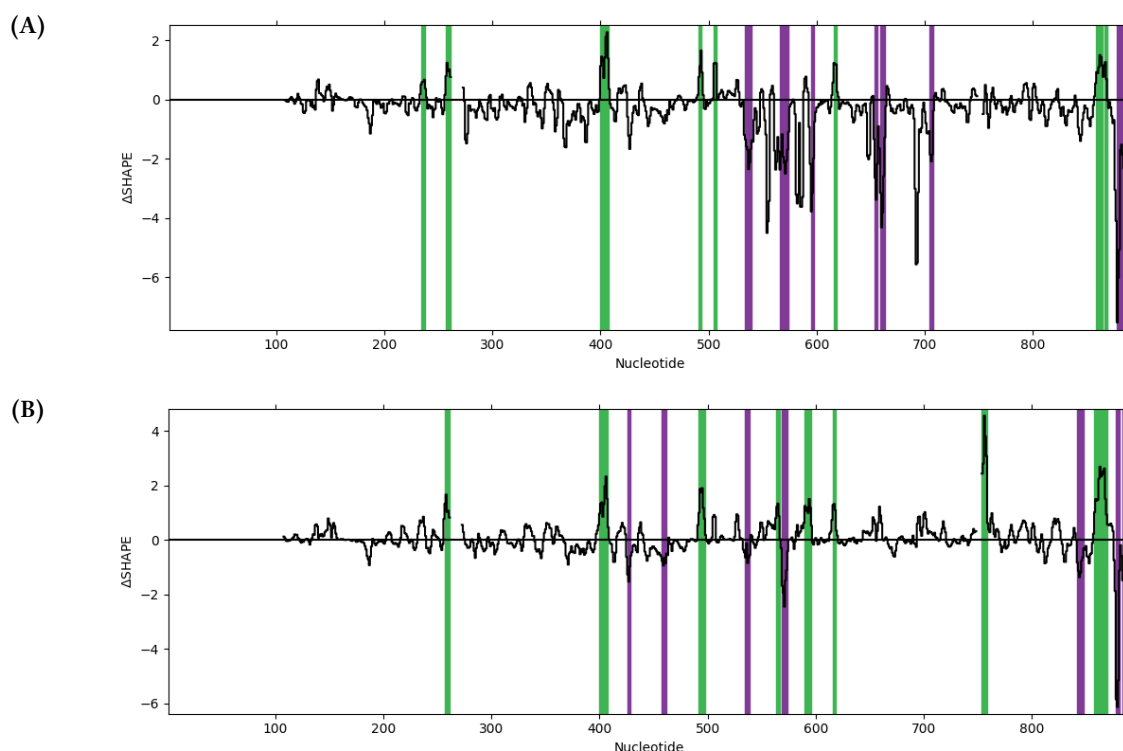


FIGURE 40. – Représentation des différences obtenues avec ΔSHAPE entre les conditions « NoNP » et « Comp » ($\Delta\text{NoNP-Comp}$) (A) ou « ProtK » et « Comp » ($\Delta\text{ProtK-Comp}$) (B) pour le segment NS. Les barres vertes et violettes représentent respectivement des différences significatives de réactivités positives et négatives. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

On retrouve des observations similaires pour le segment M, avec des profils $\Delta\text{NoNP-Comp}$ (fig. 41A) et

Δ ProtK-Comp (fig. 41B) très proches. Comme pour le segment NS, une partie des différences visibles entre les conditions Δ NoNP-Comp et Δ ProtK-Comp sont attribuables à des erreurs standards trop élevées, d'autres à des différences de profil, comme les régions 98-102, 142-144, 166-169, 212-214, 246-248, 259-261, 300-309, 478-480, 501-503, 525-528, 924-926 et 950-954. Par exemple, dans la région 550-552, on a le même profil, avec uniquement une différence d'intensité. On n'observe pas, comme pour le segment NS, de différences marquées de profil.

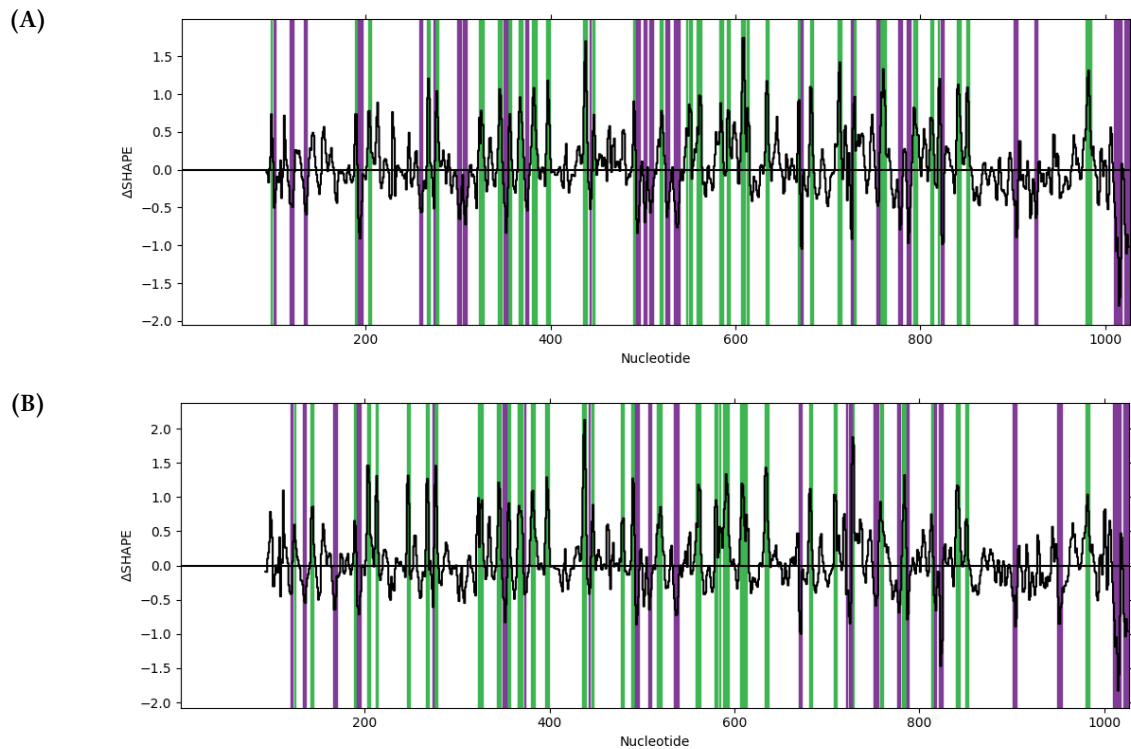


FIGURE 41. – Représentation des différences obtenues avec Δ SHAPE entre les conditions « NoNP » et « Comp » (Δ NoNP-Comp) (A) ou « ProtK » et « Comp » (Δ ProtK-Comp) (B) pour le segment M. Les barres vertes et violettes représentent respectivement des différences significatives de réactivités positives et négatives. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

Le logiciel Δ SHAPE bien qu'utilisable pour nos expériences, a été conçu dans l'optique d'analyser des données issues de la technique SHAPE-MaP. La technique de RT utilisée dans nos travaux peut entraîner une variabilité plus importante des données, ce qui influe sur l'erreur utilisée par Δ SHAPE pour déterminer quelles sont les différences significatives ou non. Ainsi pour obtenir le maximum d'informations sans en perdre, les données obtenues dans les différentes conditions ont été traitées de la sorte : chaque nucléotide s'est vu attribué une catégorie (0 : non réactif; 1 : moyennement réactif (entre 0,4 et 0,8); 2 : fortement réactif (supérieur à 0,8)) puis les changements de catégories ont été calculés pour les nucléotides dont la réactivité varie d'au moins 40%. Ces données ont ensuite été recoupées avec les résultats issus de la comparaison avec Δ SHAPE.

Afin d'essayer d'avoir des informations sur l'environnement des nucléotides détectés avec Δ SHAPE, les nucléotides pour lesquels on n'observe pas de changement de catégories sont également représentés, une

couleur spécifique est donc associée aux nucléotides qui ne sont pas réactifs dans les deux conditions (gris), faiblement réactifs (jaune) ou fortement réactifs (magenta). Les nucléotides pour lesquels on observe un changement de catégorie mais une variation de la réactivité inférieure à 40% ou bien une variation importante de la réactivité sans changement de catégorie sont représentés en blanc.

Les données entre les conditions « NoNP » et « ProtK » étant fortement similaires, seules seront abordées ici les comparaisons entre les données « ProtK » et « Comp » pour les segments NS (fig. 42) et M (fig. 43), les comparaisons entre « NoNP » et « Comp » sont présentées dans les annexes (annexe 15) et (annexe 16).

Il est ainsi possible de chercher la présence de régions qui pourraient être impliquées dans des interactions ARN-ARN ou ARN-NP ou bien des régions qui seront exposées en présence de la protéine NP, même si avec uniquement des données obtenues avec le réactif NMIA, il sera impossible de déterminer si il s'agit d'une interaction ARN-ARN ou ARN-NP. Dans le premier cas, il va s'agir de régions comportant plusieurs nucléotides pour lesquels on observe une diminution d'une ou deux catégories, dans l'idéal ayant été également détectée par Δ SHAPE. Il est possible que des nucléotides non-réactifs soient présents dans cette région : la fixation de la NP pourrait ouvrir une région double-brin, dans laquelle une autre protéine NP se fixerait. Dans ce cas, le nucléotide serait en interaction dans les deux conditions et posséderait donc une réactivité inférieure à 0,4. Dans le second, il s'agira de régions qui pourront être de potentiels sites d'interaction avec les autres segments.

Pour le segment NS (fig. 42), une première région est détectée par Δ SHAPE aux positions 235-237. On pourrait supposer l'existence d'une liaison à la NP sur les douze nucléotides allant de 231 à 242, comportant les différents nucléotides détectés ici ainsi que d'autres nucléotides avec des variations forte de réactivité. Une seconde région est détectée aux positions 257-261. Une protéine NP pourrait se fixer à partir du G en position 257 jusqu'à la position 268. La région 262-272 n'étant pas analysable, il est difficile de confirmer cela. Cependant la fixation de la NP à cette région la placerait à 14 nucléotides de la région précédemment identifiée. On sait que l'on compte 24 nucléotides pour une protéine NP, si l'on suppose que les nucléotides sont répartis en deux groupes de six nucléotides qui ne sont pas en interaction avec la protéine NP encadrant douze nucléotides interagissant avec la protéine, cette région serait placée correctement. La région 330-341 pourrait également être un site de fixation de la protéine NP. Les nucléotides 397-408 sont une des cibles les plus probables pour la fixation de la protéine NP. On retrouve cette région à proximité du milieu du segment, que la fixation de la NP permettrait d'ouvrir. De plus, il s'agit de l'une des longues régions détectées par Δ SHAPE. Il est possible d'imaginer une liaison de la protéine aux positions 490-497, mais l'environnement étant très peu réactif, il est difficile de tirer des conclusions. Bien que Δ SHAPE détecte des différences significatives pour les nucléotides 505-507, il s'agit d'une région qui ne varie que très peu. Les variations mesurées ne correspondant pas à des changements de catégories, il est donc difficile de conclure sur le sens biologique de ces variations. On retrouve la même situation pour la tige boucle 564-579. Pour la boucle 613-620, on voit une diminution de la réactivité pour les nucléotides présents dans la boucle, alors que les deux U qui forment les débuts de la boucle voient

leur réactivité augmenter. Bien qu'il soit possible que la NP se fixe à cette boucle, il est aussi probable que suite aux réarrangements structuraux, cette boucle interagissent avec une autre région du segment. Pour la boucle 651-660 et la boucle interne 684-689/704-707, Δ SHAPE détecte des variations significatives pour des nucléotides qui restent non réactifs dans les deux conditions, mais cela ne doit pas avoir de significations concernant la structure et plus être dû aux variations des réactivités, où l'on retrouve des alternances entre des valeurs fortement négatives et des valeurs positives également très fortes. En prenant notamment en compte les variations de réactivité, on ne peut exclure l'existence de deux sites de fixation de la protéine. Le premier allant de la position 791 à la position 802 et le second se situant dans la région 816-829. Ces deux régions seraient situées à un peu plus de douze nucléotides de distance, donc en cohérence avec les connaissances actuelles, mais on retrouve également une ouverture des régions situées à proximité de ces deux régions (804-815 et 830-839). Enfin, un site potentiel de liaison à la NP se trouverait dans la région 858-877, ou s'apparierait à une autre région de l'ARN.

Dans le cas du segment M (fig. 43), on retrouve plus de régions détectées par Δ SHAPE comme présentant des différences significatives, mais beaucoup sont très courtes (ex : position 124) et plusieurs ne sont pas dans des contextes qui sont facilement interprétables (régions 134-136 et 142-144). On pourrait estimer la position d'un premier site de fixation de la protéine NP dans la région 197-208, également détectée en partie par deltaSHAPE. La région 246-255 pourrait également faire partie d'un site de fixation à la protéine il y a beaucoup de nucléotides non réactifs et quelques nucléotides qui voient des augmentations assez fortes de réactivité, ainsi il est difficile de l'affirmer. Les nucléotides 277-279 subissent des changements probablement dû à des réarrangements structuraux étant donné la proportion des nucléotides environnants à voir leur réactivité augmenter. Les positions 322-333 pourraient faire partie d'un site de liaison à la NP, tout en étant détectées par Δ SHAPE. Des réarrangements structuraux pourraient expliquer les changements des nucléotides 444-447. Il en est de même pour les nucléotides 350-357. Étant donné la diminution de réactivité des nucléotides de la boucle 379-386, il serait possible de retrouver un site de liaison à la NP dans la région 364-386, mais il est difficile de le situer précisément. Les nucléotides 396-399 pourraient faire partie d'un site de liaison mais il y a trop peu d'information pour l'envisager fortement. La région 428-439 est une bonne candidate pour la fixation de la NP. Les positions 516-523 pourraient faire partie d'un site de liaison. La fixation de la protéine dans cette région pourrait permettre l'exposition des nucléotides situés en 534-539. La boucle 556-563 pourrait interagir avec une région de l'ARN après fixation de la protéine à proximité (par exemple dans la région 540-551). Un site de fixation potentiel peut être retrouvé dans la région 584-595. Un autre site serait envisageable aux positions 606-617 mais ces deux sites sont assez proches, dans une région de 33 nucléotides. Soit la structure de l'ARN permet la fixation de deux NP relativement proches, soit l'on peut retrouver localement un ratio différent de 1 protéine NP pour 24 nucléotide tel que rapporté dans la littérature. De même, la région 627-638 pourrait être un site de fixation mais est également proche de la seconde région. Le manque de nucléotides démontrant des augmentations ou diminutions fortes de réactivités dans cette région empêche de pouvoir définir précisément si un, deux ou trois sites de liaisons pourraient être présents ici. Une autre région candidate serait la région 726-737. On peut envisager un autre site de fixation dans la région 840-851. La boucle 974-987 pourrait soit comporter un site de fixation, soit se réappairer avec une

	ARN NS	Protection NS	Changement	ARN M	Protection M	Changement
A	208 (23,4%)	27 (26,5%)	+3,1%	253 (24,6%)	23 (25,6%)	+1%
U	285 (32%)	40 (39,2%)	+7,2%	298 (29%)	21 (23,3%)	-5,7%
C	215 (24,2%)	20 (19,6%)	-4,6%	263 (25,6%)	35 (38,9%)	+13,3%
G	182 (20,4%)	15 (14,7%)	-5,7%	213 (20,8%)	11 (12,2%)	-8,6%

TABLEAU 7. – Nucléotides présents dans les régions où se fixe potentiellement la protéine NP.

autre région de l'ARN. Comme pour le segment NS, on retrouve une augmentation des réactivités sur la fin du segment (1011-1025) mais il se peut que cela soit dû au fait que l'on approche de l'extrémité de l'ARN, ce qui générerait un signal plus important, sans que cela soit lié au fait que les bases soient en interaction ou non.

Plusieurs limitations subsistent pour les deux ARN. Tous les sites de fixation de la protéine NP ne sont pas visible : en se fixant sur une région simple brin, la NP peut permettre l'ouverture d'une région précédemment double brin; si la NP se fixe sur l'un ou les deux brin(s) devenu(s) simple brin, la fixation de la NP sur cette région ne sera pas visualisable avec le réactif SHAPE. De plus, la majorité des sites de fixation de la protéine semblent être situés au sein des boucles; si la protéine NP se fixe dans une région précédemment double brin, son effet ne sera pas visible, sauf si l'hélice d'ARN est déstabilisée.

La répartition des nucléotides impliqués dans des diminutions de réactivité en présence de la protéine NP a été analysée (table 7). On retrouve majoritairement des C et des U (respectivement 28,6% et 31,8%), 26% de A et des G en minorité (13,5%). D. Ferhadian avait obtenu des résultats très similaires en utilisant une NP sous forme essentiellement trimérique [98]. Par rapport à la composition des ARNv NS et M, les régions protégées par la protéine NP sont légèrement enrichies en C (28,6% contre 25% dans l'ensemble des nucléotides des deux ARN) et appauvries en G (13,5% contre 20,6%), alors que les nucléotides A (26% contre 24%) et U (31,8% contre 30,4%) sont dans la même proportion que la composition des ARN.

La composition en dinucléotides a également été analysée (fig. 44) afin d'essayer de déterminer l'existence de motifs préférentiellement liés à la protéine NP. On retrouve ainsi une proportion plus importante de dinucléotides AC, UA et CC dans les sites potentiels de liaison à la NP que dans les ARN alors que les dinucléotides AG, UG, GA et GG sont moins représentés. Ces données sont en corrélation avec l'enrichissement des sites potentiels de liaison à la NP en G et leur appauvrissement en C, ce qui avait également été observé par D. Ferhadian.

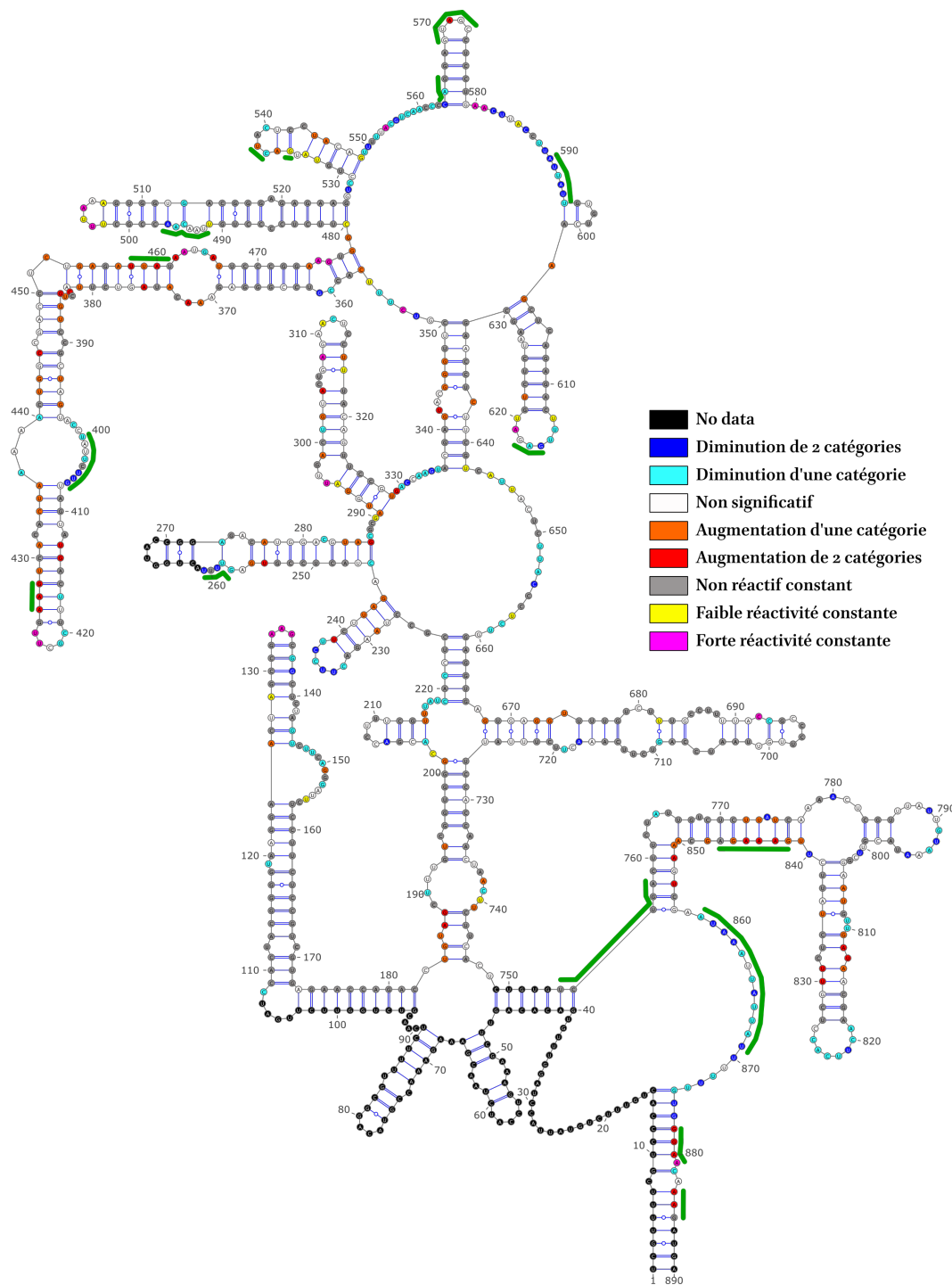


FIGURE 42. – Représentation des changements de catégories et des régions identifiées à l'aide de Δ SHAPE pour le segment NS entre les conditions « ProtK » et « Comp ». La structure secondaire a été prédite à l'aide des données « ProtK » et les catégories ont été définies selon les valeurs de réactivité : 0 : non réactif; 1 : moyennement réactif (entre 0,4 et 0,8); 2 : fortement réactif (supérieur à 0,8). Les changements de catégories ont été calculés pour les positions pour lesquels un changement de la réactivité d'au moins 40% a été observé. Un changement de catégorie sans un changement suffisant de la réactivité est indiqué en blanc. Les nucléotides pour lesquels les catégories ne changent pas entre les deux conditions sont également indiqués. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

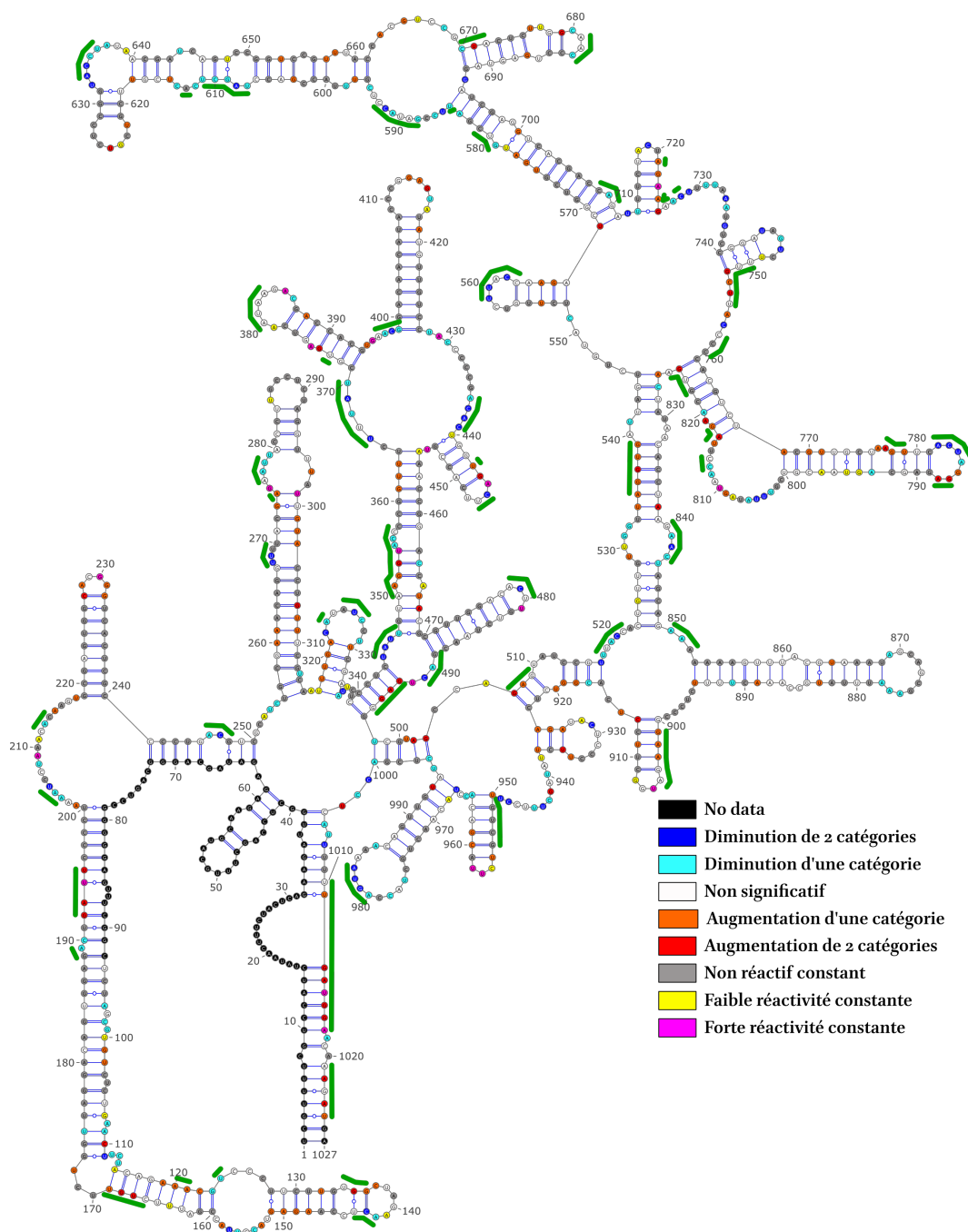


FIGURE 43. – Représentation des changements de catégories et des régions identifiées à l'aide de Δ SHAPE pour le segment M entre les conditions « ProtK » et « Comp ». La structure secondaire a été prédite à l'aide des données « ProtK » et les catégories ont été définies selon les valeurs de réactivité : 0 : non réactif; 1 : moyennement réactif (entre 0,4 et 0,8); 2 : fortement réactif (supérieur à 0,8). Les changements de catégories ont été calculés pour les positions pour lesquels un changement de la réactivité d'au moins 40% a été observé. Un changement de catégorie sans un changement suffisant de la réactivité est indiqué en blanc. Les nucléotides pour lesquels les catégories ne changent pas entre les deux conditions sont également indiqués. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

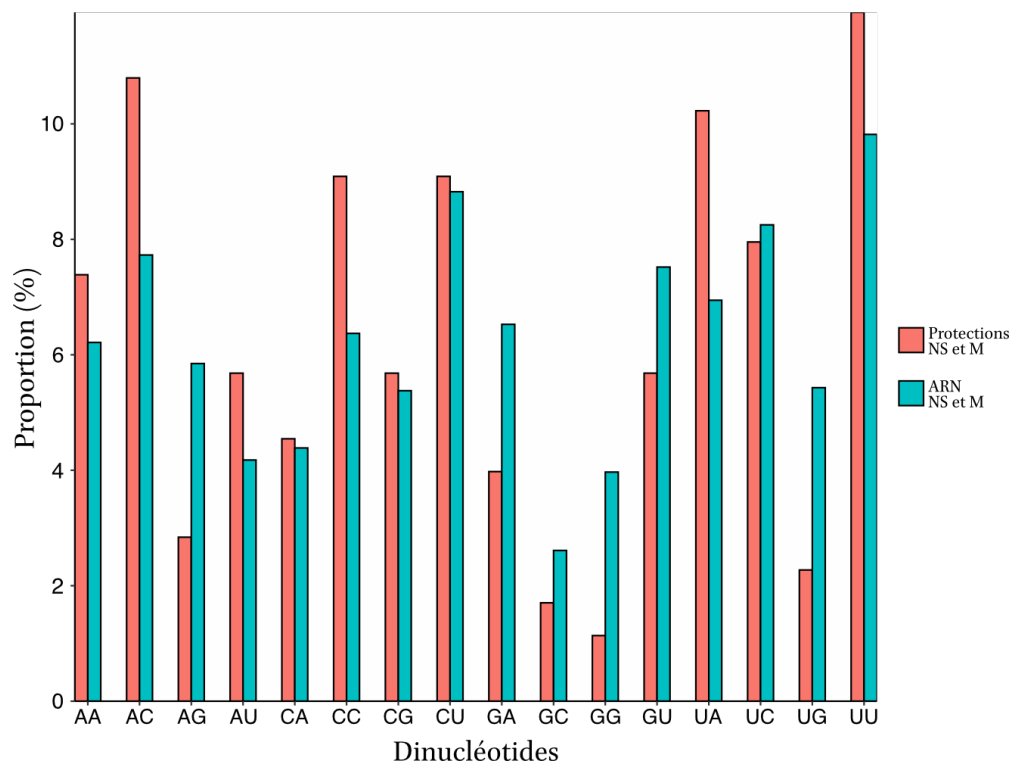


FIGURE 44. – Composition des régions potentiellement protégées par la protéine NP en dinucléotides. Les pourcentages de dinucléotides observés dans les régions protégées des segments NS et M cumulés sont représentés par les barres rouges et la composition en dinucléotide des deux segments cumulés est représentée par les barres bleues.

d. Comparaison avec la littérature

La structures des ARN viraux de différentes souches de virus Influenza ont été étudiés par diverses techniques, notamment des techniques de SHAPE-MaP [277]. Les données obtenues au cours de cette étude ont été récupérées afin de les comparer à celles obtenues au cours de cette thèse. Le travail dans cette étude a été mené sur des ARN nu produits par transcription *in vitro*, des ARN au sein de particules virales et sur des ARN obtenus après traitement des particules virales à la protéinase K.

Dans un premier temps, les réactivités ont été comparées entre les différentes conditions : « NoNP » comparé aux ARN produits *in vitro*, « ProtK » comparé aux ARN obtenus après traitement à la protéinase K et « Comp » aux ARN modifiés à l'intérieur de la particule virale (fig. 45). Les R^2 calculés ainsi sont tous faibles, montrant de fortes divergences entre les données obtenues au cours de cette thèse et celles obtenues par B. Dadonaite *et al.* [277]. Il est à noter que dans cette étude, il n'a pas été réalisé de façon systématique un triplicat pour chaque condition.

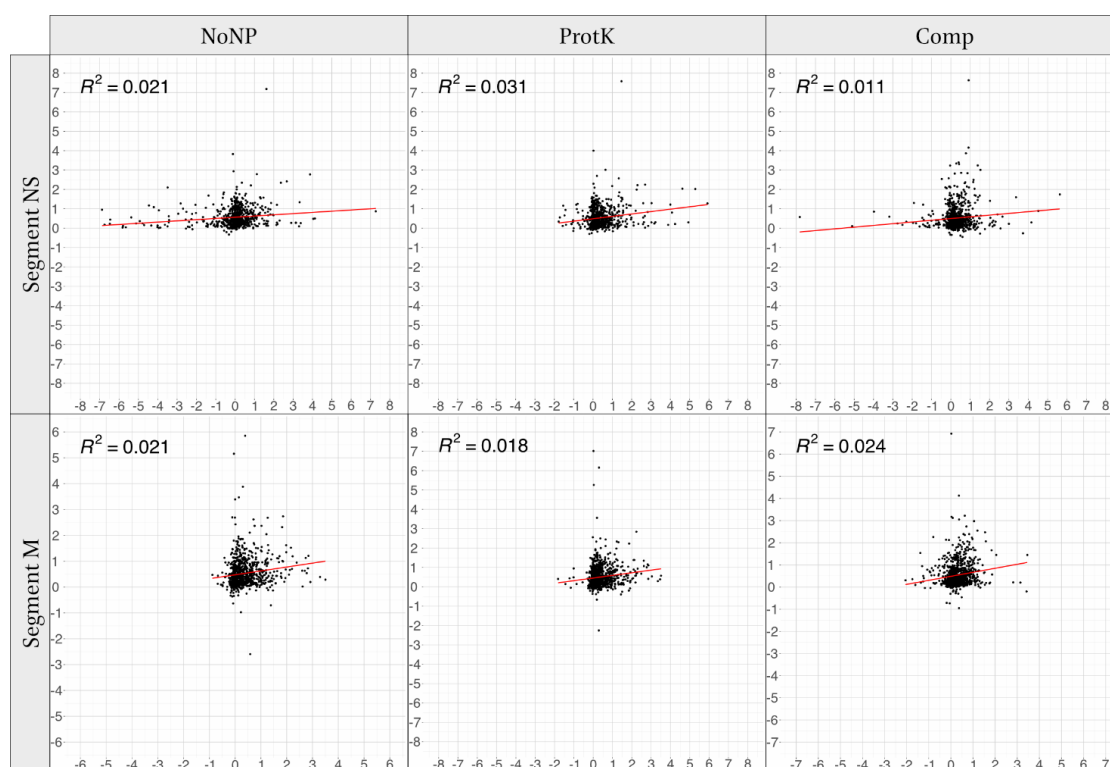


FIGURE 45. – Corrélations des données NS et M avec les données de B. Dadonaite *et al.* [277]. Les données de cette thèse sont en abscisse et les données de B. Dadonaite sont en ordonnée.

Cependant la répartition des réactivités est sensiblement la même entre les données de B. Dadonaite *et al.* [277] et celles obtenues au cours de cette thèse (fig. 46), mais les données de B. Dadonaite *et al.* [277] comportent plus de nucléotides moyennement réactifs. On retrouve également des différences pour les valeurs extrêmes des différentes conditions : les valeurs extrêmes sont plus élevées dans les données obtenues par B. Dadonaite *et al.* [277] par rapport à celles de cette thèse.

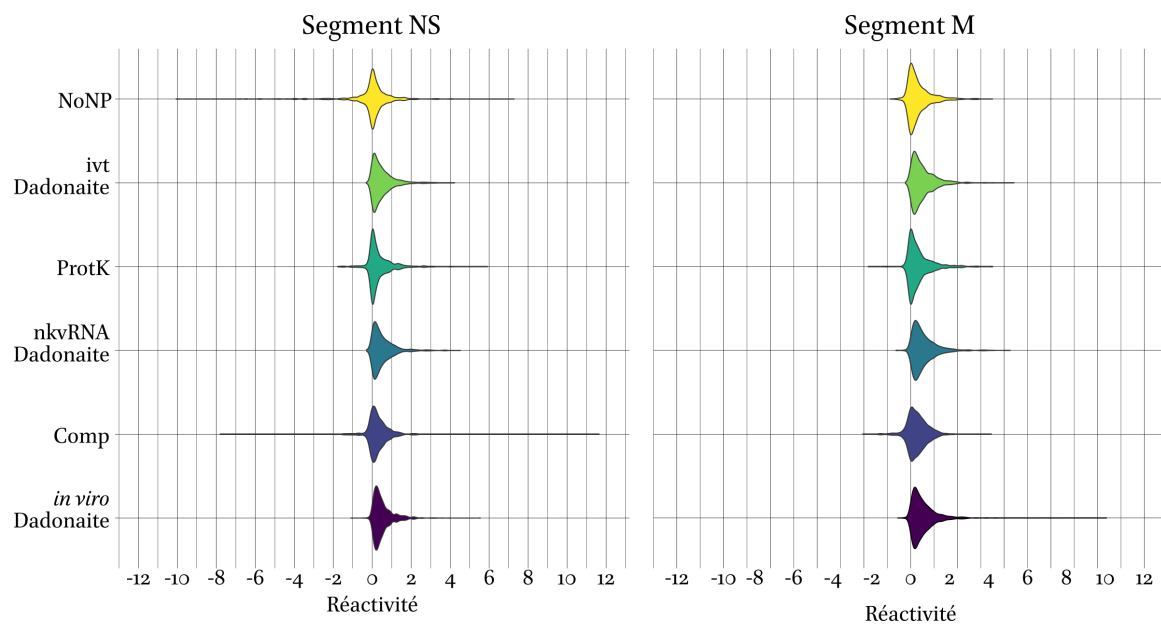


FIGURE 46. – Diagramme en violons des réactivités NS et M avec les données de B. Dadonaite *et al.* [277]. NoNP, ProtK et Comp correspondent aux différentes conditions des données présentées dans cette thèse; ivt Dadonaite correspond aux données de B. Dadonaite obtenues *in vitro*; nkvRNA Dadonaite aux données obtenues après traitement des particules virales à la protéinase K; *in viro* Dadonaite aux données obtenues sur les RNPv.

Comme il y a une certaine variation dans les valeurs de réactivité potentiellement attribuable aux techniques utilisées, les réactivités des nucléotides ont été classifiées en trois catégories : non réactif (entre 0 et 0,4, noté "0"), moyennement réactif (entre 0,4 et 0,8, noté "1") et fortement réactif (supérieur à 0,8, noté "2").

La corrélation entre les différents jeux de données a été testée en calculant le coefficient de corrélation polychorique. Pour la condition « NoNP » on obtient des coefficients de corrélation de 0,44 pour le segment NS et de 0,45 pour le segment M. Les coefficients sont assez bas mais c'est une amélioration notable par rapport aux corrélations des valeurs de réactivité. Dans la condition « ProtK », la corrélation est meilleure pour NS (0,52) mais moins bonne pour M (0,31). Pour ces données, il faut garder en tête que dans l'étude de B. Dadonaite *et al.* [277], elles ont été obtenues à partir de particules virales traitées à la protéinase K alors que dans cette thèse, les données ont été obtenues entièrement *in vitro*. Enfin, il est possible de comparer la condition « Comp » aux données obtenues *in viro* par B. Dadonaite *et al.* [277] bien que nos données aient été obtenues *in vitro* sur des complexes ne comportant pas la polymérase virale. Les coefficients de corrélation sont à 0,24 pour NS et 0,29 pour M, ce qui semble montrer que les nucléoprotéines formées *in vitro* ont une structure qui diffère des nucléoprotéines au sein de la particule virale. Ces comparaisons montrent que bien que les données de B. Dadonaite *et al.* [277] diffèrent très fortement des nôtres, il pourrait s'agir principalement d'un problème d'échelle des réactivités ainsi que d'une plus forte sensibilité de la technique SHAPE-MaP par rapport au SHAPE classique.

Différents signaux d'empaquetage ont été identifiés précédemment dans la littérature sur différents

segments, notamment dans une étude de mutagenèse de la région 27-56 du segment NS de la souche A/WSN/33 (H₁N₁) [236] montrant que des mutations dans cette région ont un effet délétère sur l'incorporation de cet ARN. D'après nos données, cette région qui est impliquée dans une structure secondaire stable en absence de protéine et devient exposée en présence de protéine, ce qui est cohérent avec la possibilité que cette région comporte un signal d'empaquetage (fig. 47). Il serait même possible d'affiner la région identifiée par Fujii *et. al* en la restreignant aux nucléotides exposés par la protéine NP. L'étude par cartographie chimique de la structure des ARN pourrait donc bien nous permettre de détecter de potentiels signaux d'empaquetage.

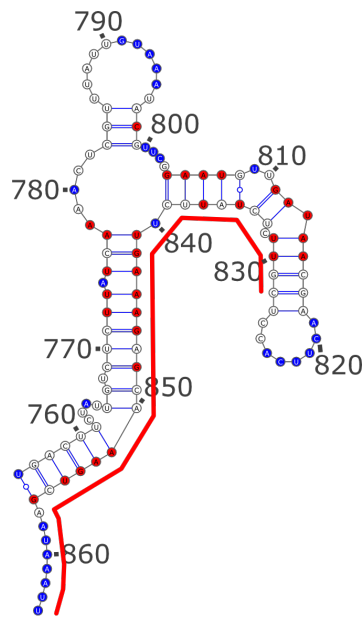


FIGURE 47. – Région de l'ARN NS comportant un site d'empaquetage potentiel d'après la structure prédite avec les données « NoNP ». La région identifiée dans la littérature [236] est représentée par la ligne rouge.

D. Mise en évidence des interactions entre les segments NS et M en présence de la protéine NP

Le modèle communément admis concernant l'empaquetage est celui d'un empaquetage sélectif, au cours duquel s'établissent des interactions entre les signaux d'empaquetage présents au sein des ARNv. Il existe un réseau d'interactions redondantes entre les différents ARNv qui permet la formation d'un complexe supramoléculaire comportant un exemplaire de chaque ARNv qui est ensuite empaqueté dans la particule virale. L'existence d'une de ces interactions a pu être démontré *in cellulo* par le laboratoire [243, 244] et une autre a été identifiée par B. Dadonaite *et al.* [277].

Parmi toutes les interactions possibles, celles qui vont réellement se former vont dépendre de l'accessibilité des signaux d'empaquetage. Celle-ci va dépendre de la fixation de la protéine NP qui se lie à l'ARN en exposant certaines bases et en protégeant d'autres [95].

Jusqu'à présent, la plupart des interactions entre ARNv ont été déterminées *in vitro* en utilisant des ARN nus [243, 244] et il est vraisemblable que certaines des interactions identifiées dans ces conditions ne peuvent pas s'établir entre RNPv. Nous avons donc cherché à savoir si en appliquant la technique de SHAPE à deux ARNv complexés à la protéine NP et en comparant ces données à celles de chaque segment individuellement complexé à la NP, il était possible de mettre en évidence des régions d'interaction entre ces deux segments. Nous avons testé cette idée sur les segments NS et M qui sont capables d'interagir ensemble [279, 363], bien que cette interaction soit faible et n'est pas forcément observée dans toutes les études [278].

Dans un premier temps, les ARNv NS et M ont été repliés indépendamment avant d'être complexés à la protéine NP. Les ARN complexés ont ensuite été incubés ensemble puis ont été analysés sur un gel d'agarose en conditions natives (fig. 48). On observe que le mélange des deux ARN complexés migre à mi-chemin entre les deux bandes des ARN individuels complexés à la NP, ce qui montre une interaction entre les deux, bien que l'observation d'une bande supérieure aux autres était attendue.

Les réactivités SHAPE obtenues ont été traitées de la même manière que précédemment en calculant les pourcentages de variation des valeurs. Nous avons pris comme convention que si l'une des valeurs utilisées pour le calcul du pourcentage était égale à 0, alors la variation serait de +100% ou -100%.

Pour le segment NS, sur 760 nucléotides analysés, on retrouve 63,3% de nucléotides non réactifs avec ou sans le segment M, 24,6% de nucléotides réactifs dont la réactivité varie peu, 3,8% de nucléotides dont la réactivité augmente en présence du segment M et 8,3% de nucléotides dont la réactivité diminue en présence du segment M. Pour le segment M, sur 933 nucléotides analysés, on retrouve 49,8% de nucléotides non réactifs avec ou sans le segment NS, 15,43% de nucléotides réactifs dont la réactivité varie peu, 14,7% de nucléotides dont la réactivité augmente en présence du segment NS et 20% de nucléotides dont la réactivité diminue en présence du segment NS.

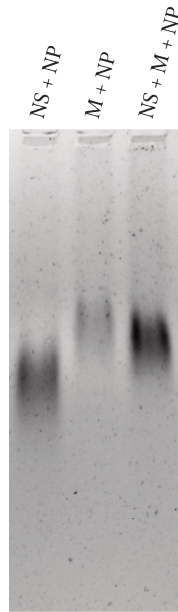


FIGURE 48. – Migration sur gel d'agarose natif 0,8% de 15 cm pendant 4 h à 150 V des ARN NS et M complexés à la protéine NP, soit pour chaque segment individuel, soit pour les deux segments rassemblés.

Une part plus importante de nucléotides voient leur réactivité changer dans le segment M par rapport au segment NS. Les changements de réactivités sont relativement regroupés au sein du segment NS (fig. 49A), tandis qu'ils sont plus uniformément distribués au sein du segment M, sans qu'aucune région ou motif ne se distingue (fig. 49B). Il est donc difficile de distinguer des régions potentiellement impliquées dans des interactions entre les segments NS et M. On peut cependant distinguer une région du segment NS (545-551) où la réactivité diminue sur plusieurs nucléotides successifs et qui serait donc potentiellement une région d'interaction.

Chercher manuellement à travers l'intégralité de la séquence du segment NS les régions comportant un nombre x de nucléotides qui présentent une forte diminution serait une tâche complexe et fastidieuse. C'est pourquoi j'ai développé un script Python afin d'automatiser cette recherche.

Ce script récupère des fichiers FASTA contenant les séquences de chaque segment afin de constituer deux tableaux à deux dimensions (= *DataFrame*), contenant la position du nucléotide, le nucléotide ainsi que la valeur associée à ce nucléotide.

```
def add_dataset(df, file, value):
    for seq_record in SeqIO.parse(file, "fasta"):
        df = pd.DataFrame(index = np.arange(1, len(seq_record.seq)+1), columns =
                           ['Nucleotide', 'Reactivity'])

        x = 1
        for p in seq_record.seq:
            df.at[x, 'Nucleotide'] = p
            x += 1
```

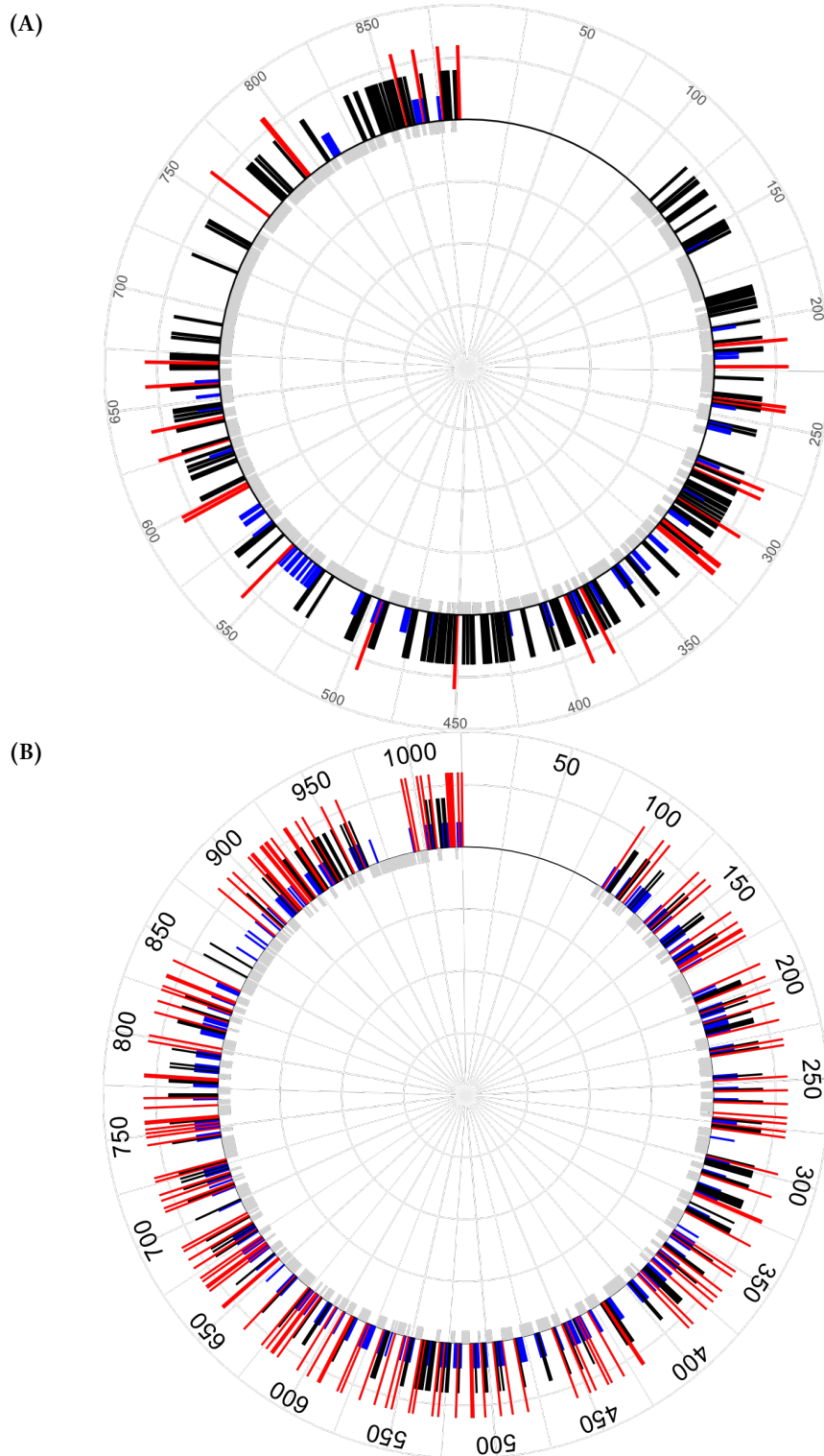


FIGURE 49. – Représentation des changements de réactivité en présence des deux ARN complexés. Les changements sont indiqués par du gris pour des nucléotides non réactifs en présence et en absence du second segment, du noir pour les nucléotides réactifs dont la variation est inférieure à 40%, du bleu pour les nucléotides dont la réactivité diminue d’au moins 40% en présence du second segment et du noir pour les nucléotides dont la réactivité augmente d’au moins 40% en présence du second segment. Les données sont représentées soit pour le segment NS (A) ou pour le segment M (B). La numérotation s’effectue du 3’ au 5’, comme le veut la convention dans le domaine des virus à ARN (-).

```

with open(value) as f:
    for line in csv.reader(f, dialect = "excel-tab"):
        if(line[1] == "" or line[1] == "#N/A"):
            df.at[int(line[0]), 'Reactivity'] = float('NaN')
        else:
            df.at[int(line[0]), 'Reactivity'] = float(line[1])
return df

```

Le script parcourt l'entièreté de la séquence du premier segment *via* une fenêtre glissante dont la taille est déterminée par l'utilisateur. Dans cette fenêtre glissante, le nombre de nucléotides atteignant la valeur recherchée est calculé. Si le nombre de nucléotides atteint la valeur seuil, aussi déterminée par l'utilisateur, alors la séquence est récupérée.

```

def check_value(args, df, seq, i, j):
nb_found = 0
while j < (i + args.window):
    # Generating sequence contained in the windows for sequence A
    seq += df.at[j, 'Nucleotide']

    # Counting the number of nucleotides that respect the threshold given
    if(pd.isna(df.at[j, 'Reactivity'])):
        pass
    else:
        if(float(df.at[j, 'Reactivity']) == args.value_searched):
            nb_found += 1
        j += 1

return seq, nb_found

```

A ce moment, le script parcourt le second segment afin de rechercher la ou les séquence(s) capable(s) de s'apparier à celle identifiée précédemment. Les appariements de type G-U sont pris en compte. Si une séquence complémentaire à celle du premier segment est trouvée, le nombre de nucléotides ayant les conditions recherchées est calculé. Si la valeur seuil est atteinte, les séquences ainsi que les positions sont stockées dans un fichier .txt.

```

def compare_sequence(args, seq_A, seq_B):
c = 1
length = 0
while c <= args.window:
    if(seq_A[c-1] == "A"):
        if(seq_B[c-1] == "U"):
            length += 1
    else:
        pass
    if(seq_A[c-1] == "U"):
        if seq_B[c-1] == "A" or seq_B[c-1] == "G":

```

```

        length += 1
    else:
        pass
    if(seq_A[c-1] == "G"):
        if seq_B[c-1] == "C" or seq_B[c-1] == "U":
            length += 1
        else:
            pass
    if(seq_A[c-1] == "C"):
        if seq_B[c-1] == "G":
            length += 1
        else:
            pass
    c += 1
if(length == args.window):
    return True
else:
    return False

```

Le script peut également préparer des fichiers formatés pour pouvoir être utilisés avec R et la librairie *circize* [364] afin de générer des diagrammes à cordes. Le script est disponible à l'adresse suivante : <https://github.com/equignon> Il a ainsi été possible d'analyser facilement les résultats obtenus. En comparant les données obtenues sur le segment NS et le segment M, plusieurs régions d'interactions ARN-ARN potentielles ont été identifiées, allant de 4 à 7 nucléotides (fig. 50).

De manière surprenante, la région du segment NS pour laquelle on voit une diminution conséquente de la réactivité (425-475) n'apparaît pas dans la liste des interactions potentielles identifiées : ce serait donc une région dont la structure est modifiée suite à l'interaction entre les deux ARN. La région la plus longue avec le plus grand nombre de nucléotides où l'on observe des diminutions franches de réactivité est une région de 7 nucléotides, allant de la position 341 à 347 du segment NS (sens 3' -> 5') "3'-GUACGGG-5'" et pouvant interagir avec la séquence allant de la position 545 à 551 du segment M (sens 3' -> 5') "3'-CAUGUCU-5'". Les deux séquences seraient parallèles dans cette configuration [365]. Afin de vérifier si une interaction existe entre ces deux régions, il faudrait soit induire des mutations dans cette région, soit utiliser des oligonucléotides les ciblant. Bien que cette interaction entre les segments NS et M n'ait pas été validée par manque de temps, ces résultats sont prometteurs dans l'optique d'une recherche fine des interactions entre les différents ARNv, en permettant de rapidement déterminer quelles sont les régions les plus susceptibles d'interagir ensemble. Il est également intéressant de noter que plusieurs régions avec des changements de réactivité sont observées; il est donc possible que les interactions entre deux segments modifient légèrement la structure de l'ARN même si ce dernier est lié à la protéine NP.

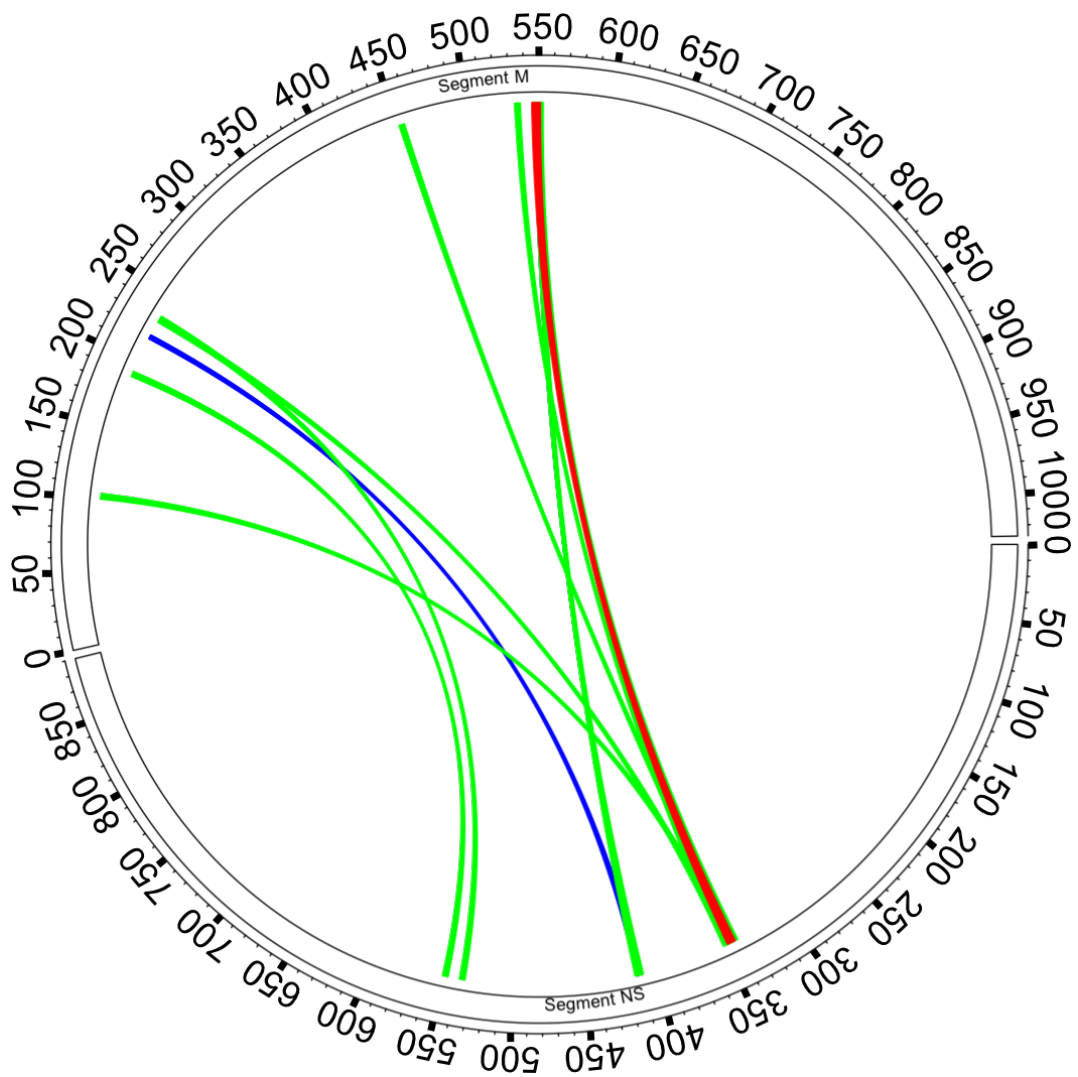


FIGURE 50. – Représentation des interactions potentielles entre les segments NS et M. Les arcs bleus représentent les interactions de 4 nucléotides (au moins 3 nucléotides avec une diminution importante de réactivité), les arcs verts les interactions de 5 nucléotides (avec au moins 3 nucléotides avec une diminution importante de réactivité) et l'arc rouge une interaction de 7 nucléotides (avec au moins 4 nucléotides avec une diminution importante de réactivité.). La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

E. Microscopie électronique

Afin d'observer la formation des ribonucléoprotéines *in vitro*, une collaboration avec l'équipe du Dr. Thibaut Crépin (IBS, Grenoble) a été établie. Les complexes ont été formés selon le même protocole que lors de la modification des ARN, dans des conditions favorisant soit la forme monomérique de la protéine, soit la forme trimérique.

Deux concentrations de protéine (2 mg/mL et 5 mg/mL) et deux concentrations de sels (50 mM de NaCl et 150 mg/mL de NaCl) ont été testées. Après centrifugation et récupération des culots, seuls quelques agrégats sont visibles (fig.51A-D). Comme contrôle, 6 mg de protéine NP ont été incubés en présence d'un dodécamère (UC)₆ porteur d'un fluorophore FAM à l'extrémité 3' en présence de 150 mM de NaCl. Comme précédemment montré par l'équipe de Thibaut Crépin [257], après centrifugation et récupération du culot, de très long filaments sont visibles (fig.51E). Ces structures sont de bien plus grande taille que les ribonucléoprotéines et sont dues à l'oligomérisation de la protéine NP en présence d'ARN 12-mer porteur d'un fluorophore.

N'ayant pas observé, à partir de nos échantillons, de complexes dans les culots, nous avons mesuré l'absorbance à 260 et 280 nm des culots remis en suspension et des surnageants de la centrifugation pour attester de la présence ou de l'absence de l'ARN et des protéines dans les échantillons (table 8). Il en ressort qu'à l'exception de la condition contrôle et de la condition 5 mg/mL de protéine avec 50 mM NaCl, l'essentiel de la densité optique à 260 et 280 nm est retrouvée dans les surnageants. Contrairement aux complexes formés avec le dodécamère fluorescent, il semble que les complexes entre la NP et l'ARN NS soient solubles.

	Longueur d'onde	Culot	Surnageant
A/NP 2 mg/mL + NS 50 mM NaCl	260 nm	0,193	8,567
	280 nm	0,125	5,59
A/NP 2 mg/mL + NS 150 mM NaCl	260 nm	0,332	7,821
	280 nm	0,206	5,172
A/NP 5 mg/mL + NS 50 mM NaCl	260 nm	5,103	3,087
	280 nm	5,166	2,335
A/NP 5 mg/mL + NS 150 mM NaCl	260 nm	1,093	9,423
	280 nm	1,016	8,395
A/NP 6 mg/mL + 12-mer polyUC-FAM	260 nm	18,252	8,474
	280 nm	15,325	7,711

TABLEAU 8. – Quantification du matériel biologique présent dans les surnageants et les culots après centrifugation.

Pour des raisons d'accès au microscope, les surnageants ont été conservés à 4°C pendant une semaine puis analysés en microscopie électronique. Le premier échantillon (fig. 52A) est trop concentré et il n'est pas possible de distinguer des structures individuelles. Le deuxième échantillon (fig. 52B) quant à lui

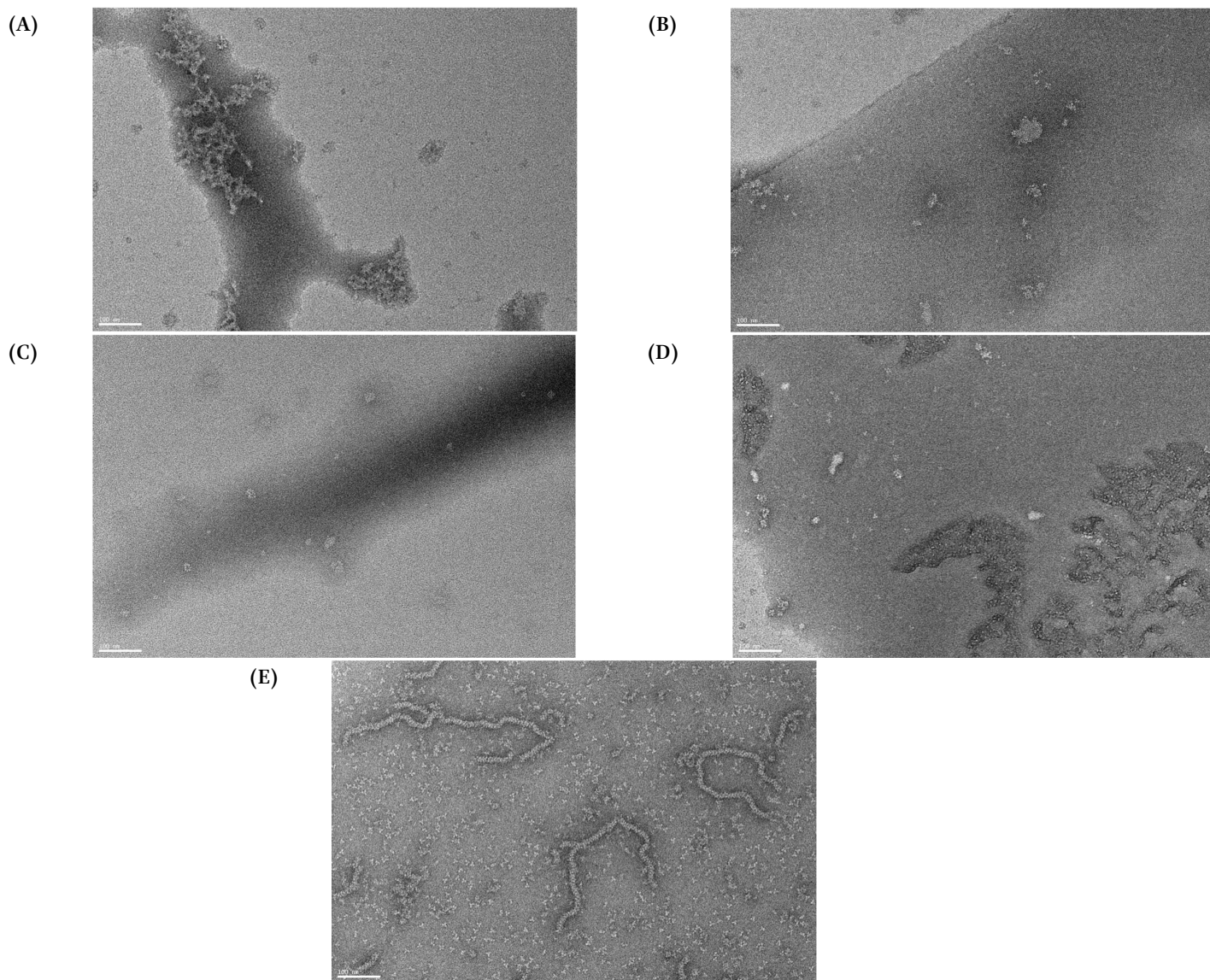


FIGURE 51. – Images capturées par microscopie électronique des complexes ARN NS-NP après centrifugation et récupération des culots. (A) ARN en présence d'environ 2 mg/mL de protéine et 50 mM NaCl, (B) ARN NS en présence d'environ 2 mg/mL de protéine et 150 mM NaCl, (C) ARN NS en présence d'environ 5 mg/mL de protéine et 50 mM NaCl, (D) ARN NS en présence d'environ 5 mg de protéine et 150 mM NaCl, (E) contrôle avec le dodécamère (UC)₆-3'FAM en présence d'environ 6 mg/mL de protéine et 150 mM NaCl.

montre principalement des anneaux de protéine NP ainsi que des trimères de protéine. Quelques structures pourraient évoquer des ribonucléoprotéines, cependant il est difficile de distinguer s'il s'agit d'anneaux et de bruit de fond ou bien de véritables nucléoprotéines. Un échantillon plus dilué fraîchement préparé apporterait sans doute plus d'informations. Il serait également intéressant d'utiliser un autre ARN que le segment NS, afin d'obtenir des ribonucléoprotéines d'une plus grande taille. L'un des culots de l'expérience précédente avait également été passé de nouveau au microscope électronique (fig. 52C) comme contrôle. Comme lors de l'expérience précédente, très peu de matériel est visible, ce qui confirme que les ribonucléoprotéines formées avec l'ARN NS sont solubles.

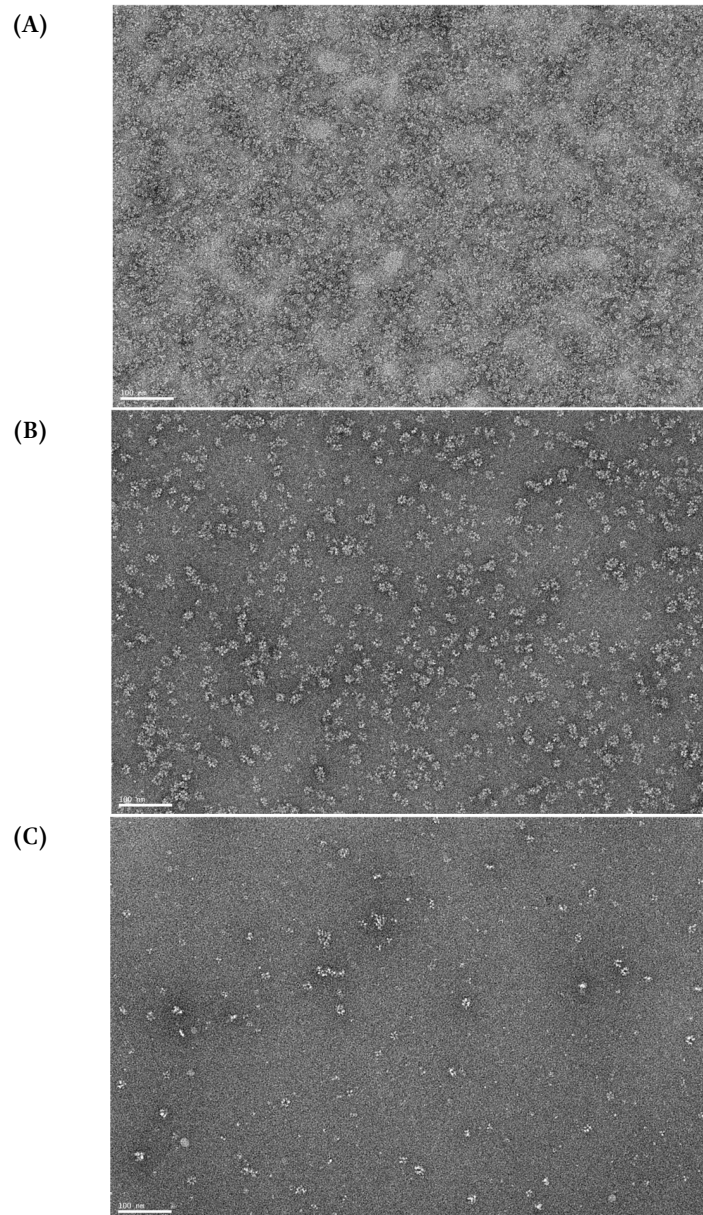


FIGURE 52. – Images en microscopie électronique des complexes ARN NS-NP après centrifugation et récupération des surnageants (A) et (B) ou des culots (C). (A) ARN NS en présence de 2 mg/mL de protéine NP et 150 mM de NaCl, (B) ARN NS en présence de 5 mg/mL de protéine NP et 50 mM NaCl, (C) ARN NS en présence de 5 mg/mL de protéine NP et 50 mM NaCl.

V | Discussion, perspectives et conclusions

Au cours de cette thèse, j'ai travaillé sur la structure des ARN viraux des virus *Influenza A*. En utilisant des techniques de cartographie chimique, nous avons étudié la fixation de la protéine NP sur les ARN NS et M *in vitro*. Un ensemble de trois réactifs (NMIA, DMS et EDC) a été utilisé afin d'essayer de discriminer les régions en interaction avec l'ARN de celles liées à la protéine NP.

La première partie de ce travail a consisté en l'optimisation de la modification des ARN avec les différents réactifs. L'amélioration des conditions de rétrotranscription a permis d'obtenir des résultats plus reproductibles tout en réduisant la quantité de réactif nécessaire pour le réactif NMIA. L'obtention des données complètes avec le DMS n'a pu être réalisée, cependant les résultats préliminaires semblent montrer que les données au DMS peuvent être recoupées avec celles obtenues avec le NMIA. L'EDC a posé plus de soucis, montrant que les modifications sont bien présentes sur l'ARN, mais le bruit de fond est trop important pour que les données soient exploitables à ce stade. L'optimisation des conditions de modification devrait permettre de faire ressortir les modifications par rapport au bruit de fond, ce qui améliorerait l'analyse et devrait rendre les données exploitables. Bien qu'il reste des optimisations à réaliser, ces résultats préliminaires sont prometteurs concernant l'utilisation des trois réactifs en combinaison afin de pouvoir discriminer les régions d'interaction ARN-ARN et ARN-NP.

La seconde partie a consisté en l'amélioration de la production de NP telle qu'établie par D. Ferhadian [98]. L'utilisation de bactéries BL21 DE3 pLysS, capables de décoder des codons rares lui avaient permis d'augmenter les rendements de production, cependant ces dernières manquaient d'homogénéité. L'ajout d'étapes supplémentaires de purification, notamment l'ajout d'une colonne de gel filtration, a permis d'améliorer l'homogénéité et les rendements de production de la protéine. L'application des techniques de cartographie chimique aux ARN liés à la protéine NP a nécessité quelques ajustements mais la possibilité d'utiliser les réactifs NMIA et DMS sur les complexes ARN-NP a pu être montrée. Il reste encore à acquérir les données avec le réactif DMS, ainsi qu'à mettre au point les modifications avec l'EDC. Avec les avancées récentes des techniques de cryo-microscopie électronique, il pourrait être intéressant d'observer les complexes ARN-NP avec ces techniques, que ce soient les ribonucléoprotéines séparées ou bien en interaction les unes avec les autres.

Les difficultés rencontrées au cours de cette thèse ont cependant empêché la mise au point complète du DMS et de l'EDC et l'acquisition des données par ces méthodes. Les conditions pour l'utilisation du DMS ont été déterminées, cependant un travail important reste à mener avec l'EDC, pour lequel des résultats préliminaires encourageants ont été obtenus mais n'ont pas pu être reproduits. Il est possible

qu'une dégradation du réactif se soit produite au cours du temps et que celui-ci ait perdu en efficacité.

L'hypothèse selon laquelle la protéine NP aurait une activité chaperon avait été émise avant le début de cette thèse. En considérant les deux définitions possibles de cette activité chaperon, deux conclusions peuvent en être tirées. Si l'on se base sur la définition d'origine, la plus stricte qui stipule qu'une protéine chaperon de l'ARN se fixe de façon transitoire afin de remodeler un ARN, il faut se baser sur les données de la comparaison NoNP-ProtK, qui comporte l'enlèvement de la protéine. Dans ce cas de figure, le fait que la NP ne soit pas retirée de l'ARN viral durant la quasi-totalité du cycle viral, ainsi que le peu, voir l'absence, de modifications structurales persistant après l'enlèvement de la NP, démontrent une absence d'activité chaperon de la protéine NP. Si l'on se réfère à la définition plus large, qui n'inclue pas la notion de fixation transitoire, alors la comparaison à prendre en compte est celle entre les conditions « NoNP » et « Comp », du fait que la NP n'est, à l'exception des étapes de transcription et de réplication, pas retirée de l'ARN. Dans ce cas de figure, on observe clairement des modifications structurales qui se prononcent en faveur d'un effet chaperon analogue à celui d'autres protéines chaperons virales.

Les données obtenues au cours de cette thèse sont en grande partie cohérentes avec les données publiées dans la littérature. Bien que des différences soient visibles avec les données issues de SHAPE-MaP [277], ces différences proviennent probablement d'une sensibilité plus grande de l'approche MaP. Il faut cependant remarquer que dans cette étude des triplicats n'ont été obtenus que pour les RNPv et que certains jeux de données ne remplissent pas tous les critères de qualité définis par SHAPMapper 2 [366]. Par ailleurs, la comparaison des réactivités avec un signal d'empaquetage déjà identifié dans la littérature [236] montre que ce signal se trouve dans une région identifiée comme simple brin et exposée par la fixation de la protéine NP, ce qui montre que l'utilisation de la cartographie chimique pour déterminer les signaux d'empaquetage des ARNv est prometteuse. En outre, l'utilisation de la cartographie chimique a permis de mettre en évidence une potentielle interaction entre deux ARNv complexés à la NP. Bien que le temps ait manqué afin de confirmer l'existence de cette interaction à l'aide d'oligonucléotides antisens ou par mutagenèse, ces premiers résultats sont prometteurs et représentent une avancée importante par rapport à l'identification d'interactions en ARNv nus initialement développée au laboratoire [243, 244], car il existe de plus en plus d'éléments indiquant que la protéine NP module les interactions entre ARNv [252]. De manière intéressante, nous avons observé plusieurs changements structuraux importants dans des régions qui ne sont, à priori, pas en interaction; il est donc possible que les interactions entre les RNPv modifient l'exposition de certaines des bases, ce qui pourrait masquer ou exposer des signaux d'empaquetage. Dans ce cas, il est envisageable que l'assemblage des huit RNPv en complexe supramoléculaire soit séquentiel et coopératif, comme semble l'indiquer la littérature [250].

Les perspectives pour la poursuite de ce travail peuvent être classées en deux catégories : *in vitro* et *in viro*. Les premières seront plus destinées à étudier des mécanismes précis ou difficiles à étudier dans un contexte viral ou cellulaire, tandis que les secondes seront destinées à des études de plus grande échelle.

Plusieurs structure secondaires de l'ARN ont pu être identifiées comme étant conservées au sein de

plusieurs souches virales [295]. Il serait donc possible via des études *in vitro* d'essayer d'étudier la conservation de certaines structures entre des souches de différents sous-types. L'importance des structures des ARN pour leur fonction est connue et des régions identifiées comme exposées au cours des études pourraient jouer un rôle aussi bien de signal d'emballage que d'autres rôles. La dynamique de la structure de l'ARN lors de la formation de la RNPv au cours de synthèse de l'ARN serait un sujet d'étude intéressant car il est possible d'hypothétiser l'existence de structures transitoires qui se formeraient au cours de la synthèse qui serviraient de signaux d'initiation de la liaison de la protéine NP à l'ARN.

Les expériences menées au cours de cette thèse l'ont été en présence de l'ARN viral et de la protéine NP. Cependant une RNPv comporte également le complexe réplicase (PB2, PB1 et PA). Il est possible d'envisager que la fixation du complexe réplicase à l'extrémité de la RNPv entraîne des changements de la structure de l'ARN, ce qui par conséquent, pourrait modifier les signaux d'emballage disponibles.

La possibilité que la formation du complexe supramoléculaire soit séquentielle et les possibles conséquences pourraient également être étudiées *in vitro*. En se basant sur une souche dont le réseau d'interaction a été décrit, il serait intéressant d'incuber les ARN en les rajoutant l'un après l'autre afin de voir si l'ajout d'un troisième ou quatrième segment modifie à nouveau la structure de l'ARN.

Les données obtenues à l'aide des techniques de cartographie chimique apportent des informations 2D, qui pourraient être complétées par l'apport de données sur la structure 3D des ARNv. Pour cela, l'emploi de techniques telles le MOHCA-Seq serait une piste intéressante [367].

La méthode développée ici *in vitro* afin de déterminer et discriminer les régions en interaction avec l'ARN ou la protéine NP pourrait également être appliquée aux particules virales intactes et désassemblées. Cela permettrait (1) de voir les différences entre les approches *in vitro* et *in viro* et (2) d'identifier des interactions présentes uniquement au sein des particules virales. Ce développement d'approches *in viro*, associé au développement de techniques à haut débit permettra d'obtenir des données sur le génome complet, en absence ou en présence des autres segments. Il sera donc plus facile de comparer l'état des réseaux d'interaction entre les différentes souches de virus *Influenza A*, afin de déterminer précisément les positions impliquées dans des interactions inter-RNPv.

Le développement des techniques de cryo-microscopie électronique sur l'analyse des structures ARN [368] pourrait permettre, en l'appliquant à l'ARN ainsi qu'aux RNPv individuelles ou complexées, l'obtention d'informations sur la structure 3D de l'ARN au sein des particules virales, complétant ainsi les données 2D obtenues à l'aide des techniques de cartographie chimique.

Un défi est posé par l'analyse et l'interprétation des données. Avec un génome totalisant 13 588 nucléotides, huit ARN génomiques et par conséquent un grand nombre d'interactions intra- ou inter-segments, la complexité de l'analyse rendrait difficile l'interprétation des données. A cette fin, le développement d'un logiciel de prédiction d'interactions à longue distance sera nécessaire afin d'obtenir une analyse robuste et complète des données.

A l'instar du script que j'ai développé au cours de ma thèse, il devrait être possible en comparant des diminutions de réactivité au sein de certaines régions, d'estimer si une interaction entre un ou plusieurs ARN peut s'y réaliser. L'étude de ces régions d'interaction, une fois confirmée, pourra permettre de

déterminer si certains motifs ou certains biais sont présents dans les différentes régions impliquées dans des interactions entre RNPv.

Cette analyse pourrait être complétée par une analyse comparative des données NMIA, DMS et EDC. Une fois ces données normalisées et comparées, la détermination des régions ayant des différences significatives permettrait d'estimer les sites de fixation à la protéine NP. En cumulant cette analyse comparative à la prédiction des interactions inter-RNPv, il serait donc possible de dresser une carte des interactions ARN-ARN et ARN-NP au sein d'une souche virale d'Influenza A.

En intégrant l'ensemble de ces analyses dans un *pipeline*, il serait possible de grandement simplifier et automatiser l'analyse de la fixation de la protéine NP et des interactions inter-RNPv. Cela permettrait de tester un nombre plus important de souches virales différentes afin de collecter une quantité massive d'informations sur la structure des RNPv des virus Influenza, ouvrant la porte à une compréhension beaucoup plus fine du mécanisme d'empaquetage sélectif de ces virus.

Ce *pipeline* pourrait également être adapté à l'étude de la structure et de la fixation d'une ou plusieurs protéines virales sur d'autres virus.

Afin de confirmer la fonctionnalité des interactions ARN-ARN entre différents ARNv, deux méthodes ont été majoritairement employées jusqu'à présent : l'utilisation d'oligonucléotides antisens et la mutation des régions potentiellement impliquées. Il serait également possible de tirer parti des techniques d'études de l'interaction directe entre ARN [369–371] pour confirmer l'existence fonctionnelle de ces interactions.

Une compréhension plus fine des règles qui régissent les interactions entre segments permettrait de prédire avec une efficacité accrue les différents réassortants les plus susceptibles d'être générés entre plusieurs souches virales. Comprendre la fixation de la NP aux ARNv et la formation des RNPv apportera des informations complémentaires à ce sujet, la NP pouvant exposer ou masquer des signaux d'empaquetage. En intégrant l'ensemble de ces données dans des modèles prédictifs, il devrait être possible de développer des outils performants pour la prévention des prochaines épidémies. Plusieurs outils ont déjà été développés qui visent à prédire les virus réassortants qui pourraient devenir pandémiques [372–376] mais n'incluent pas à ce jour d'informations sur l'empaquetage des RNPv, ce dont ils pourraient grandement bénéficier.

Bibliographie

- [1] N. M. Bouvier and P. Palese, “The biology of Influenza viruses,” *Vaccine*, vol. 26, pp. D49–D53, Sept. 2008.
- [2] “Influenza A subtypes and the species affected | seasonal Influenza (Flu) | CDC.” <https://www.cdc.gov/flu/other/animal-flu.html>, Sept. 2018.
- [3] S. Tong, Y. Li, P. Rivaille, C. Conrardy, D. A. A. Castillo, L.-M. Chen, S. Recuenco, J. A. Ellison, C. T. Davis, I. A. York, A. S. Turmelle, D. Moran, S. Rogers, M. Shi, Y. Tao, M. R. Weil, K. Tang, L. A. Rowe, S. Sammons, X. Xu, M. Frace, K. A. Lindblade, N. J. Cox, L. J. Anderson, C. E. Rupprecht, and R. O. Donis, “A distinct lineage of Influenza A virus from bats,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, pp. 4269–4274, Mar. 2012.
- [4] S. Tong, X. Zhu, Y. Li, M. Shi, J. Zhang, M. Bourgeois, H. Yang, X. Chen, S. Recuenco, J. Gomez, L.-M. Chen, A. Johnson, Y. Tao, C. Dreyfus, W. Yu, R. McBride, P. J. Carney, A. T. Gilbert, J. Chang, Z. Guo, C. T. Davis, J. C. Paulson, J. Stevens, C. E. Rupprecht, E. C. Holmes, I. A. Wilson, and R. O. Donis, “New world bats harbor diverse Influenza A viruses,” *PLoS Pathog*, vol. 9, p. e1003657, Oct. 2013.
- [5] D. Dou, R. Revol, H. Östbye, H. Wang, and R. Daniels, “Influenza A Virus cell entry, replication, virion assembly and movement,” *Front. Immunol.*, vol. 9, p. 1581, July 2018.
- [6] W. Hao, L. Wang, and S. Li, “Roles of the non-structural proteins of Influenza A virus,” *Pathogens*, vol. 9, p. 812, Oct. 2020.
- [7] P. J. Walker, S. G. Siddell, E. J. Lefkowitz, A. R. Mushegian, E. M. Adriaenssens, P. Alfenas-Zerbini, A. J. Davison, D. M. Dempsey, B. E. Dutilh, M. L. García, B. Harrach, R. L. Harrison, R. C. Hendrickson, S. Junglen, N. J. Knowles, M. Krupovic, J. H. Kuhn, A. J. Lambert, M. Łobocka, M. L. Nibert, H. M. Oksanen, R. J. Orton, D. L. Robertson, L. Rubino, S. Sabanadzovic, P. Simmonds, D. B. Smith, N. Suzuki, K. Van Doorslaer, A.-M. Vandamme, A. Varsani, and F. M. Zerbini, “Changes to virus taxonomy and to the international code of virus classification and nomenclature ratified by the International Committee on Taxonomy of Viruses (2021),” *Arch Virol*, vol. 166, pp. 2633–2648, Sept. 2021.
- [8] A. D. M. E. Osterhaus, G. F. Rimmelzwaan, B. E. E. Martina, T. M. Bestebroer, and R. A. M. Fouchier, “Influenza B virus in seals,” *Science*, vol. 288, pp. 1051–1053, May 2000.
- [9] R. A. Elderfield, M. Koutsakos, R. Frise, K. Bradley, J. Ashcroft, S. Miah, A. Lackenby, and W. S. Barclay, “NB protein does not affect Influenza B virus replication in vitro and is not required for replication in or transmission between ferrets,” *J Gen Virol*, vol. 97, pp. 593–601, Mar. 2016.

- [10] J. A. Mould, R. G. Paterson, M. Takeda, Y. Ohigashi, P. Venkataraman, R. A. Lamb, and L. H. Pinto, "Influenza B virus BM2 protein has ion channel activity that conducts protons across membranes," *Developmental Cell*, vol. 5, pp. 175–184, July 2003.
- [11] C. C. Sreenivasan, Z. Sheng, D. Wang, and F. Li, "Host range, biology, and species specificity of seven-segmented Influenza viruses - a comparative review on Influenza C and D," *Pathogens*, vol. 10, p. 1583, Dec. 2021.
- [12] S. Hongo, K. Sugawara, H. Nishimura, Y. Muraki, F. Kitame, and K. . Nakamura, "Identification of a second protein encoded by Influenza C virus RNA segment 6," *Journal of General Virology*, vol. 75, pp. 3503–3510, Dec. 1994.
- [13] S. Hongo, K. Sugawara, Y. Muraki, F. Kitame, and K. Nakamura, "Characterization of a second protein (CM2) encoded by RNA segment 6 of Influenza C virus.," *Journal of Virology*, vol. 71, pp. 2786–2792, Apr. 1997.
- [14] T. Furukawa, Y. Muraki, T. Noda, E. Takashita, R. Sho, K. Sugawara, Y. Matsuzaki, Y. Shimotai, and S. Hongo, "Role of the CM2 protein in the Influenza C virus replication cycle," *J Virol*, vol. 85, pp. 1322–1329, Feb. 2011.
- [15] T. Goto, Y. Shimotai, Y. Matsuzaki, Y. Muraki, R. Sho, K. Sugawara, and S. Hongo, "Effect of phosphorylation of CM2 protein on Influenza C virus replication," *J Virol*, vol. 91, Oct. 2017.
- [16] B. M. Hause, M. Ducatez, E. A. Collin, Z. Ran, R. Liu, Z. Sheng, A. Armien, B. Kaplan, S. Chakravarty, A. D. Hoppe, R. J. Webby, R. R. Simonson, and F. Li, "Isolation of a novel swine Influenza virus from Oklahoma in 2011 which is distantly related to Human Influenza C viruses," *PLoS Pathog*, vol. 9, Feb. 2013.
- [17] S. Su, X. Fu, G. Li, F. Kerlin, and M. Veit, "Novel Influenza D virus : Epidemiology, pathology, evolution and biological characteristics," *Virulence*, vol. 8, pp. 1580–1591, Aug. 2017.
- [18] C. M. Trombetta, "Influenza D virus : A potential threat for humans?," *European Journal of Public Health*, vol. 30, p. ckaa165.1097, Sept. 2020.
- [19] "Orthomyxoviridae," in *Fenner's Veterinary Virology*, pp. 389–410, Elsevier, 2017.
- [20] W. N. Batts, S. E. LaPatra, R. Katona, E. Leis, T. F. F. Ng, M. S. Briec, R. B. Breyta, M. K. Purcell, C. M. Conway, T. B. Waltzek, E. Delwart, and J. R. Winton, "Molecular characterization of a novel orthomyxovirus from rainbow and steelhead trout (*Oncorhynchus mykiss*)," *Virus Research*, vol. 230, pp. 38–49, Feb. 2017.
- [21] S. Sameroff, R. Tokarz, K. Jain, A. Oleynik, C. V. Carrington, W. I. Lipkin, and C. A. Oura, "Novel Quaranjavirus and other viral sequences identified from ticks parasitizing hunted wildlife in Trinidad and Tobago," *Ticks and Tick-borne Diseases*, vol. 12, p. 101730, July 2021.

- [22] A. B. Allison, J. R. Ballard, R. B. Tesh, J. D. Brown, M. G. Ruder, M. K. Keel, B. A. Munk, R. M. Mickley, S. E. J. Gibbs, A. P. A. Travassos da Rosa, J. C. Ellis, H. S. Ip, V. I. Shearn-Bochsler, M. B. Rogers, E. Ghedin, E. C. Holmes, C. R. Parrish, and C. Dwyer, "Cyclic avian mass mortality in the northeastern United States is associated with a novel Orthomyxovirus," *J Virol*, vol. 89, pp. 1389–1403, Jan. 2015.
- [23] M. Morse, A. Marriott, and P. Nuttall, "The glycoprotein of Thogoto virus (a tick-borne orthomyxo-like virus) is related to the Baculovirus glycoprotein GP64," *Virology*, vol. 186, pp. 640–646, Feb. 1992.
- [24] P. Mohr, M. Crane, J. Hoad, L. Williams, D. Cummins, M. Neave, B. Shiell, G. Beddome, W. Michalski, G. Peck, F. Samsing, J. Wynne, S. Cramer, A. Hyatt, and N. Moody, "Pilchard Orthomyxovirus (POMV). I. Characterisation of an emerging virus isolated from pilchards *Sardinops sagax* and Atlantic salmon *Salmo salar*," *Dis. Aquat. Org.*, vol. 139, pp. 35–50, Apr. 2020.
- [25] F. Samsing, H. K. Tengesdal, A. Nylund, M. Rigby, K. H. Wiltshire, R. N. Morrison, S. Godwin, C. Giles, T. Wilson, P. G. Mohr, J. Hoad, and J. W. Wynne, "Development of a new real-time PCR for the detection of pilchard orthomyxovirus (POMV) in apparently healthy fish," *Aquaculture*, vol. 547, p. 737404, Jan. 2022.
- [26] A. V. Bussetti, G. Palacios, A. Travassos da Rosa, N. Savji, K. Jain, H. Guzman, S. Hutchison, V. L. Popov, R. B. Tesh, and W. I. Lipkin, "Genomic and antigenic characterization of Jos virus," *Journal of General Virology*, vol. 93, pp. 293–298, Feb. 2012.
- [27] M. Redlberger-Fritz, M. Kundi, and T. Popow-Kraupp, "Heterogeneity of circulating Influenza viruses and their impact on Influenza virus vaccine effectiveness during the Influenza seasons 2016/17 to 2018/19 in Austria," *Front Immunol*, vol. 11, Mar. 2020.
- [28] H. Zaraket, A. C. Hurt, B. Clinch, I. Barr, and N. Lee, "Burden of Influenza B virus infection and considerations for clinical management," *Antiviral Research*, vol. 185, p. 104970, Jan. 2021.
- [29] S. Boivin, S. Cusack, R. W. Ruigrok, and D. J. Hart, "Influenza A virus Polymerase : Structural insights into replication and host adaptation mechanisms," *Journal of Biological Chemistry*, vol. 285, pp. 28411–28417, Sept. 2010.
- [30] M. F. Boni, "Vaccination and antigenic drift in influenza," *Vaccine*, vol. 26, pp. C8–C14, July 2008.
- [31] U. Desselberger, K. Nakajima, P. Alfino, F. S. Pedersen, W. A. Haseltine, C. Hannoun, and P. Palese, "Biochemical evidence that "new" Influenza virus strains in nature may arise by recombination (reassortment).," *Proc Natl Acad Sci U S A*, vol. 75, pp. 3341–3345, July 1978.
- [32] H. Kim, R. G. Webster, and R. J. Webby, "Influenza virus : Dealing with a drifting and shifting pathogen," *Viral Immunology*, vol. 31, pp. 174–183, Mar. 2018.
- [33] S.-W. Yoon, R. J. Webby, and R. G. Webster, "Evolution and ecology of Influenza A viruses," in *Influenza Pathogenesis and Control - Volume I* (R. W. Compans and M. B. A. Oldstone, eds.), Current Topics in Microbiology and Immunology, pp. 359–375, Cham : Springer International Publishing, 2014.

- [34] J. S. Long, B. Mistry, S. M. Haslam, and W. S. Barclay, "Host and viral determinants of Influenza A virus species specificity," *Nat Rev Microbiol*, vol. 17, pp. 67–81, Feb. 2019.
- [35] A. J. Thompson and J. C. Paulson, "Adaptation of Influenza viruses to human airway receptors," *Journal of Biological Chemistry*, vol. 296, p. 100017, Jan. 2021.
- [36] I. Kosik and J. W. Yewdell, "Influenza hemagglutinin and neuraminidase : Yin–yang proteins coevolving to thwart immunity," *Viruses*, vol. 11, p. 346, Apr. 2019.
- [37] M. Matrosovich, G. Herrler, and H. D. Klenk, "Sialic acid receptors of viruses," *SialoGlyco Chemistry and Biology II*, vol. 367, pp. 1–28, July 2013.
- [38] J. M. Nicholls, A. J. Bourne, H. Chen, Y. Guan, and J. M. Peiris, "Sialic acid receptor detection in the human respiratory tract : Evidence for widespread distribution of potential binding sites for human and avian Influenza viruses," *Respir Res*, vol. 8, no. 1, p. 73, 2007.
- [39] J. J. Skehel and D. C. Wiley, "Receptor binding and membrane fusion in virus entry : The Influenza hemagglutinin," *Annu. Rev. Biochem.*, vol. 69, pp. 531–569, June 2000.
- [40] E. K. Subbarao, W. London, and B. R. Murphy, "A single amino acid in the PB2 gene of Influenza A virus is a determinant of host range," *J Virol*, vol. 67, pp. 1761–1764, Apr. 1993.
- [41] J. S. Long, E. S. Giotis, O. Moncorgé, R. Frise, B. Mistry, J. James, M. Morisson, M. Iqbal, A. Vignal, M. A. Skinner, and W. S. Barclay, "Species difference in ANP32A underlies Influenza A virus polymerase host restriction," *Nature*, vol. 529, pp. 101–104, Jan. 2016.
- [42] L. Carrique, H. Fan, A. P. Walker, J. R. Keown, J. Sharps, E. Staller, W. S. Barclay, E. Fodor, and J. M. Grimes, "Host ANP32A mediates the assembly of the Influenza virus replicase," *Nature*, vol. 587, pp. 638–643, Nov. 2020.
- [43] A. R. Camacho-Zarco, S. Kalayil, D. Maurin, N. Salvi, E. Delaforge, S. Milles, M. R. Jensen, D. J. Hart, S. Cusack, and M. Blackledge, "Molecular basis of host-adaptation interactions between influenza virus polymerase PB2 subunit and ANP32A," *Nat Commun*, vol. 11, July 2020.
- [44] S. E. Galloway, B. Liang, and D. A. Steinhauer, "Activation of the hemagglutinin of Influenza viruses," in *Activation of Viruses by Host Proteases* (E. Böttcher-Friebertshäuser, W. Garten, and H. D. Klenk, eds.), pp. 3–26, Cham : Springer International Publishing, 2018.
- [45] J. Veits, S. Weber, O. Stech, A. Breithaupt, M. Gräber, S. Gohrbandt, J. Bogs, J. Hundt, J. P. Teifke, T. C. Mettenleiter, and J. Stech, "Avian Influenza virus hemagglutinins H2, H4, H8, and H14 support a highly pathogenic phenotype," *Proc Natl Acad Sci USA*, vol. 109, pp. 2579–2584, Feb. 2012.
- [46] L. Jia, J. Xie, J. Zhao, D. Cao, Y. Liang, X. Hou, L. Wang, and Z. Li, "Mechanisms of severe mortality-associated bacterial co-infections following Influenza virus infection," *Front. Cell. Infect. Microbiol.*, vol. 7, p. 338, Aug. 2017.

- [47] H. M. Rowe, V. A. Meliopoulos, A. Iverson, P. Bomme, S. Schultz-Cherry, and J. W. Rosch, "Direct interactions with Influenza promote bacterial adherence during respiratory infections," *Nat Microbiol*, vol. 4, pp. 1328–1336, Aug. 2019.
- [48] S. Sarkar, P. Khanna, and A. K. Singh, "Impact of COVID-19 in patients with concurrent co-infections : A systematic review and meta-analyses," *J Med Virol*, vol. 93, pp. 2385–2395, Apr. 2021.
- [49] Z. Guan, C. Chen, Y. Li, D. Yan, X. Zhang, D. Jiang, S. Yang, and L. Li, "Impact of coinfection with SARS-CoV-2 and Influenza on disease severity : A systematic review and meta-analysis," *Front. Public Health*, vol. 9, p. 773130, Dec. 2021.
- [50] J. Haney, S. Vijaykrishnan, J. Streetley, K. Dee, D. M. Goldfarb, M. Clarke, M. Mullin, S. D. Carter, D. Bhella, and P. R. Murcia, "In vitro coinfection by Influenza A virus and respiratory syncytial virus generates hybrid viral particles with altered structure and tropism," preprint, Microbiology, Aug. 2021.
- [51] J. S. Robertson, "5' and 3' terminal nucleotide sequences of the RNA genome segments of Influenza virus," *Nucl Acids Res*, vol. 6, no. 12, pp. 3745–3758, 1979.
- [52] U. Desselberger, V. R. Racaniello, J. J. Zazra, and P. Palese, "The 3' and 5'-terminal sequences of Influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity," *Gene*, vol. 8, pp. 315–328, Feb. 1980.
- [53] D. Ferhadian, M. Contrant, A. Printz-Schweigert, R. P. Smyth, J.-C. Paillart, and R. Marquet, "Structural and functional motifs in Influenza virus RNAs," *Front. Microbiol.*, vol. 9, p. 559, Mar. 2018.
- [54] A. Vasin, O. Temkina, V. Egorov, S. Klotchenko, M. Plotnikova, and O. Kiselev, "Molecular mechanisms enhancing the proteome of Influenza A viruses : An overview of recently discovered proteins," *Virus Research*, vol. 185, pp. 53–63, June 2014.
- [55] J. Benkaroun, G. Robertson, H. Whitney, and A. Lang, "Analysis of the variability in the non-coding regions of Influenza A viruses," *Veterinary Sciences*, vol. 5, p. 76, Aug. 2018.
- [56] L. L. M. Poon, D. C. Pritlove, E. Fodor, and G. G. Brownlee, "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*, vol. 73, pp. 3473–3476, Apr. 1999.
- [57] E. Fodor, D. C. Pritlove, and G. G. Brownlee, "The Influenza virus panhandle is involved in the initiation of transcription," *J Virol*, vol. 68, pp. 4092–4096, June 1994.
- [58] G. G. Brownlee and J. L. Sharps, "The RNA Polymerase of Influenza A virus is stabilized by interaction with its viral RNA promoter," *J Virol*, vol. 76, pp. 7103–7113, July 2002.
- [59] A. Pflug, D. Guilligay, S. Reich, and S. Cusack, "Structure of Influenza A polymerase bound to the viral RNA promoter," *Nature*, vol. 516, pp. 355–360, Dec. 2014.

- [60] A. R. Taylor, D. G. Sharp, D. Beard, J. W. Beard, J. H. Dingle, and A. E. Feller, "Isolation and characterization of Influenza A virus (PR8 Strain)," *The Journal of Immunology*, vol. 47, pp. 261–282, Sept. 1943.
- [61] V. M. Mosley and R. W. G. Wyckoff, "Electron micrography of the virus of Influenza," *Nature*, vol. 157, pp. 263–263, Mar. 1946.
- [62] C. M. Chu, I. M. Dawson, and W. J. Elford, "Filamentous forms associated with newly isolated Influenza virus," *The Lancet*, vol. 253, pp. 602–603, Apr. 1949.
- [63] A. P. Kendal, K. Apostolov, and G. Belyavin, "The effect of protease treatment on the morphology of Influenza A, B and C viruses," *Journal of General Virology*, vol. 5, pp. 141–143, July 1969.
- [64] R. W. Horne, A. P. Waterson, P. Wildy, and A. E. Farnham, "The structure and composition of the Myxoviruses : I. Electron microscope studies of the structure of Myxovirus particles by negative staining techniques," *Virology*, vol. 11, pp. 79–98, May 1960.
- [65] F. P. Booy, R. W. H. Ruigrok, and E. F. J. van Bruggen, "Electron microscopy of influenza virus : A comparison of negatively stained and ice-embedded particles," *Journal of Molecular Biology*, vol. 184, pp. 667–676, Aug. 1985.
- [66] B. E. Eddy and R. W. G. Wyckoff, "Influenza virus in sectioned tissues," *Proceedings of the Society for Experimental Biology and Medicine*, vol. 75, pp. 290–293, Oct. 1950.
- [67] C. J. Elleman and W. S. Barclay, "The M₁ matrix protein controls the filamentous phenotype of Influenza A virus," *Virology*, vol. 321, pp. 144–153, Mar. 2004.
- [68] P. C. Roberts and R. W. Compans, "Host cell dependence of viral morphology," *PNAS*, vol. 95, pp. 5746–5751, May 1998.
- [69] M. D. Vahey and D. A. Fletcher, "Influenza A virus surface proteins are organized to help penetrate host mucus," *eLife*, vol. 8, p. e43764, May 2019.
- [70] P. J. Campbell, S. Danzy, C. S. Kyriakis, M. J. Deymier, A. C. Lowen, and J. Steel, "The M segment of the 2009 pandemic Influenza virus confers increased neuraminidase activity, filamentous morphology, and efficient contact transmissibility to A/Puerto Rico/8/1934-based reassortant viruses," *J Virol*, vol. 88, pp. 3802–3814, Apr. 2014.
- [71] S. S. Lakdawala, E. W. Lamirande, A. L. Suguitan, W. Wang, C. P. Santos, L. Vogel, Y. Matsuoka, W. G. Lindsley, H. Jin, and K. Subbarao, "Eurasian-origin gene segments contribute to the transmissibility, aerosol release, and morphology of the 2009 pandemic H₁N₁ Influenza virus," *PLoS Pathog*, vol. 7, p. e1002443, Dec. 2011.
- [72] S. L. Zebedee and R. A. Lamb, "Growth restriction of Influenza A virus by M₂ protein antibody is genetically linked to the M₁ protein," *PNAS*, vol. 86, pp. 1061–1065, Feb. 1989.

- [73] P. C. Roberts, R. A. Lamb, and R. W. Compans, "The M₁ and M₂ proteins of Influenza A virus are important determinants in filamentous particle formation," *Virology*, vol. 240, pp. 127–137, Jan. 1998.
- [74] S. V. Bourmakina and A. Y. R. García-Sastre, "Reverse genetics studies on the filamentous morphology of influenza A virus," *Journal of General Virology*, vol. 84, no. 3, pp. 517–527, 2003.
- [75] L. M. Burleigh, L. J. Calder, J. J. Skehel, and D. A. Steinhauer, "Influenza A Viruses with mutations in the M₁ helix six domain display a wide variety of morphological phenotypes," *J Virol*, vol. 79, pp. 1262–1270, Jan. 2005.
- [76] J. Seladi-Schulman, J. Steel, and A. C. Lowen, "Spherical Influenza viruses have a fitness advantage in embryonated eggs, while filament-producing strains are selected *in vivo*," *J Virol*, vol. 87, pp. 13343–13353, Dec. 2013.
- [77] K. M. Bialas, K. A. Bussey, R. L. Stone, and T. Takimoto, "Specific nucleoprotein residues affect Influenza virus morphology," *J Virol*, vol. 88, pp. 2227–2234, Feb. 2014.
- [78] H. Kong, S. Ma, J. Wang, C. Gu, Z. Wang, J. Shi, G. Deng, Y. Guan, and H. Chen, "Identification of key amino acids in the PB₂ and M₁ proteins of H₇N₉ Influenza virus that affect its transmission in guinea pigs," *J Virol*, vol. 94, pp. e01180–19, Dec. 2019.
- [79] J. Guo, J. Chen, Y. Li, Y. Li, G. Deng, J. Shi, L. Liu, H. Chen, and X. Li, "SUMOylation of matrix protein M₁ and filamentous morphology collectively contribute to the replication and virulence of highly pathogenic H₅N₁ avian Influenza viruses in mammals," *J Virol*, vol. 96, pp. e01630–21, Feb. 2022.
- [80] T. Liu, J. Muller, and Z. Ye, "Association of Influenza virus matrix protein with ribonucleoproteins may control viral growth and morphology," *Virology*, vol. 304, pp. 89–96, Dec. 2002.
- [81] J. Seladi-Schulman, P. J. Campbell, S. Suppiah, J. Steel, and A. C. Lowen, "Filament-producing mutants of Influenza A/Puerto Rico/8/1934 (H₁N₁) virus have higher neuraminidase activities than the spherical wild-type," *PLoS ONE*, vol. 9, p. e112462, Nov. 2014.
- [82] A. Harris, G. Cardone, D. C. Winkler, J. B. Heymann, M. Brecher, J. M. White, and A. C. Steven, "Influenza virus pleiomorphy characterized by cryoelectron tomography," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 103, pp. 19123–19127, Dec. 2006.
- [83] R. A. Lamb, S. L. Zebedee, and C. D. Richardson, "Influenza virus M₂ protein is an integral membrane protein expressed on the infected-cell surface," *Cell*, vol. 40, pp. 627–633, Mar. 1985.
- [84] S. C. Inglis, A. R. Carroll, R. A. Lamb, and B. W. Mahy, "Polypeptides specified by the Influenza virus genome," *Virology*, vol. 74, pp. 489–503, Oct. 1976.
- [85] M. Getie-Kehtie, I. Sultana, M. Eichelberger, and M. Alterman, "Label-free mass spectrometry-based quantification of hemagglutinin and neuraminidase in Influenza virus preparations and vaccines : Label-free mass spectrometry-based quantification of HA," *Influenza and Other Respiratory Viruses*, vol. 7, pp. 521–530, July 2013.

- [86] L. J. Calder, S. Wasilewski, J. A. Berriman, and P. B. Rosenthal, "Structural organization of a filamentous Influenza A virus," *Proc Natl Acad Sci U S A*, vol. 107, pp. 10685–10690, June 2010.
- [87] S. Wasilewski, L. J. Calder, T. Grant, and P. B. Rosenthal, "Distribution of surface glycoproteins on Influenza A virus determined by electron cryotomography," *Vaccine*, vol. 30, pp. 7368–7373, Dec. 2012.
- [88] K. G. Murti, R. G. Webster, and I. M. Jones, "Localization of RNA polymerases on Influenza viral ribonucleoproteins by immunogold labeling," *Virology*, vol. 164, pp. 562–566, June 1988.
- [89] R. W. Compans, J. Content, and P. H. Duesberg, "Structure of the ribonucleoprotein of Influenza virus," *J Virol*, vol. 10, pp. 795–800, Oct. 1972.
- [90] M. H. Heggeness, P. Smith, I. Ulmanen, R. M. Krug, and P. W. Choppin, "Studies on the helical nucleocapsid of Influenza virus," *Virology*, vol. 118, pp. 466–470, Apr. 1982.
- [91] R. Arranz, R. Coloma, F. J. Chichón, J. J. Conesa, J. L. Carrascosa, J. M. Valpuesta, J. Ortín, and J. Martín-Benito, "The structure of native Influenza virion ribonucleoproteins," *Science*, vol. 338, pp. 1634–1637, Dec. 2012.
- [92] A. Moeller, R. N. Kirchdoerfer, C. S. Potter, B. Carragher, and I. A. Wilson, "Organization of the Influenza virus replication machinery," *Science*, vol. 338, pp. 1631–1634, Dec. 2012.
- [93] T. Noda, H. Sagara, A. Yen, A. Takada, H. Kida, R. H. Cheng, and Y. Kawaoka, "Architecture of ribonucleoprotein complexes in Influenza A virus particles," *Nature*, vol. 439, pp. 490–492, Jan. 2006.
- [94] E. Fournier, V. Moules, B. Essere, J.-C. Paillart, J.-D. Sirbat, C. Isel, A. Cavalier, J.-P. Rolland, D. Thomas, B. Lina, and R. Marquet, "A supramolecular assembly formed by Influenza A virus genomic RNA segments," *Nucleic Acids Research*, vol. 40, pp. 2197–2209, Mar. 2012.
- [95] F. Baudin, C. Bach, S. Cusack, and R. W. Ruigrok, "Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent," *EMBO J*, vol. 13, pp. 3158–3165, July 1994.
- [96] K. Klumpp, "Roles of the Influenza virus polymerase and nucleoprotein in forming a functional RNP structure," *The EMBO Journal*, vol. 16, pp. 1248–1257, Mar. 1997.
- [97] T. Noda and Y. Kawaoka, "Structure of Influenza virus ribonucleoprotein complexes and their packaging into virions," *Rev. Med. Virol.*, vol. 20, pp. 380–391, Nov. 2010.
- [98] D. Ferhadian, *Structure de l'ARN Au Sein Des Ribonucléoprotéines Des Influenzavirus A*. These de doctorat, Strasbourg, Sept. 2018.
- [99] S. Nakatsu, S. Murakami, K. Shindo, T. Horimoto, H. Sagara, T. Noda, and Y. Kawaoka, "Influenza C and D viruses package eight organized ribonucleoprotein complexes," *J Virol*, vol. 92, pp. e02084–17, Mar. 2018.

- [100] S. Nakatsu, H. Sagara, Y. Sakai-Tagawa, N. Sugaya, T. Noda, and Y. Kawaoka, "Complete and incomplete genome packaging of Influenza A and B viruses," *mBio*, vol. 7, pp. e01248–16, /mbio/7/5/e01248–16.atom, Nov. 2016.
- [101] Y.-y. Chou, R. Vafabakhsh, S. Doganay, Q. Gao, T. Ha, and P. Palese, "One Influenza virus particle packages eight unique viral RNAs as shown by FISH analysis," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 9101–9106, June 2012.
- [102] A. Inagaki, H. Goto, S. Kakugawa, M. Ozawa, and Y. Kawaoka, "Competitive incorporation of homologous gene segments of Influenza A virus into virions," *J. Virol.*, vol. 86, pp. 10200–10202, Sept. 2012.
- [103] T. Noda, Y. Sugita, K. Aoyama, A. Hirase, E. Kawakami, A. Miyazawa, H. Sagara, and Y. Kawaoka, "Three-dimensional analysis of ribonucleoprotein complexes in Influenza A virus," *Nat Commun*, vol. 3, p. 639, Jan. 2012.
- [104] C. Gavazzi, C. Isel, E. Fournier, V. Moules, A. Cavalier, D. Thomas, B. Lina, and R. Marquet, "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 Research*, vol. 41, pp. 1241–1254, Jan. 2013.
- [105] T. Noda, S. Murakami, S. Nakatsu, H. Imai, Y. Muramoto, K. Shindo, H. Sagara, and Y. Kawaoka, "Importance of the 1+7 configuration of ribonucleoprotein complexes for Influenza A virus genome packaging," *Nat Commun*, vol. 9, p. 54, Dec. 2018.
- [106] M. Zanin, B. Marathe, S.-S. Wong, S.-W. Yoon, E. Collin, C. Oshansky, J. Jones, B. Hause, and R. Webby, "Pandemic swine H1N1 Influenza viruses with almost undetectable neuraminidase activity are not transmitted via aerosols in ferrets and are inhibited by human mucus but not swine mucus," *Journal of Virology*, Mar. 2015.
- [107] R. Xu, X. Zhu, R. McBride, C. M. Nycholat, W. Yu, J. C. Paulson, and I. A. Wilson, "Functional balance of the hemagglutinin and neuraminidase activities accompanies the emergence of the 2009 H1N1 Influenza pandemic," *Journal of Virology*, Sept. 2012.
- [108] T. Sakai, H. Takagi, Y. Muraki, and M. Saito, "Unique directional motility of Influenza C virus controlled by Its filamentous morphology and short-range motions," *Journal of Virology*, Nov. 2017.
- [109] L. Byrd-Leotis, N. Jia, S. Dutta, J. F. Trost, C. Gao, S. F. Cummings, T. Bräulke, S. Müller-Loennies, J. Heimburg-Molinaro, D. A. Steinhauer, and R. D. Cummings, "Influenza binds phosphorylated glycans from human lung," *Sci Adv*, vol. 5, Feb. 2019.
- [110] U. Karakus, T. Thamamongood, K. Ciminski, W. Ran, S. C. Günther, M. O. Pohl, D. Eletto, C. Jeney, D. Hoffmann, S. Reiche, J. Schinköthe, R. Ulrich, J. Wiener, M. G. B. Hayes, M. W. Chang, A. Hunziker, E. Yángüez, T. Aydililo, F. Krammer, J. Oderbolz, M. Meier, A. Oxenius, A. Halenius, G. Zimmer, C. Benner, B. G. Hale, A. García-Sastre, M. Beer, M. Schwemmle, and S. Stertz, "MHC class II proteins mediate cross-species entry of bat Influenza viruses," *Nature*, vol. 567, pp. 109–112, Mar. 2019.

- [111] H. Guo, H. Rabouw, A. Slomp, M. Dai, F. van der Vegt, J. W. M. van Lent, R. McBride, J. C. Paulson, R. J. de Groot, F. J. M. van Kuppeveld, E. de Vries, and C. A. M. de Haan, “Kinetic analysis of the influenza A virus HA/NA balance reveals contribution of NA to virus-receptor binding and NA-dependent rolling on receptor-containing surfaces,” *PLoS Pathog*, vol. 14, p. e1007233, Aug. 2018.
- [112] U. Karakus, M. O. Pohl, and S. Stertz, “Breaking the convention : Sialoglycan variants, coreceptors, and alternative receptors for Influenza A virus entry,” *J Virol*, vol. 94, pp. e01357–19, Jan. 2020.
- [113] M. Sempere Borau and S. Stertz, “Entry of Influenza A virus into host cells - recent progress and remaining challenges,” *Current Opinion in Virology*, vol. 48, pp. 23–29, June 2021.
- [114] S. Patterson, J. S. Oxford, and R. R. Dourmashkin, “Studies on the mechanism of Influenza virus entry into cells,” *Journal of General Virology*, vol. 43, pp. 223–229, Apr. 1979.
- [115] K. S. Matlin, H. Reggio, A. Helenius, and K. Simons, “Infectious entry pathway of Influenza virus in a canine kidney cell line,” *Journal of Cell Biology*, vol. 91, pp. 601–613, Dec. 1981.
- [116] C. Chen and X. Zhuang, “Epsin 1 is a cargo-specific adaptor for the clathrin-mediated endocytosis of the Influenza virus,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 105, pp. 11790–11795, Aug. 2008.
- [117] S. B. Sieczkarski and G. R. Whittaker, “Influenza virus can enter and infect cells in the absence of clathrin-mediated endocytosis,” *J Virol*, vol. 76, pp. 10455–10464, Oct. 2002.
- [118] M. J. Rust, M. Lakadamyali, F. Zhang, and X. Zhuang, “Assembly of endocytic machinery around individual Influenza viruses during viral entry,” *Nat Struct Mol Biol*, vol. 11, pp. 567–573, June 2004.
- [119] E. de Vries, D. M. Tscherne, M. J. Wienholts, V. Cobos-Jiménez, F. Scholte, A. García-Sastre, P. J. M. Rottier, and C. A. M. de Haan, “Dissection of the Influenza A virus endocytic routes reveals macropinocytosis as an alternative entry pathway,” *PLoS Pathog*, vol. 7, p. e1001329, Mar. 2011.
- [120] A. G. Bukrinskaya, N. K. Vorkunova, G. V. Kornilayeva, R. A. Narmanbetova, and G. K. Vorkunova, “Influenza virus uncoating in infected cells and effect of rimantadine,” *J Gen Virol*, vol. 60, pp. 49–59, May 1982.
- [121] O. Zhirnov, “Solubilization of matrix protein M₁/M from virions occurs at different pH for Orthomyxo- and Paramyxoviruses,” *Virology*, vol. 176, pp. 274–279, May 1990.
- [122] S. Stauffer, Y. Feng, F. Nebioglu, R. Heilig, P. Picotti, and A. Helenius, “Stepwise priming by acidic pH and a high K⁺ concentration is required for efficient uncoating of Influenza A virus cores after penetration,” *J Virol*, vol. 88, pp. 13029–13046, Nov. 2014.
- [123] F. Boulay, R. W. Doms, I. Wilson, and A. Helenius, “The influenza hemagglutinin precursor as an acid-sensitive probe of the biosynthetic pathway,” *The EMBO Journal*, vol. 6, pp. 2643–2650, Sept. 1987.
- [124] E. Böttcher-Friebertshäuser, W. Garten, M. Matrosovich, and H. D. Klenk, “The hemagglutinin : A determinant of pathogenicity,” in *Influenza Pathogenesis and Control - Volume I* (R. W. Compans and

- M. B. A. Oldstone, eds.), *Current Topics in Microbiology and Immunology*, pp. 3–34, Cham : Springer International Publishing, 2014.
- [125] K. S. Lim, M. S. Mohamed, H. Wang, Hartono, M. Hazawa, A. Kobayashi, D. C.-C. Voon, N. Kodera, T. Ando, and R. W. Wong, “Direct visualization of avian Influenza H5N1 hemagglutinin precursor and its conformational change by high-speed atomic force microscopy,” *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1864, p. 129313, Feb. 2020.
- [126] D. K. Das, R. Govindan, I. Nikić-Spiegel, F. Krammer, E. A. Lemke, and J. B. Munro, “Direct visualization of the conformational dynamics of single Influenza hemagglutinin trimers,” *Cell*, vol. 174, pp. 926–937.e12, Aug. 2018.
- [127] I. Banerjee, Y. Miyake, S. P. Nobs, C. Schneider, P. Horvath, M. Kopf, P. Matthias, A. Helenius, and Y. Yamauchi, “Influenza A virus uses the aggresome processing machinery for host cell entry,” *Science*, vol. 346, pp. 473–477, Oct. 2014.
- [128] Y. Miyake, J. J. Keusch, L. Decamps, H. Ho-Xuan, S. Iketani, H. Gut, U. Kutay, A. Helenius, and Y. Yamauchi, “Influenza virus uses transportin 1 for vRNP debundling during cell entry,” *Nat Microbiol*, vol. 4, pp. 578–586, Apr. 2019.
- [129] K. Martin and A. Helenius, “Transport of incoming Influenza virus nucleocapsids into the nucleus,” *J Virol*, vol. 65, pp. 232–244, Jan. 1991.
- [130] I. Kemler, G. Whittaker, and A. Helenius, “Nuclear import of microinjected Influenza virus ribonucleoproteins,” *Virology*, vol. 202, pp. 1028–1033, Aug. 1994.
- [131] R. E. O’Neill, R. Jaskunas, G. Blobel, P. Palese, and J. Moroianu, “Nuclear import of Influenza virus RNA can be mediated by viral nucleoprotein and transport factors required for protein import,” *J Biol Chem*, vol. 270, pp. 22701–22704, Sept. 1995.
- [132] J. F. Cros and P. Palese, “Trafficking of viral genomic RNA into and out of the nucleus : Influenza, Thogoto and Borna disease viruses,” *Virus Research*, vol. 95, pp. 3–12, Sept. 2003.
- [133] I. Jones, P. Reay, and K. Philpott, “Nuclear location of all three Influenza polymerase proteins and a nuclear signal in polymerase PB2,” *The EMBO Journal*, vol. 5, pp. 2371–2376, Sept. 1986.
- [134] S. T. Nath and D. P. Nayak, “Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 Influenza virus (H1 N1),” *Mol. Cell. Biol.*, vol. 10, pp. 4139–4145, Aug. 1990.
- [135] Y. Asano and A. Ishihama, “Identification of two nucleotide-binding domains on the PB1 subunit of Influenza virus RNA Polymerase,” *Journal of Biochemistry*, vol. 122, pp. 627–634, Sept. 1997.
- [136] A. Nieto, S. de la Luna, J. Barcena, A. Portela, and J. Ortin, “Complex structure of the nuclear translocation signal of Influenza virus polymerase PA subunit,” *Journal of General Virology*, vol. 75, pp. 29–36, Jan. 1994.

- [137] F. Tarendeau, J. Boudet, D. Guilligay, P. J. Mas, C. M. Bougault, S. Boulo, F. Baudin, R. W. H. Ruigrok, N. Daigle, J. Ellenberg, S. Cusack, J.-P. Simorre, and D. J. Hart, "Structure and nuclear import function of the C-terminal domain of Influenza virus polymerase PB2 subunit," *Nat Struct Mol Biol*, vol. 14, pp. 229–233, Mar. 2007.
- [138] J. Davey, N. Dimmock, and A. Colman, "Identification of the sequence responsible for the nuclear accumulation of the Influenza virus nucleoprotein in *Xenopus* oocytes," *Cell*, vol. 40, pp. 667–675, Mar. 1985.
- [139] G. Gabriel, K. Klingel, A. Otte, S. Thiele, B. Hudjetz, G. Arman-Kalcek, M. Sauter, T. Shmidt, F. Rother, S. Baumgarte, B. Keiner, E. Hartmann, M. Bader, G. G. Brownlee, E. Fodor, and H.-D. Klenk, "Differential use of importin- α isoforms governs cell tropism and host adaptation of influenza virus," *Nat Commun*, vol. 2, p. 156, Sept. 2011.
- [140] K. Ninpan, O. Suptawiwat, C. Boonarkart, P. Phuangphung, S. Sathirareuangchai, M. Uiprasertkul, and P. Auewarakul, "Expression of importin- α isoforms in human nasal mucosa : Implication for adaptation of avian Influenza A viruses to human host," *Virol J*, vol. 13, p. 90, Dec. 2016.
- [141] A. Donchet, E. Vassal-Stermann, F. C. A. Gérard, R. W. H. Ruigrok, and T. Crépin, "Differential behaviours and preferential bindings of Influenza nucleoproteins on Importins- α ," *Viruses*, vol. 12, p. 834, July 2020.
- [142] M. Bouloy, S. J. Plotch, and R. M. Krug, "Globin mRNAs are primers for the transcription of influenza viral RNA *in vitro*," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 75, pp. 4886–4890, Oct. 1978.
- [143] S. J. Plotch, M. Bouloy, and R. M. Krug, "Transfer of 5'-terminal cap of globin mRNA to Influenza viral complementary RNA during transcription *in vitro*," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 76, pp. 1618–1622, Apr. 1979.
- [144] S. J. Plotch, M. Bouloy, I. Ulmanen, and R. M. Krug, "A unique cap(m7GpppXm)-dependent Influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription," *Cell*, vol. 23, pp. 847–858, Mar. 1981.
- [145] A. Dias, D. Bouvier, T. Crépin, A. A. McCarthy, D. J. Hart, F. Baudin, S. Cusack, and R. W. H. Ruigrok, "The cap-snatching endonuclease of Influenza virus polymerase resides in the PA subunit," *Nature*, vol. 458, pp. 914–918, Apr. 2009.
- [146] P. Yuan, M. Bartlam, Z. Lou, S. Chen, J. Zhou, X. He, Z. Lv, R. Ge, X. Li, T. Deng, E. Fodor, Z. Rao, and Y. Liu, "Crystal structure of an avian Influenza polymerase PAN reveals an endonuclease active site," *Nature*, vol. 458, pp. 909–913, Apr. 2009.
- [147] S. K. Biswas and D. P. Nayak, "Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1," *J Virol*, vol. 68, pp. 1819–1826, Mar. 1994.

- [148] S. Reich, D. Guilligay, A. Pflug, H. Malet, I. Berger, T. Crépin, D. Hart, T. Lunardi, M. Nanao, R. W. H. Ruigrok, and S. Cusack, “Structural insight into cap-snatching and RNA synthesis by Influenza polymerase,” *Nature*, vol. 516, pp. 361–366, Dec. 2014.
- [149] A. Pflug, S. Gaudon, P. Resa-Infante, M. Lethier, S. Reich, W. M. Schulze, and S. Cusack, “Capped RNA primer binding to Influenza polymerase and implications for the mechanism of cap-binding inhibitors,” *Nucleic Acids Research*, vol. 46, pp. 956–971, Jan. 2018.
- [150] T. Kouba, P. Drncová, and S. Cusack, “Structural snapshots of actively transcribing Influenza polymerase,” *Nat Struct Mol Biol*, vol. 26, pp. 460–470, June 2019.
- [151] O. G. Engelhardt, M. Smith, and E. Fodor, “Association of the Influenza A virus RNA-dependent RNA Polymerase with cellular RNA Polymerase II,” *J Virol*, vol. 79, pp. 5812–5818, May 2005.
- [152] M. Martínez-Alonso, N. Hengrung, and E. Fodor, “RNA-free and ribonucleoprotein-associated Influenza virus Polymerases directly bind the serine-5-phosphorylated carboxyl-terminal domain of host RNA Polymerase II,” *J Virol*, vol. 90, pp. 6014–6021, July 2016.
- [153] J. M. Wandzik, T. Kouba, M. Karuppasamy, A. Pflug, P. Drncova, J. Provaznik, N. Azevedo, and S. Cusack, “A structure-based model for the complete transcription cycle of Influenza Polymerase,” *Cell*, vol. 181, pp. 877–893.e21, May 2020.
- [154] J. S. Robertson, M. Schubert, and R. A. Lazzarini, “Polyadenylation sites for Influenza virus mRNA,” *J Virol*, vol. 38, pp. 157–163, Apr. 1981.
- [155] G. E. Mark, J. M. Taylor, B. Broni, and R. M. Krug, “Nuclear accumulation of Influenza viral RNA transcripts and the effects of Cycloheximide, Actinomycin D, and α -Amanitin,” *J Virol*, vol. 29, pp. 744–752, Feb. 1979.
- [156] N. Jorba, R. Coloma, and J. Ortín, “Genetic trans-complementation establishes a new model for Influenza virus RNA transcription and replication,” *PLoS Pathog*, vol. 5, p. e1000462, May 2009.
- [157] M.-J. Amorim, E. K. Read, R. M. Dalton, L. Medcalf, and P. Digard, “Nuclear export of Influenza A virus mRNAs requires ongoing RNA Polymerase II activity,” *Traffic*, vol. 8, pp. 1–11, Jan. 2007.
- [158] J. P. Gales, J. Kubina, A. Geldreich, and M. Dimitrova, “Strength in diversity : Nuclear export of viral RNAs,” *Viruses*, vol. 12, p. 1014, Sept. 2020.
- [159] S. Larsen, S. Bui, V. Perez, A. Mohammad, H. Medina-Ramirez, and L. L. Newcomb, “Influenza polymerase encoding mRNAs utilize atypical mRNA nuclear export,” *Virology*, vol. 11, p. 154, Dec. 2014.
- [160] Z. Chen and R. M. Krug, “Selective nuclear export of viral mRNAs in Influenza-virus-infected cells,” *Trends in Microbiology*, vol. 8, pp. 376–383, Aug. 2000.
- [161] Z.-x. Ji, X.-q. Wang, and X.-f. Liu, “NS1 : A key protein in the “Game” between Influenza A virus and host in innate immunity,” *Front. Cell. Infect. Microbiol.*, vol. 11, p. 670177, July 2021.

- [162] F. Wang, C. M. Sheppard, B. Mistry, E. Staller, W. S. Barclay, J. M. Grimes, E. Fodor, and H. Fan, "The C-terminal LCAR of host ANP32 proteins interacts with the Influenza A virus nucleoprotein to promote the replication of the viral RNA genome," *Nucleic Acids Research*, vol. 50, pp. 5713–5725, June 2022.
- [163] F. T. Vreede and G. G. Brownlee, "Influenza virion-derived viral ribonucleoproteins synthesize both mRNA and cRNA in vitro," *J Virol*, vol. 81, pp. 2196–2204, Mar. 2007.
- [164] N. C. Robb, M. Smith, F. T. Vreede, and E. Fodor, "NS2/NEP protein regulates transcription and replication of the Influenza virus RNA genome," *Journal of General Virology*, vol. 90, pp. 1398–1407, June 2009.
- [165] J. Li, M. Yu, W. Zheng, and W. Liu, "Nucleocytoplasmic shuttling of Influenza A virus proteins," *Viruses*, vol. 7, pp. 2668–2682, May 2015.
- [166] E. Fodor and M. Smith, "The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the Influenza A virus RNA Polymerase complex," *J Virol*, vol. 78, pp. 9144–9153, Sept. 2004.
- [167] T. Deng, O. G. Engelhardt, B. Thomas, A. V. Akoulitchev, G. G. Brownlee, and E. Fodor, "Role of Ran Binding Protein 5 in nuclear import and assembly of the Influenza virus RNA Polymerase complex," *J Virol*, vol. 80, pp. 11911–11919, Dec. 2006.
- [168] S. Huet, S. V. Avilov, L. Ferbitz, N. Daigle, S. Cusack, and J. Ellenberg, "Nuclear import and assembly of Influenza A virus RNA Polymerase studied in live cells by fluorescence cross-correlation spectroscopy," *J Virol*, vol. 84, pp. 1254–1264, Feb. 2010.
- [169] J. F. Cros, A. García-Sastre, and P. Palese, "An unconventional NLS is critical for the nuclear import of the Influenza A virus nucleoprotein and ribonucleoprotein : Role of the unconventional NLS of the Influenza A virus NP," *Traffic*, vol. 6, pp. 205–213, Mar. 2005.
- [170] S. Gao, S. Wang, S. Cao, L. Sun, J. Li, Y. Bi, G. F. Gao, and W. Liu, "Characteristics of nucleocytoplasmic transport of H1N1 Influenza A virus nuclear export protein," *J Virol*, vol. 88, pp. 7455–7463, July 2014.
- [171] Z. Ye, T. Liu, D. P. Offringa, J. McInnis, and R. A. Levandowski, "Association of Influenza virus matrix protein with ribonucleoproteins," *J Virol*, vol. 73, pp. 7467–7473, Sept. 1999.
- [172] H. Akarsu, "Crystal structure of the M1 protein-binding domain of the influenza A virus nuclear export protein (NEP/NS2)," *The EMBO Journal*, vol. 22, pp. 4646–4655, Sept. 2003.
- [173] T. Shimizu, N. Takizawa, K. Watanabe, K. Nagata, and N. Kobayashi, "Crucial role of the Influenza virus NS2 (NEP) C-terminal domain in M1 binding and nuclear export of vRNP," *FEBS Letters*, vol. 585, pp. 41–46, Jan. 2011.
- [174] F. Momose, Y. Kikuchi, K. Komase, and Y. Morikawa, "Visualization of microtubule-mediated transport of Influenza viral progeny ribonucleoprotein," *Microbes and Infection*, vol. 9, pp. 1422–1433, Oct. 2007.

- [175] C. Chaimayo, T. Hayashi, A. Underwood, E. Hodges, and T. Takimoto, "Selective incorporation of vRNP into Influenza A virions determined by its specific interaction with M₁ protein," *Virology*, vol. 505, pp. 23–32, May 2017.
- [176] E. A. Bruce, P. Digard, and A. D. Stuart, "The Rab11 pathway is required for Influenza A virus budding and filament formation," *J Virol*, vol. 84, pp. 5848–5859, June 2010.
- [177] A. J. Eisfeld, E. Kawakami, T. Watanabe, G. Neumann, and Y. Kawaoka, "RAB11A is essential for transport of the Influenza virus genome to the plasma membrane," *Journal of Virology*, July 2011.
- [178] M. J. Amorim, E. A. Bruce, E. K. C. Read, Á. Foeglein, R. Mahen, A. D. Stuart, and P. Digard, "A Rab11- and microtubule-dependent mechanism for cytoplasmic transport of Influenza A virus viral RNA," *J Virol*, vol. 85, pp. 4143–4156, May 2011.
- [179] J. Han, K. Ganti, V. K. Sali, C. Twigg, Y. Zhang, S. Manivasagam, C.-Y. Liang, O. A. Vogel, I. Huang, S. N. Emmanuel, J. Plung, L. Radoshevich, J. T. Perez, A. C. Lowen, and B. Manicassamy, "Host factor Rab11a is critical for efficient assembly of Influenza A virus genomic segments," *PLoS Pathog*, vol. 17, p. e1009517, May 2021.
- [180] K. Ganti, J. Han, B. Manicassamy, and A. C. Lowen, "Rab11a mediates cell-cell spread and reassortment of Influenza A virus genomes via tunneling nanotubes," *PLoS Pathog*, vol. 17, p. e1009321, Sept. 2021.
- [181] A. Kumar, J. H. Kim, P. Ranjan, M. G. Metcalfe, W. Cao, M. Mishina, S. Gangappa, Z. Guo, E. S. Boyden, S. Zaki, I. York, A. García-Sastre, M. Shaw, and S. Sambhara, "Influenza virus exploits tunneling nanotubes for cell-to-cell spread," *Sci Rep*, vol. 7, p. 40360, Feb. 2017.
- [182] K. L. Roberts, B. Manicassamy, and R. A. Lamb, "Influenza A virus uses intercellular connections to spread to neighboring cells," *J Virol*, vol. 89, pp. 1537–1549, Feb. 2015.
- [183] I. F. de Castro Martin, G. Fournier, M. Sachse, J. Pizarro-Cerda, C. Risco, and N. Naffakh, "Influenza virus genome reaches the plasma membrane via a modified endoplasmic reticulum and Rab11-dependent vesicles," *Nat Commun*, vol. 8, p. 1396, Dec. 2017.
- [184] E. Nturibi, A. R. Bhagwat, S. Coburn, M. M. Myerburg, and S. S. Lakdawala, "Intracellular colocalization of Influenza viral RNA and Rab11A is dependent upon microtubule filaments," *J Virol*, vol. 91, pp. e01179–17, Oct. 2017.
- [185] S. Giese, H. Bolte, and M. Schwemmle, "The feat of packaging eight unique genome segments," *Viruses*, vol. 8, p. 165, June 2016.
- [186] J. A. Schalken, A. J. Roebroek, P. P. Oomen, S. S. Wagenaar, F. M. Debruyne, H. P. Bloemers, and W. J. Van de Ven, "Fur gene expression as a discriminating marker for small cell and nonsmall cell lung carcinomas," *J. Clin. Invest.*, vol. 80, pp. 1545–1549, Dec. 1987.
- [187] A. Stieneke-Gröber, M. Vey, H. Angliker, E. Shaw, G. Thomas, C. Roberts, H. Klenk, and W. Garten, "Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease," *The EMBO Journal*, vol. 11, pp. 2407–2414, July 1992.

- [188] D. Lingwood and K. Simons, "Lipid rafts as a membrane-organizing principle," *Science*, vol. 327, pp. 46–50, Jan. 2010.
- [189] J. Zhang and R. A. Lamb, "Characterization of the membrane association of the Influenza virus matrix protein in living cells," *Virology*, vol. 225, pp. 255–266, Nov. 1996.
- [190] A. Ali, R. T. Avalos, E. Ponimaskin, and D. P. Nayak, "Influenza virus assembly: Effect of Influenza virus glycoproteins on the membrane association of M₁ protein," *J Virol*, vol. 74, pp. 8709–8719, Sept. 2000.
- [191] M. L. Shaw, K. L. Stone, C. M. Colangelo, E. E. Gulcicek, and P. Palese, "Cellular proteins in Influenza virus particles," *PLoS Pathog*, vol. 4, p. e1000085, June 2008.
- [192] M. F. McCown and A. Pekosz, "The Influenza A virus M₂ cytoplasmic tail is required for infectious virus production and efficient genome packaging," *J Virol*, vol. 79, pp. 3595–3605, Mar. 2005.
- [193] K. Iwatsuki-Horimoto, T. Horimoto, T. Noda, M. Kiso, J. Maeda, S. Watanabe, Y. Muramoto, K. Fujii, and Y. Kawaoka, "The cytoplasmic tail of the Influenza A virus M₂ protein plays a role in viral assembly," *J Virol*, vol. 80, pp. 5233–5240, June 2006.
- [194] B. J. Chen, G. P. Leser, D. Jackson, and R. A. Lamb, "The Influenza virus M₂ protein cytoplasmic tail interacts with the M₁ protein and influences virus assembly at the site of virus budding," *J Virol*, vol. 82, pp. 10059–10070, Oct. 2008.
- [195] P. Gómez-Puertas, C. Albo, E. Pérez-Pastrana, A. Vivo, and A. Portela, "Influenza Virus matrix protein is the major driving force in virus budding," *J Virol*, vol. 74, pp. 11538–11547, Dec. 2000.
- [196] D. Wang, A. Harmon, J. Jin, D. H. Francis, J. Christopher-Hennings, E. Nelson, R. C. Montelaro, and F. Li, "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*, vol. 84, pp. 4673–4681, May 2010.
- [197] B. J. Chen, G. P. Leser, E. Morita, and R. A. Lamb, "Influenza virus hemagglutinin and neuraminidase, but not the matrix protein, are required for assembly and budding of plasmid-derived virus-like particles," *J Virol*, vol. 81, pp. 7111–7123, July 2007.
- [198] J. C. C. Lai, W. W. L. Chan, F. Kien, J. M. Nicholls, J. S. M. Peiris, and J.-M. Garcia, "Formation of virus-like particles from human cell lines exclusively expressing Influenza neuraminidase," *Journal of General Virology*, vol. 91, pp. 2322–2330, Sept. 2010.
- [199] M. A. Yondola, F. Fernandes, A. Belicha-Villanueva, M. Uccellini, Q. Gao, C. Carter, and P. Palese, "Budding capability of the Influenza virus neuraminidase can be modulated by tetherin," *J Virol*, vol. 85, pp. 2480–2491, Mar. 2011.
- [200] J. Lohmeyer, L. T. Talens, and H.-D. Klenk, "Biosynthesis of the Influenza virus envelope in abortive infection," *Journal of General Virology*, vol. 42, pp. 73–88, Jan. 1979.

- [201] A. K. Pattnaik, D. J. Brown, and D. P. Nayak, "Formation of Influenza virus particles lacking hemagglutinin on the viral envelope," *J Virol*, vol. 60, pp. 994–1001, Dec. 1986.
- [202] C. Liu, M. C. Eichelberger, R. W. Compans, and G. M. Air, "Influenza type A virus neuraminidase does not play a role in viral entry, replication, assembly, or budding," *J Virol*, vol. 69, pp. 1099–1106, Feb. 1995.
- [203] J. S. Rossman, X. Jing, G. P. Leser, and R. A. Lamb, "Influenza virus M2 protein mediates ESCRT-independent membrane scission," *Cell*, vol. 142, pp. 902–913, Sept. 2010.
- [204] E. C. Hutchinson, J. C. von Kirchbach, J. R. Gog, and P. Digard, "Genome packaging in Influenza A virus," *Journal of General Virology*, vol. 91, pp. 313–328, Feb. 2010.
- [205] N. T. Jacobs, N. O. Onuoha, A. Antia, J. Steel, R. Antia, and A. C. Lowen, "Incomplete Influenza A virus genomes occur frequently but are readily complemented during localized viral spread," *Nat Commun*, vol. 10, p. 3526, Dec. 2019.
- [206] R. Tellier, "Aerosol transmission of Influenza A virus : A review of new studies," *J. R. Soc. Interface.*, vol. 6, Dec. 2009.
- [207] C. B. Brooke, W. L. Ince, J. Wrammert, R. Ahmed, P. C. Wilson, J. R. Bennink, and J. W. Yewdell, "Most Influenza A virions fail to express at least one essential viral protein," *J Virol*, vol. 87, pp. 3155–3162, Mar. 2013.
- [208] F. G. Alnaji, W. K. Reiser, J. Rivera-Cardona, A. J. W. te Velhuis, and C. B. Brooke, "Influenza A virus defective viral genomes are inefficiently packaged into virions relative to wild-type genomic RNAs," *mBio*, vol. 12, pp. e02959–21, Dec. 2021.
- [209] A. R. Davis and D. P. Nayak, "Sequence relationships among defective interfering influenza viral RNAs.," *Proceedings of the National Academy of Sciences*, vol. 76, pp. 3092–3096, July 1979.
- [210] P. A. Jennings, J. T. Finch, G. Winter, and J. S. Robertson, "Does the higher order structure of the Influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA?," *Cell*, vol. 34, pp. 619–627, Sept. 1983.
- [211] M. Enami, G. Sharma, C. Benham, and P. Palese, "An Influenza virus containing nine different RNA segments," *Virology*, vol. 185, pp. 291–298, Nov. 1991.
- [212] M. J. Sekellick, S. A. Carra, A. Bowman, D. A. Hopkins, and P. I. Marcus, "Transient resistance of Influenza virus to interferon action attributed to random multiple packaging and activity of NS genes," *Journal of Interferon & Cytokine Research*, vol. 20, pp. 963–970, Nov. 2000.
- [213] C. T. Bancroft and T. G. Parslow, "Evidence for segment-nonspecific packaging of the Influenza A virus genome," *J Virol*, vol. 76, pp. 7133–7139, July 2002.
- [214] S. D. Duhaut and J. W. Mccauley, "Defective RNAs inhibit the assembly of Influenza virus genome segments in a segment-specific manner," *Virology*, vol. 216, pp. 326–337, Feb. 1996.

- [215] G. L. Smith and A. J. Hay, "Replication of the Influenza virus genome," *Virology*, vol. 118, pp. 96–108, Apr. 1982.
- [216] M. Bergmann and T. Muster, "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," *Journal of General Virology*, vol. 76, pp. 3211–3215, Dec. 1995.
- [217] D. McGeoch, P. Fellner, and C. Newton, "Influenza virus genome consists of eight distinct RNA species," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 73, pp. 3045–3049, Sept. 1976.
- [218] E. Hatada, M. Hasegawa, J. Mukaigawa, K. Shimizu, and R. Fukuda, "Control of Influenza virus gene expression : Quantitative analysis of each viral RNA species in infected cells," *The Journal of Biochemistry*, vol. 105, pp. 537–546, Apr. 1989.
- [219] W. Laver and J. C. Downie, "Influenza virus recombination," *Virology*, vol. 70, pp. 105–117, Mar. 1976.
- [220] K. Nakajima and A. Sugiura, "Three-factor cross of Influenza virus," *Virology*, vol. 81, pp. 486–489, Sept. 1977.
- [221] E. Dos Santos Afonso, N. Escriou, I. Leclercq, S. van der Werf, and N. Naffakh, "The generation of recombinant Influenza A viruses expressing a PB2 fusion protein requires the conservation of a packaging signal overlapping the coding and noncoding regions at the 5' end of the PB2 segment," *Virology*, vol. 341, pp. 34–46, Oct. 2005.
- [222] Y. Liang, Y. Hong, and T. G. Parslow, "Cis-acting packaging signals in the Influenza virus PB1, PB2, and PA genomic RNA segments," *J Virol*, vol. 79, pp. 10348–10355, Aug. 2005.
- [223] Y. Muramoto, A. Takada, K. Fujii, T. Noda, K. Iwatsuki-Horimoto, S. Watanabe, T. Horimoto, H. Kida, and Y. Kawaoka, "Hierarchy among viral RNA (vRNA) segments in their role in vRNA incorporation into Influenza A virions," *J Virol*, vol. 80, pp. 2318–2325, Mar. 2006.
- [224] J. R. Gog, E. D. S. Afonso, R. M. Dalton, I. Leclercq, L. Tiley, D. Elton, J. C. von Kirchbach, N. Naffakh, N. Escriou, and P. Digard, "Codon conservation in the Influenza A virus genome defines RNA packaging signals," *Nucleic Acids Research*, vol. 35, pp. 1897–1907, Mar. 2007.
- [225] G. A. Marsh, R. Rabadán, A. J. Levine, and P. Palese, "Highly conserved regions of Influenza A virus Polymerase gene segments are critical for efficient viral RNA packaging," *J Virol*, vol. 82, pp. 2295–2304, Mar. 2008.
- [226] T. Odagiri and M. Tashiro, "Segment-specific noncoding sequences of the Influenza virus genome RNA are involved in the specific competition between defective interfering RNA and its progenitor RNA segment at the virion assembly step," *Journal of Virology*, vol. 71, pp. 2138–2145, Mar. 1997.
- [227] T. Watanabe, S. Watanabe, T. Noda, Y. Fujii, and Y. Kawaoka, "Exploitation of nucleic acid packaging signals to generate a novel Influenza virus-based vector stably expressing two foreign genes," *J Virol*, vol. 77, pp. 10575–10583, Oct. 2003.

- [228] M. Ozawa, K. Fujii, Y. Muramoto, S. Yamada, S. Yamayoshi, A. Takada, H. Goto, T. Horimoto, and Y. Kawaoka, "Contributions of two nuclear localization signals of Influenza A virus nucleoprotein to viral replication," *Journal of Virology*, Jan. 2007.
- [229] E. C. Hutchinson, H. M. Wise, K. Kudryavtseva, M. D. Curran, and P. Digard, "Characterisation of Influenza A viruses with mutations in segment 5 packaging signals," *Vaccine*, vol. 27, pp. 6270–6275, Oct. 2009.
- [230] M. Bergmann and T. Muster, "Mutations in the nonconserved noncoding sequences of the influenza A virus segments affect viral vRNA formation," *Virus Research*, vol. 44, pp. 23–31, Sept. 1996.
- [231] Y. Fujii, H. Goto, T. Watanabe, T. Yoshida, and Y. Kawaoka, "Selective incorporation of Influenza virus RNA segments into virions," *Proceedings of the National Academy of Sciences*, vol. 100, pp. 2002–2007, Feb. 2003.
- [232] G. A. Marsh, R. Hatami, and P. Palese, "Specific residues of the Influenza A virus hemagglutinin viral RNA are important for efficient packaging into budding virions," *J Virol*, vol. 81, pp. 9727–9736, Sept. 2007.
- [233] E. Seshimo, F. Momose, and Y. Morikawa, "Identification of the 5'-terminal packaging signal of the H1N1 Influenza A virus neuraminidase segment at single-nucleotide resolution," *Front. Microbiol.*, vol. 12, p. 709010, Aug. 2021.
- [234] E. C. Hutchinson, M. D. Curran, E. K. Read, J. R. Gog, and P. Digard, "Mutational analysis of *cis*-acting RNA signals in segment 7 of Influenza A virus," *J Virol*, vol. 82, pp. 11869–11879, Dec. 2008.
- [235] M. Ozawa, J. Maeda, K. Iwatsuki-Horimoto, S. Watanabe, H. Goto, T. Horimoto, and Y. Kawaoka, "Nucleotide sequence requirements at the 5' end of the Influenza A virus M RNA segment for efficient virus replication," *J Virol*, vol. 83, pp. 3384–3388, Apr. 2009.
- [236] K. Fujii, Y. Fujii, T. Noda, Y. Muramoto, T. Watanabe, A. Takada, H. Goto, T. Horimoto, and Y. Kawaoka, "Importance of both the coding and the segment-specific noncoding regions of the Influenza A virus NS segment for its efficient incorporation into virions," *J. Virol.*, vol. 79, pp. 3766–3774, Mar. 2005.
- [237] K. Fujii, M. Ozawa, K. Iwatsuki-Horimoto, T. Horimoto, and Y. Kawaoka, "Incorporation of Influenza A virus genome segments does not absolutely require wild-type sequences," *Journal of General Virology*, vol. 90, pp. 1734–1740, July 2009.
- [238] S. Duhaut and N. J. Dimmock, "Approximately 150 nucleotides from the 5' End of an Influenza A segment 1 defective virion RNA are needed for genome stability during passage of defective virus in infected cells," *Virology*, vol. 275, pp. 278–285, Sept. 2000.
- [239] S. D. Duhaut and N. J. Dimmock, "Defective segment 1 RNAs that interfere with production of infectious Influenza A virus require at least 150 nucleotides of 5' sequence: Evidence from a plasmid-driven system," *Journal of General Virology*, vol. 83, pp. 403–411, Feb. 2002.

- [240] L. Sherry, K. Punovuori, L. E. Wallace, E. Pranglely, S. DeFries, and D. Jackson, “Identification of cis-acting packaging signals in the coding regions of the Influenza B virus HA gene segment,” *Journal of General Virology*, vol. 97, pp. 306–315, Feb. 2016.
- [241] Y. Liang, T. Huang, H. Ly, T. G. Parslow, and Y. Liang, “Mutational analyses of packaging signals in Influenza virus PA, PB1, and PB2 genomic RNA segments,” *JVI*, vol. 82, pp. 229–236, Jan. 2008.
- [242] N. Takizawa, K. Higashi, R. K. Kawaguchi, Y. Gotoh, Y. Suzuki, T. Hayashi, and K. Kurokawa, “Comprehensive in virio structure probing analysis of the Influenza A virus identifies a functional RNA structure involved in replication and segment interactions,” preprint, *Microbiology*, Mar. 2020.
- [243] E. Fournier, V. Moules, B. Essere, J.-C. Paillart, J.-D. Sirbat, A. Cavalier, J.-P. Rolland, D. Thomas, B. Lina, C. Isel, and R. Marquet, “Interaction network linking the human H3N2 Influenza A virus genomic RNA segments,” *Vaccine*, vol. 30, pp. 7359–7367, Dec. 2012.
- [244] C. Gavazzi, M. Yver, C. Isel, R. P. Smyth, M. Rosa-Calatrava, B. Lina, V. Moulès, and R. Marquet, “A functional sequence-specific interaction between Influenza A virus genomic RNA segments,” *PNAS*, vol. 110, pp. 16604–16609, Oct. 2013.
- [245] Y.-y. Chou, N. S. Heaton, Q. Gao, P. Palese, R. Singer, and T. Lionnet, “Colocalization of different Influenza viral RNA segments in the cytoplasm before viral budding as shown by single-molecule sensitivity FISH analysis,” *PLoS Pathog*, vol. 9, p. e1003358, May 2013.
- [246] S. S. Lakdawala, Y. Wu, P. Wawrzusin, J. Kabat, A. J. Broadbent, E. W. Lamirande, E. Fodor, N. Altan-Bonnet, H. Shroff, and K. Subbarao, “Influenza A virus assembly intermediates fuse in the cytoplasm,” *PLoS Pathog*, vol. 10, p. e1003971, Mar. 2014.
- [247] B. Essere, M. Yver, C. Gavazzi, O. Terrier, C. Isel, E. Fournier, F. Giroux, J. Textoris, T. Julien, C. Socratous, M. Rosa-Calatrava, B. Lina, R. Marquet, and V. Moules, “Critical role of segment-specific packaging signals in genetic reassortment of Influenza A viruses,” *Proceedings of the National Academy of Sciences*, vol. 110, pp. E3840–E3848, Oct. 2013.
- [248] B. Crescenzo-Chaigne, C. V. S. Barbezange, S. Léandri, C. Roquin, C. Berthault, and S. van der Werf, “Incorporation of the Influenza A virus NA segment into virions does not require cognate non-coding sequences,” *Sci Rep*, vol. 7, p. 43462, Apr. 2017.
- [249] E. de Wit, M. I. Spronken, G. F. Rimmelzwaan, A. D. Osterhaus, and R. A. Fouchier, “Evidence for specific packaging of the Influenza A virus genome from conditionally defective virus particles lacking a polymerase gene,” *Vaccine*, vol. 24, pp. 6647–6650, Nov. 2006.
- [250] Q. Gao, Y.-Y. Chou, S. Doğanay, R. Vafabakhsh, T. Ha, and P. Palese, “The Influenza A virus PB2, PA, NP, and M segments play a pivotal role during genome packaging,” *J Virol*, vol. 86, pp. 7043–7051, July 2012.

- [251] H. Goto, Y. Muramoto, T. Noda, and Y. Kawaoka, "The genome-packaging signal of the Influenza A virus genome comprises a genome incorporation signal and a genome-bundling signal," *J Virol*, vol. 87, pp. 11316–11322, Nov. 2013.
- [252] É. A. Moreira, A. Weber, H. Bolte, L. Kolesnikova, S. Giese, S. Lakdawala, M. Beer, G. Zimmer, A. García-Sastre, M. Schwemmle, and M. Juozapaitis, "A conserved Influenza A virus nucleoprotein code controls specific viral genome packaging," *Nat Commun*, vol. 7, p. 12861, Nov. 2016.
- [253] M. W. Pons, I. T. Schulze, and G. K. Hirst, "Isolation and characterization of the ribonucleoprotein of Influenza virus," *Virology*, vol. 39, pp. 250–259, Oct. 1969.
- [254] R. W. H. Ruigrok and F. Baudin, "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," *Journal of General Virology*, vol. 76, pp. 1009–1014, Apr. 1995.
- [255] B. Tarus, O. Bakowicz, S. Chenavas, L. Duchemin, L. Estrozi, C. Bourdieu, N. Lejal, J. Bernard, M. Moudjou, C. Chevalier, B. Delmas, R. Ruigrok, C. Di Primo, and A. Slama-Schwok, "Oligomerization paths of the nucleoprotein of Influenza A virus," *Biochimie*, vol. 94, pp. 776–785, Mar. 2012.
- [256] S. Chenavas, L. F. Estrozi, A. Slama-Schwok, B. Delmas, C. Di Primo, F. Baudin, X. Li, T. Crépin, and R. W. H. Ruigrok, "Monomeric nucleoprotein of Influenza A virus," *PLoS Pathog*, vol. 9, p. e1003275, Mar. 2013.
- [257] A. Labaronne, C. Swale, A. Monod, G. Schoehn, T. Crépin, and R. Ruigrok, "Binding of RNA by the nucleoproteins of Influenza viruses A and B," *Viruses*, vol. 8, p. 247, Sept. 2016.
- [258] E. C. Hutchinson, E. M. Denham, B. Thomas, D. C. Trudgian, S. S. Hester, G. Ridlova, A. York, L. Turrell, and E. Fodor, "Mapping the phosphoproteome of Influenza A and B viruses by mass spectrometry," *PLoS Pathog*, vol. 8, p. e1002993, Nov. 2012.
- [259] L. Turrell, E. C. Hutchinson, F. T. Vreede, and E. Fodor, "Regulation of Influenza A virus nucleoprotein oligomerization by phosphorylation," *J Virol*, vol. 89, pp. 1452–1455, Jan. 2015.
- [260] K. Yamanaka, A. Ishihama, and K. Nagata, "Reconstitution of Influenza virus RNA-nucleoprotein complexes structurally resembling native viral ribonucleoprotein cores," *J Biol Chem*, vol. 265, pp. 11151–11155, July 1990.
- [261] Q. Ye, R. M. Krug, and Y. J. Tao, "The mechanism by which Influenza A virus nucleoprotein forms oligomers and binds RNA," *Nature*, vol. 444, pp. 1078–1082, Dec. 2006.
- [262] A. K.-L. Ng, H. Zhang, K. Tan, Z. Li, J.-h. Liu, P. K.-S. Chan, S.-M. Li, W.-Y. Chan, S. W.-N. Au, A. Joachimiak, T. Walz, J.-H. Wang, and P.-C. Shaw, "Structure of the Influenza virus A H5N1 nucleoprotein : Implications for RNA binding, oligomerization, and vaccine design," *FASEB j*, vol. 22, pp. 3638–3647, Oct. 2008.

- [263] M. L. Knight, H. Fan, D. L. V. Bauer, J. M. Grimes, E. Fodor, and J. R. Keown, "Structure of an H₃N₂ Influenza virus nucleoprotein," *Acta Crystallogr F Struct Biol Commun*, vol. 77, pp. 208–214, July 2021.
- [264] Y.-S. Tang, S. Xu, Y.-W. Chen, J.-H. Wang, and P.-C. Shaw, "Crystal structures of Influenza nucleoprotein complexed with nucleic acid provide insights into the mechanism of RNA interaction," *Nucleic Acids Research*, vol. 49, pp. 4144–4154, Apr. 2021.
- [265] W.-H. Chan, A. K.-L. Ng, N. C. Robb, M. K.-H. Lam, P. K.-S. Chan, S. W.-N. Au, J.-H. Wang, E. Fodor, and P.-C. Shaw, "Functional analysis of the Influenza virus H₅N₁ nucleoprotein tail loop reveals amino acids that are crucial for oligomerization and ribonucleoprotein activities," *J Virol*, vol. 84, pp. 7337–7345, July 2010.
- [266] S. Boulo, H. Akarsu, V. Lotteau, C. W. Müller, R. W. Ruigrok, and F. Baudin, "Human importin alpha and RNA do not compete for binding to influenza A virus nucleoprotein," *Virology*, vol. 409, pp. 84–90, Jan. 2011.
- [267] M. Kobayashi, T. Toyoda, D. M. Adyshev, Y. Azuma, and A. Ishihama, "Molecular dissection of Influenza virus nucleoprotein : Deletion mapping of the RNA binding domain," *J Virol*, vol. 68, pp. 8433–8436, Dec. 1994.
- [268] C. Albo, A. Valencia, and A. Portela, "Identification of an RNA binding region within the N-terminal third of the influenza A virus nucleoprotein," *J Virol*, vol. 69, pp. 3799–3806, June 1995.
- [269] A. K.-L. Ng, M. K.-H. Lam, H. Zhang, J. Liu, S. W.-N. Au, P. K.-S. Chan, J. Wang, and P.-C. Shaw, "Structural basis for RNA binding and homo-oligomer formation by Influenza B virus nucleoprotein," *J Virol*, vol. 86, pp. 6758–6767, June 2012.
- [270] A. Donchet, J. Oliva, A. Labaronne, L. Tengo, M. Miloudi, F. C.A. Gerard, C. Mas, G. Schoehn, R. W.H. Ruigrok, M. Ducatez, and T. Crépin, "The structure of the nucleoprotein of Influenza D shows that all Orthomyxoviridae nucleoproteins have a similar NPCORE, with or without a NPTAIL for nuclear transport," *Sci Rep*, vol. 9, p. 600, Dec. 2019.
- [271] W. Zheng, J. Olson, V. Vakharia, and Y. J. Tao, "The crystal structure and RNA-binding of an Orthomyxovirus nucleoprotein," *PLoS Pathog*, vol. 9, p. e1003624, Sept. 2013.
- [272] D. W. Kingsbury, I. M. Jones, and K. Murti, "Assembly of Influenza ribonucleoprotein in vitro using recombinant nucleoprotein," *Virology*, vol. 156, pp. 396–403, Feb. 1987.
- [273] J. Ortega, J. Martín-Benito, T. Zürcher, J. M. Valpuesta, J. L. Carrascosa, and J. Ortín, "Ultrastructural and functional analyses of recombinant Influenza virus ribonucleoproteins suggest dimerization of nucleoprotein during virus amplification," *J Virol*, vol. 74, pp. 156–163, Jan. 2000.
- [274] G. D. Williams, D. Townsend, K. M. Wylie, P. J. Kim, G. K. Amarasinghe, S. B. Kutluay, and A. C. M. Boon, "Nucleotide resolution mapping of Influenza A virus nucleoprotein-RNA interactions reveals RNA features required for replication," *Nat Commun*, vol. 9, p. 465, Dec. 2018.

- [275] N. Lee, V. Le Sage, A. V. Nanni, D. J. Snyder, V. S. Cooper, and S. S. Lakdawala, "Genome-wide analysis of Influenza viral RNA and nucleoprotein association," *Nucleic Acids Research*, vol. 45, pp. 8968–8977, Sept. 2017.
- [276] V. Le Sage, A. Nanni, A. Bhagwat, D. Snyder, V. Cooper, S. Lakdawala, and N. Lee, "Non-uniform and non-random binding of nucleoprotein to Influenza A and B viral RNA," *Viruses*, vol. 10, p. 522, Sept. 2018.
- [277] B. Dadonaite, B. Gilbertson, M. L. Knight, S. Trifkovic, S. Rockman, A. Laederach, L. E. Brown, E. Fodor, and D. L. V. Bauer, "The structure of the Influenza A virus genome," *Nat Microbiol*, vol. 4, pp. 1781–1789, Nov. 2019.
- [278] V. Le Sage, J. P. Kanarek, D. J. Snyder, V. S. Cooper, S. S. Lakdawala, and N. Lee, "Mapping of Influenza virus RNA-RNA interactions reveals a flexible network," *Cell Reports*, vol. 31, p. 107823, June 2020.
- [279] S. Miyamoto and T. Noda, "In vitro vRNA–vRNA interactions in the H1N1 Influenza A virus genome," *Microbiol Immunol*, vol. 64, pp. 202–209, Mar. 2020.
- [280] S. Miyamoto, Y. Muramoto, K. Shindo, Y. Fujita-Fujiharu, T. Morikawa, R. Tamura, J. L. Gilmore, M. Nakano, and T. Noda, "Contribution of RNA-RNA interactions mediated by the genome packaging signals for the selective genome packaging of Influenza A virus," *J Virol*, vol. 96, pp. e01641–21, Mar. 2022.
- [281] R. J. Hagey, M. Elazar, S. Tian, E. A. Pham, W. Kladwang, L. Ben-Avi, K. Nguyen, A. Xiong, M. Rabinovich, S. Schaffert, T. Avisar, B. Fram, P. Liu, P. Khatri, J. K. Taubenberger, R. Das, and J. S. Glenn, "Identification and targeting of a pan-genotypic Influenza A virus RNA structure that mediates packaging and disease," preprint, Biorxiv, Aug. 2021.
- [282] B. Gilbertson, T. Zheng, M. Gerber, A. Printz-Schweigert, C. Ong, R. Marquet, C. Isel, S. Rockman, and L. Brown, "Influenza NA and PB1 gene segments interact during the formation of viral progeny : Localization of the binding region within the PB1 gene," *Viruses*, vol. 8, p. 238, Aug. 2016.
- [283] S. Jackson, N. Van Hoeven, L.-M. Chen, T. R. Maines, N. J. Cox, J. M. Katz, and R. O. Donis, "Reassortment between Avian H5N1 and Human H3N2 Influenza Viruses in ferrets : A public health risk assessment," *J Virol*, vol. 83, pp. 8131–8140, Aug. 2009.
- [284] C. P. Octaviani, M. Ozawa, S. Yamada, H. Goto, and Y. Kawaoka, "High level of genetic compatibility between swine-origin H1N1 and highly pathogenic avian H5N1 Influenza viruses," *J Virol*, vol. 84, pp. 10918–10922, Oct. 2010.
- [285] J. C. A. Cobbin, C. Ong, E. Verity, B. P. Gilbertson, S. P. Rockman, and L. E. Brown, "Influenza virus PB1 and neuraminidase gene segments can cosegregate during vaccine reassortment driven by interactions in the PB1 coding region," *J Virol*, vol. 88, pp. 8971–8980, Aug. 2014.
- [286] C. Bergeron, M. Valette, B. Lina, and M. Ottmann, "Genetic content of Influenza H3N2 vaccine seeds," *PLoS Curr*, vol. 2, p. RRN1165, Sept. 2010.

- [287] A. A. Fulvini, M. Ramanunninair, J. Le, B. A. Pokorny, J. M. Arroyo, J. Silverman, R. Devis, and D. Bucher, “Gene constellation of Influenza A virus reassortants with high growth phenotype prepared as seed candidates for vaccine production,” *PLoS ONE*, vol. 6, p. e20823, June 2011.
- [288] J. C. A. Cobbin, E. E. Verity, B. P. Gilbertson, S. P. Rockman, and L. E. Brown, “The source of the PB1 gene in Influenza vaccine reassortants selectively alters the hemagglutinin content of the resulting seed virus,” *J Virol*, vol. 87, pp. 5577–5585, May 2013.
- [289] X. Li, M. Gu, Q. Zheng, R. Gao, and X. Liu, “Packaging signal of Influenza A virus,” *Virology*, vol. 18, p. 36, Dec. 2021.
- [290] M. I. Nelson and E. C. Holmes, “The evolution of epidemic Influenza,” *Nat Rev Genet*, vol. 8, pp. 196–205, Mar. 2007.
- [291] J. R. Gallagher, U. Torian, D. M. McCraw, and A. K. Harris, “Structural studies of Influenza virus RNPs by electron microscopy indicate molecular contortions within NP supra-structures,” *Journal of Structural Biology*, vol. 197, pp. 294–307, Mar. 2017.
- [292] Y. Kobayashi, B. Dadonaite, N. van Doremalen, Y. Suzuki, W. S. Barclay, and O. G. Pybus, “Computational and molecular analysis of conserved Influenza A virus RNA secondary structures involved in infectious virion production,” *RNA Biology*, vol. 13, pp. 883–894, Sept. 2016.
- [293] A. P. Gultyaev, M. I. Spronken, M. Richard, E. J. A. Schrauwen, R. C. L. Olsthoorn, and R. A. M. Fouchier, “Subtype-specific structural constraints in the evolution of Influenza A virus hemagglutinin genes,” *Sci Rep*, vol. 6, p. 38892, Dec. 2016.
- [294] A. P. Gultyaev, A. Tsyganov-Bodounov, M. I. Spronken, S. van der Kooij, R. A. Fouchier, and R. C. Olsthoorn, “RNA structural constraints in the evolution of the Influenza A virus genome NP segment,” *RNA Biology*, vol. 11, pp. 942–952, July 2014.
- [295] Y. Kobayashi, O. G. Pybus, T. Itou, and Y. Suzuki, “Conserved secondary structures predicted within the 5′ packaging signal region of Influenza A virus PB2 segment,” *Meta Gene*, vol. 15, pp. 75–79, Feb. 2018.
- [296] J. Nowakowski and I. Tinoco, “RNA structure and stability,” *Seminars in Virology*, vol. 8, no. 3, pp. 153–165, 1997.
- [297] D. Herschlag, “RNA chaperones and the RNA folding problem,” *Journal of Biological Chemistry*, vol. 270, pp. 20871–20874, Sept. 1995.
- [298] S. Chauhan and S. A. Woodson, “Tertiary interactions determine the accuracy of RNA folding,” *J. Am. Chem. Soc.*, vol. 130, pp. 1296–1303, Jan. 2008.
- [299] K. Musier-Forsyth, “RNA remodeling by chaperones and helicases,” *RNA Biology*, vol. 7, pp. 632–633, Nov. 2010.

- [300] G. Cristofari and J.-L. Darlix, "The ubiquitous nature of RNA chaperone proteins," in *Progress in Nucleic Acid Research and Molecular Biology*, vol. 72, pp. 223–268, Elsevier, 2002.
- [301] R. Schroeder, A. Barta, and K. Semrad, "Strategies for RNA folding and assembly," *Nat Rev Mol Cell Biol*, vol. 5, pp. 908–919, Nov. 2004.
- [302] P. Tompa and P. Csermely, "The role of structural disorder in the function of RNA and protein chaperones," *FASEBj*, vol. 18, pp. 1169–1175, Aug. 2004.
- [303] L. Rajkowitsch, K. Semrad, O. Mayer, and R. Schroeder, "Assays for the RNA chaperone activity of proteins," *Biochemical Society Transactions*, vol. 33, pp. 450–456, June 2005.
- [304] L. Rajkowitsch, D. Chen, S. Stampfl, K. Semrad, C. Waldsich, O. Mayer, M. F. Jantsch, R. Konrat, U. Bläsi, and R. Schroeder, "RNA chaperones, RNA annealers and RNA helicases," *RNA Biology*, vol. 4, pp. 118–130, July 2007.
- [305] R. Ivanyi-Nagy, J.-P. Lavergne, C. Gabus, D. Ficheux, and J.-L. Darlix, "RNA chaperoning and intrinsic disorder in the core proteins of Flaviviridae," *Nucleic Acids Research*, vol. 36, pp. 712–725, Feb. 2008.
- [306] S. Zúñiga, I. Sola, J. L. Cruz, and L. Enjuanes, "Role of RNA chaperones in virus replication," *Virus Research*, vol. 139, pp. 253–266, Feb. 2009.
- [307] M. Doetsch, R. Schroeder, and B. Fürtig, "Transient RNA-protein interactions in RNA folding," *FEBS Journal*, vol. 278, pp. 1634–1642, May 2011.
- [308] S. A. Woodson, S. Panja, and A. Santiago-Frangos, "Proteins that chaperone RNA regulation," *Microbiol Spectr*, vol. 6, p. 6.4.21, July 2018.
- [309] S. Mohr, J. M. Stryker, and A. M. Lambowitz, "A DEAD-Box protein functions as an ATP-dependent RNA chaperone in group I intron splicing," *Cell*, vol. 109, pp. 769–779, June 2002.
- [310] J. Yang, H. Xia, Q. Qian, and X. Zhou, "RNA chaperones encoded by RNA viruses," *Virolog. Sin.*, vol. 30, pp. 401–409, Dec. 2015.
- [311] G. Cristofari, "The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA in vitro," *Nucleic Acids Research*, vol. 32, pp. 2623–2631, Apr. 2004.
- [312] R. Ivanyi-Nagy, I. Kanevsky, C. Gabus, J.-P. Lavergne, D. Ficheux, F. Penin, P. Fosse, and J.-L. Darlix, "Analysis of Hepatitis C virus RNA dimerization and core-RNA interactions," *Nucleic Acids Research*, vol. 34, pp. 2618–2633, May 2006.
- [313] E. D. Holmstrom, Z. Liu, D. Nettels, R. B. Best, and B. Schuler, "Disordered RNA chaperones can enhance nucleic acid folding via local charge screening," *Nat Commun*, vol. 10, p. 2453, Dec. 2019.
- [314] Z.-S. Huang and H.-N. Wu, "Identification and characterization of the RNA chaperone activity of Hepatitis Delta Antigen peptides," *Journal of Biological Chemistry*, vol. 273, pp. 26455–26461, Oct. 1998.

- [315] Z.-S. Huang, W.-H. Su, J.-L. Wang, and H.-N. Wu, "Selective strand annealing and selective strand exchange promoted by the N-terminal domain of Hepatitis Delta Antigen," *Journal of Biological Chemistry*, vol. 278, pp. 5685–5693, Feb. 2003.
- [316] K. S. Jeng, P. Y. Su, and M. M. Lai, "Hepatitis Delta Antigens enhance the ribozyme activities of Hepatitis Delta virus RNA in vivo," *J Virol*, vol. 70, pp. 4205–4209, July 1996.
- [317] C.-C. Wang, "Nucleic acid binding properties of the nucleic acid chaperone domain of Hepatitis Delta Antigen," *Nucleic Acids Research*, vol. 31, pp. 6481–6492, Nov. 2003.
- [318] S. Zúñiga, I. Sola, J. L. Moreno, P. Sabella, J. Plana-Durán, and L. Enjuanes, "Coronavirus nucleocapsid protein is an RNA chaperone," *Virology*, vol. 357, pp. 215–227, Jan. 2007.
- [319] L. Cui, H. Wang, Y. Ji, J. Yang, S. Xu, X. Huang, Z. Wang, L. Qin, P. Tien, X. Zhou, D. Guo, and Y. Chen, "The nucleocapsid protein of Coronaviruses acts as a viral suppressor of RNA silencing in mammalian cells," *J Virol*, vol. 89, pp. 9029–9043, Sept. 2015.
- [320] J. J. DeStefano and O. Titilope, "Poliovirus protein 3AB displays nucleic acid chaperone and helix-destabilizing activities," *J Virol*, vol. 80, pp. 1662–1671, Feb. 2006.
- [321] F. Tang, H. Xia, P. Wang, J. Yang, T. Zhao, Q. Zhang, Y. Hu, and X. Zhou, "The identification and characterization of nucleic acid chaperone activity of human enterovirus 71 nonstructural protein 3AB," *Virology*, vol. 464–465, pp. 353–364, Sept. 2014.
- [322] T. Jackson and G. J. Belsham, "Picornaviruses : A view from 3A," *Viruses*, vol. 13, p. 456, Mar. 2021.
- [323] Z. Cheng, J. Yang, H. Xia, Y. Qiu, Z. Wang, Y. Han, X. Xia, C.-F. Qin, Y. Hu, and X. Zhou, "The nonstructural protein 2C of a Picorna-like virus displays nucleic acid helix destabilizing activity that can be functionally separated from its ATPase activity," *J Virol*, vol. 87, pp. 5205–5218, May 2013.
- [324] S.-H. Wang, K. Wang, K. Zhao, S.-C. Hua, and J. Du, "The structure, function, and mechanisms of action of Enterovirus non-structural protein 2C," *Front. Microbiol.*, vol. 11, p. 615965, Dec. 2020.
- [325] M. A. Mir and A. T. Panganiban, "The Bunyavirus nucleocapsid protein is an RNA chaperone : Possible roles in viral RNA panhandle formation and genome replication," *RNA*, vol. 12, pp. 272–282, Feb. 2006.
- [326] Z. F. Taraporewala and J. T. Patton, "Identification and characterization of the helix-destabilizing activity of Rotavirus Nonstructural Protein NSP2," *J Virol*, vol. 75, pp. 4519–4527, May 2001.
- [327] J. Yang, Z. Cheng, S. Zhang, W. Xiong, H. Xia, Y. Qiu, Z. Wang, F. Wu, C.-F. Qin, L. Yin, Y. Hu, and X. Zhou, "A cypovirus VP5 displays the RNA chaperone-like activity that destabilizes RNA helices and accelerates strand annealing," *Nucleic Acids Research*, vol. 42, pp. 2538–2554, Feb. 2014.
- [328] W.-L. Pong, Z.-S. Huang, P.-G. Teoh, C.-C. Wang, and H.-N. Wu, "RNA binding property and RNA chaperone activity of dengue virus core protein and other viral RNA-interacting proteins," *FEBS Letters*, vol. 585, pp. 2575–2581, Aug. 2011.

- [329] X. E. Yong, P. V. Raghuvamsi, G. S. Anand, T. Wohland, and K. K. Sharma, "Dengue virus strain 2 capsid protein switches the annealing pathway and reduces intrinsic dynamics of the conserved 5' untranslated region," *RNA Biology*, vol. 18, pp. 718–731, May 2021.
- [330] X. E. Yong, V. R. Palur, G. S. Anand, T. Wohland, and K. K. Sharma, "Dengue virus 2 capsid protein chaperones the strand displacement of 5'-3' cyclization sequences," *Nucleic Acids Research*, vol. 49, pp. 5832–5844, June 2021.
- [331] D. Muriaux, H. De Rocquigny, B.-P. Roques, and J. Paoletti, "NCp7 activates HIV-1Lai RNA dimerization by converting a transient Loop-Loop complex into a stable dimer," *Journal of Biological Chemistry*, vol. 271, pp. 33686–33692, Dec. 1996.
- [332] Y.-X. Feng, S. Campbell, D. Harvin, B. Ehresmann, C. Ehresmann, and A. Rein, "The Human Immunodeficiency Virus Type 1 Gag polyprotein has nucleic acid chaperone activity : Possible role in dimerization of genomic RNA and placement of tRNA on the Primer Binding Site," *J Virol*, vol. 73, pp. 4251–4256, May 1999.
- [333] S. Bernacchi, S. Stoylov, E. Piémont, D. Ficheux, B. P. Roques, J. L. Darlix, and Y. Mély, "HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence," *Journal of Molecular Biology*, vol. 317, pp. 385–399, Mar. 2002.
- [334] M. A. Urbaneja, M. Wu, J. R. Casas-Finet, and R. L. Karpel, "HIV-1 nucleocapsid protein as a nucleic acid chaperone : Spectroscopic study of its helix-destabilizing properties, structural binding specificity, and annealing activity," *Journal of Molecular Biology*, vol. 318, pp. 749–764, May 2002.
- [335] H. Beltz, J. Azoulay, S. Bernacchi, J.-P. Clamme, D. Ficheux, B. Roques, J.-L. Darlix, and Y. Mély, "Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7," *Journal of Molecular Biology*, vol. 328, pp. 95–108, Apr. 2003.
- [336] J. G. Levin, M. Mitra, A. Mascarenhas, and K. Musier-Forsyth, "Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription," *RNA Biology*, vol. 7, pp. 754–774, Nov. 2010.
- [337] M. Kuciak, C. Gabus, R. Ivanyi-Nagy, K. Semrad, R. Storchak, O. Chaloin, S. Muller, Y. Mély, and J.-L. Darlix, "The HIV-1 transcriptional activator Tat has potent nucleic acid chaperoning activities in vitro," *Nucleic Acids Research*, vol. 36, pp. 3389–3400, June 2008.
- [338] C. Boudier, R. Storchak, K. Sharma, P. Didier, A. Follenius-Wund, S. Muller, J.-L. Darlix, and Y. Mély, "The mechanism of HIV-1 Tat-directed nucleic acid annealing supports its role in reverse transcription," *Journal of Molecular Biology*, vol. 400, pp. 487–501, July 2010.
- [339] S. Henriët, L. Sinck, G. Bec, R. J. Gorelick, R. Marquet, and J.-C. Paillart, "Vif is a RNA chaperone that could temporally regulate RNA dimerization and the early steps of HIV-1 reverse transcription," *Nucleic Acids Research*, vol. 35, pp. 5141–5153, July 2007.

- [340] J. Batisse, S. Guerrero, S. Bernacchi, D. Sleiman, C. Gabus, J.-L. Darlix, R. Marquet, C. Tisné, and J.-C. Paillart, “The role of Vif oligomerization and RNA chaperone activity in HIV-1 replication,” *Virus Research*, vol. 169, pp. 361–376, Nov. 2012.
- [341] T.-F. Li, M. Hosmillo, H. Schwanke, T. Shu, Z. Wang, L. Yin, S. Curry, I. G. Goodfellow, and X. Zhou, “Human Norovirus NS3 has RNA helicase and chaperoning activities,” *J Virol*, vol. 92, pp. e01606–17, Mar. 2018.
- [342] K. R. Han, J.-H. Lee, G. G. Kotiguda, K. H. Jung, M. S. Chung, S. Kang, S. Hwang, and K. H. Kim, “Nucleotide triphosphatase and RNA chaperone activities of murine norovirus NS3,” *Journal of General Virology*, vol. 99, pp. 1482–1493, Nov. 2018.
- [343] J. K. Grohman, R. J. Gorelick, C. R. Lickwar, J. D. Lieb, B. D. Bower, B. M. Znosko, and K. M. Weeks, “A guanosine-centric mechanism for RNA chaperone function,” *Science*, vol. 340, pp. 190–195, Apr. 2013.
- [344] H. Xia, P. Wang, G.-C. Wang, J. Yang, X. Sun, W. Wu, Y. Qiu, T. Shu, X. Zhao, L. Yin, C.-F. Qin, Y. Hu, and X. Zhou, “Human Enterovirus nonstructural protein 2CATPase functions as both an RNA helicase and ATP-independent RNA chaperone,” *PLoS Pathog*, vol. 11, p. e1005067, July 2015.
- [345] O. Gilmer, E. Quignon, A.-C. Jousset, J.-C. Paillart, R. Marquet, and V. Vivet-Boudou, “Chemical and enzymatic probing of viral RNAs : From infancy to maturity and beyond,” *Viruses*, vol. 13, p. 1894, Oct. 2021.
- [346] F. Karabiber, J. L. McGinnis, O. V. Favorov, and K. M. Weeks, “QuShape : Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis,” *RNA*, vol. 19, pp. 63–73, Jan. 2013.
- [347] E. J. Merino, K. A. Wilkinson, J. L. Coughlan, and K. M. Weeks, “RNA structure analysis at single nucleotide resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE),” *J. Am. Chem. Soc.*, vol. 127, pp. 4223–4231, Mar. 2005.
- [348] P. Y. Wang, A. N. Sexton, W. J. Culligan, and M. D. Simon, “Carbodiimide reagents for the chemical probing of RNA structure in cells,” *RNA*, vol. 25, pp. 135–146, Jan. 2019.
- [349] D. Mitchell, A. J. Renda, C. A. Douds, P. Babitzke, S. M. Assmann, and P. C. Bevilacqua, “In vivo RNA structural probing of uracil and guanine base-pairing by 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC),” *RNA*, vol. 25, pp. 147–157, Jan. 2019.
- [350] S. A. Mortimer and K. M. Weeks, “A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry,” *J. Am. Chem. Soc.*, vol. 129, pp. 4144–4145, Apr. 2007.
- [351] P. Tijerina, S. Mohr, and R. Russell, “DMS footprinting of structured RNAs and RNA–protein complexes,” *Nat Protoc*, vol. 2, pp. 2608–2623, Oct. 2007.
- [352] H. D. Chirra, T. Sexton, D. Biswal, L. B. Hersh, and J. Z. Hilt, “Catalase-coupled gold nanoparticles : Comparison between the carbodiimide and biotin–streptavidin methods,” *Acta Biomaterialia*, vol. 7, pp. 2865–2872, July 2011.





- [353] C. K. Kwok, A. B. Sahakyan, and S. Balasubramanian, "Structural analysis using SHALiPE to reveal RNA G-quadruplex formation in human precursor MicroRNA," *Angew. Chem. Int. Ed.*, vol. 55, pp. 8958–8961, July 2016.
- [354] J. S. Reuter and D. H. Mathews, "RNAstructure : Software for RNA secondary structure prediction and analysis," *BMC Bioinformatics*, vol. 11, p. 129, Dec. 2010.
- [355] K. Darty, A. Denise, and Y. Ponty, "VARNA : Interactive drawing and editing of the RNA secondary structure," *Bioinformatics*, vol. 25, pp. 1974–1975, Aug. 2009.
- [356] M. Antczak, T. Zok, M. Popenda, P. Lukasiak, R. W. Adamiak, J. Blazewicz, and M. Szachniuk, "RNAPdbec— a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs," *Nucleic Acids Research*, vol. 42, pp. W368–W372, July 2014.
- [357] M. Antczak, M. Popenda, T. Zok, M. Zurkowski, R. W. Adamiak, and M. Szachniuk, "New algorithms to represent complex pseudoknotted RNA structures in dot-bracket notation," *Bioinformatics*, vol. 34, pp. 1304–1312, Apr. 2018.
- [358] T. Zok, M. Antczak, M. Zurkowski, M. Popenda, J. Blazewicz, R. W. Adamiak, and M. Szachniuk, "RNAPdbec 2.0 : Multifunctional tool for RNA structure annotation," *Nucleic Acids Research*, vol. 46, pp. W30–W35, July 2018.
- [359] V. Tsybul'skyi, M. Mounir, and I. M. Meyer, "R-chie : A web server and R package for visualizing cis and trans RNA–RNA, RNA–DNA and DNA–DNA interactions," *Nucleic Acids Research*, vol. 48, pp. e105–e105, Oct. 2020.
- [360] M. D. Schmidt, A. Kirkpatrick, and C. Heitsch, "RNAStructViz : Graphical base pairing analysis," *Bioinformatics*, vol. 37, pp. 3660–3661, Oct. 2021.
- [361] M. J. Smola, J. M. Calabrese, and K. M. Weeks, "Detection of RNA–protein Interactions in living cells with SHAPE," *Biochemistry*, vol. 54, pp. 6867–6875, Nov. 2015.
- [362] Y. Zhang, J. Wang, and Y. Xiao, "3dRNA : 3D structure prediction from linear to circular RNAs," *Journal of Molecular Biology*, vol. 434, p. 167452, June 2022.
- [363] M. Gerber, *Recherche de Signaux d'empaquetage Spécifiques Du Génome Des Virus Influenza A*. These de doctorat, Strasbourg, Sept. 2016.
- [364] Z. Gu, L. Gu, R. Eils, M. Schlesner, and B. Brors, "Circlize implements and enhances circular visualization in R," *Bioinformatics*, vol. 30, pp. 2811–2812, Oct. 2014.
- [365] M. Szabat and R. Kierzek, "Parallel-stranded DNA and RNA duplexes - structural features and potential applications," *FEBSJ*, vol. 284, pp. 3986–3998, Dec. 2017.
- [366] S. Busan and K. M. Weeks, "Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2," *RNA*, vol. 24, pp. 143–148, Feb. 2018.

- [367] C. Y. Cheng, F.-C. Chou, W. Kladwang, S. Tian, P. Cordero, and R. Das, “Consistent global structures of complex RNA states through multidimensional chemical mapping,” *eLife*, vol. 4, p. e07600, June 2015.
- [368] H. Ma, X. Jia, K. Zhang, and Z. Su, “Cryo-EM advances in RNA structure determination,” *Sig Transduct Target Ther*, vol. 7, p. 58, Dec. 2022.
- [369] G. Kudla, S. Granneman, D. Hahn, J. D. Beggs, and D. Tollervy, “Cross-linking, ligation, and sequencing of hybrids reveals RNA–RNA interactions in yeast,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, pp. 10010–10015, June 2011.
- [370] V. Ramani, R. Qiu, and J. Shendure, “High-throughput determination of RNA structure by proximity ligation,” *Nat Biotechnol*, vol. 33, pp. 980–984, Sept. 2015.
- [371] R. A. Schäfer and B. Voß, “RNANUE : Efficient data analysis for RNA–RNA interactomics,” *Nucleic Acids Research*, vol. 49, pp. 5493–5501, June 2021.
- [372] X. Qiang and Z. Kou, “Prediction of interspecies transmission for avian Influenza A virus based on a back-propagation neural network,” *Mathematical and Computer Modelling*, vol. 52, pp. 2060–2065, Dec. 2010.
- [373] J. Wang, C. Ma, Z. Kou, Y. H. Zhou, and H. L. Liu, “Predicting transmission of avian Influenza A viruses from avian to human by using informative physicochemical properties,” *IJDMB*, vol. 7, no. 2, p. 166, 2013.
- [374] C. L. Eng, J. C. Tong, and T. W. Tan, “Predicting host tropism of Influenza A virus proteins using random forest,” *BMC Med Genomics*, vol. 7, p. S1, Dec. 2014.
- [375] C. Eng, J. Tong, and T. Tan, “Predicting zoonotic risk of Influenza A viruses from host tropism protein signature using random forest,” *IJMS*, vol. 18, p. 1135, May 2017.
- [376] R. Yin, X. Zhou, S. Rashid, and C. K. Kwoh, “HopPER : An adaptive model for probability estimation of Influenza reassortment through host prediction,” *BMC Med Genomics*, vol. 13, p. 9, Dec. 2020.

Annexes

Review

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

Orian Gilmer [†], Erwan Quignon [†] , Anne-Caroline Jousset [†] , Jean-Christophe Paillart [†] , Roland Marquet ^{*} 
and Valérie Vivet-Boudou ^{*}

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

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

[†] These authors contributed equally to this review.

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

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



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

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

Received: 27 August 2021
Accepted: 16 September 2021
Published: 22 September 2021

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



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

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

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

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

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

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

2.1. Birth of RNA Structure Probing

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

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

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

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

2.2.1. Characteristics and Specificities of the Pre-SHAPE Probes

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

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

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

In the first experiments in the 1960s, detection of the modified nucleotides was achieved by paper or gel electrophoresis of the small fragments obtained after the successive digestion of RNAs and the laborious work of puzzle reconstruction [21,53]. From the 1970s, the detection of modification or cleavage sites was achieved by one of two methods, both of which involve ^{32}P (or sometimes ^{35}S) labeling. For probing experiments of small RNAs (or of the extremities of long RNAs), direct detection of strand scissions of 3' or 5' end-labeled RNAs is the method of choice. In that case, the cleavage of the ribo-phosphate backbone can, of course, be achieved by enzymatic probes or chemical probes, such as lead (II) or hydroxyl radicals. However, other chemical reagents that, under normal conditions, only modify RNA can be combined with a subsequent treatment that induces strand scission at the modification site. For instance, DMS modified guanosines, when reduced by NaBH_4 , then treated with aniline, results in strand cleavage [54]; whereas N3 alkylated cytosines need to be treated with hydrazine, and then with aniline, for strand scission [55]. It has to be noted that treatment with hydrazine is not specific, since it may induce cleavage at the U positions, although these nucleotides are not probed by DMS. DEPC is also compatible with this detection method when coupled with aniline treatment [56], as well as ENU, since the phosphotriester formed after ENU treatment are easily hydrolyzed under mild alkaline conditions [57]. The RNA fragments are separated by polyacrylamide gel electrophoresis and by comparison with a sequencing ladder; thus, the positions of the cleavages in the RNA molecule are identified.

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

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

Table 1. Cont.

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

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

2.2.2. Early Applications to Viral RNAs

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

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

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

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

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

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

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

RNA Structures of Human Pathogenic Viruses

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

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

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

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

The appearance of the acquired immunodeficiency syndrome (AIDS) pandemic in the 1980s has led to a large number of probing studies of the human immunodeficiency virus type 1 (HIV-1) genomic RNA (gRNA) at different stages of the viral cycle and with a wide range of probes. For example, the structural domains, located in the 5' terminal region, and their potential role in dimerization were studied by chemical probing with enzymatic and chemical probes [94,95]. *Neurospora crassa* endonuclease was used to study the tRNA^{Lys,3}-gRNA reverse transcription initiation complex, together with a panel of chemical probes [67,96]. In addition to the predictable interaction between the viral primer binding site (PBS) and the 3' extremity of the tRNA^{Lys,3}, the authors were able to identify additional and specific contacts between an A-rich loop (located upstream of the PBS and the anticodon loop of tRNA^{Lys,3}) by labeling the 3' extremity of the tRNA and performing a denaturing PAGE on the RNA digests. To gain insight into the three-dimensional structure of the gRNA-tRNA^{Lys,3} complex, a complementary study with hydroxyl radicals was also performed [73]. In 1997, a variety of chemical probes, including lead ions, were used to study the influence of the purines flanking the dimerization initiation site (DIS) palindromic sequence on HIV-1 gRNA dimerization [76]. Pb²⁺ ions were used to detect subtle structural alterations between monomeric and dimeric forms of gRNA. The data allowed for the conclusion that the three flanking purines played major roles in the kinetics of RNA dimerization and in the stability of the dimers formed [76]. Later on, lead ions were also used to monitor the specific binding of aminoglycosides at the DIS in vitro and in cells [97]. Structural studies of the 5'-UTR (untranslated region) of HIV-1 gRNA were important for understanding the fate of this RNA between translation and packaging. Based on enzymatic and chemical probing, different models with alternative long-range interactions have been published, which could drive gRNA to the translation or packaging pathways [98–100].

Enzymatic Viral RNAs: Self-Cleaving Ribozymes

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

3. The SHAPE (r)Evolution

Since the beginning of chemical probing, probes have been selected for their ability to highlight the exposure to the solvent, flexibility, or ss nature of the nucleotides. A combination of a few probes was usually required, in order to establish an experimentally validated secondary structure model of an RNA of interest. This changed in 2005, with the work of Merino et al. [105]. Indeed, in previous work on 2'-NH₂-nucleotides, this laboratory had noticed that the 2' position of nucleotides with a flexible sugar ring, which generally correspond to unpaired nucleotides, was more reactive than that of base-paired nucleotides [106]. Therefore, they designed a new chemical probe, the N-methylisatoic anhydride (NMIA), capable of acylating the 2'-hydroxyl group of nucleosides and tested

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

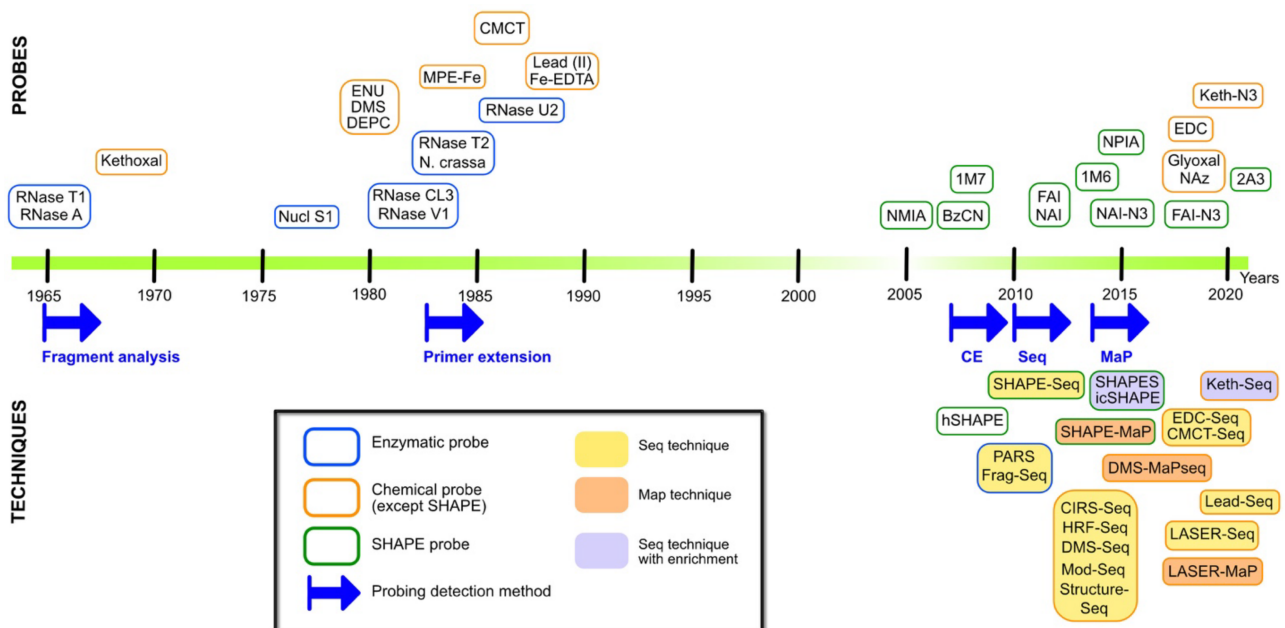


Figure 2. Timeline of RNA probes and probing techniques evolution. Upper part, enzymes are boxed in blue, chemical reagents (with the exception of SHAPE reagents) are boxed in orange, and SHAPE probes are boxed in green; the center of the box corresponds to the year of publication. Lower part, the blue arrows indicate the implementation of a new method to detect sites of cleavage or modification. CE: capillary electrophoresis; Seq: deep-sequencing; MaP: mutational profiling. Methods combining probing and deep-sequencing are highlighted in yellow. Methods based on mutational profiling are highlighted in orange, and techniques using enrichment by selection via the probe are labeled in purple. PARS: parallel analysis of RNA structure; FragSeq: fragmentation sequencing; CIRS-Seq: DMS and CMCT probing and sequencing; HRF-Seq: hydroxyl radical cleavage and sequencing; Mod-Seq: DMS probing and sequencing; Structure-Seq: DMS probing and sequencing; SHAPES: probing with NPIA and selection; icSHAPE: probing with NAI-N₃ and selection; Keth-seq: N₃-kethoxal probing, selection, and sequencing; LASER/-seq/-MaP: light-activated structural examination of RNA/analyzed by sequencing/mutational profiling.

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

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

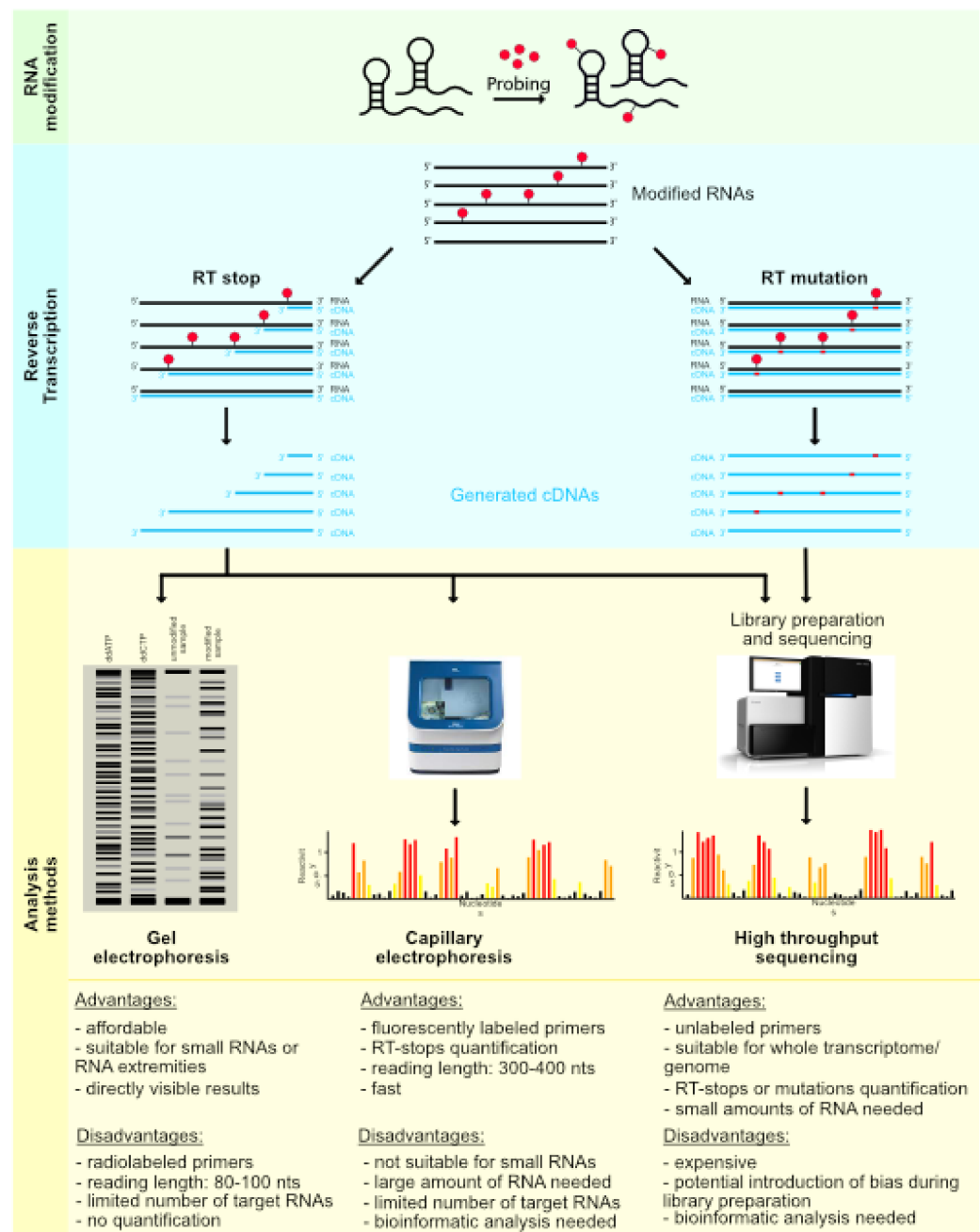


Figure 3. Schematic representation of RNA probing methods based on chemical modification. Folded RNAs were modified with the chemical probe of interest (green panel). For the RT stop strategy (left, blue panel), modified RNAs were reverse-transcribed with (1) radiolabeled primers for gel electrophoresis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

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

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

Table 2. SHAPE and post-SHAPE chemical probes.

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

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

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

4. A New Era for Enzymatic and Chemical Probing

4.1. The Deep-Sequencing Leaps

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

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

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

4.2. Old Probes Back in the Spotlight

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

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

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

4.3. Mutational Profiling

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

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

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

4.4. The Latest Developments

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

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

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

4.5. Applications of the MAP Strategies to Viral RNAs

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

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

5. Conclusions

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

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

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

6. Perspectives

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

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

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

Author Contributions: Conceptualization, R.M.; writing—original draft preparation, O.G., E.Q., A.-C.J. and V.V.-B., writing—review and editing, O.G., A.-C.J., J.-C.P., R.M. and V.V.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Agence Nationale de Recherche sur le SIDA et les hépatites virales (ANRS), SIDACTION, and the CNRS. O.G. and E.Q. were supported by fellowships from the French ministry of Research and Higher Education, as well as from SIDACTION (O.G.) and University of Strasbourg (E.Q.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

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

References

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

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

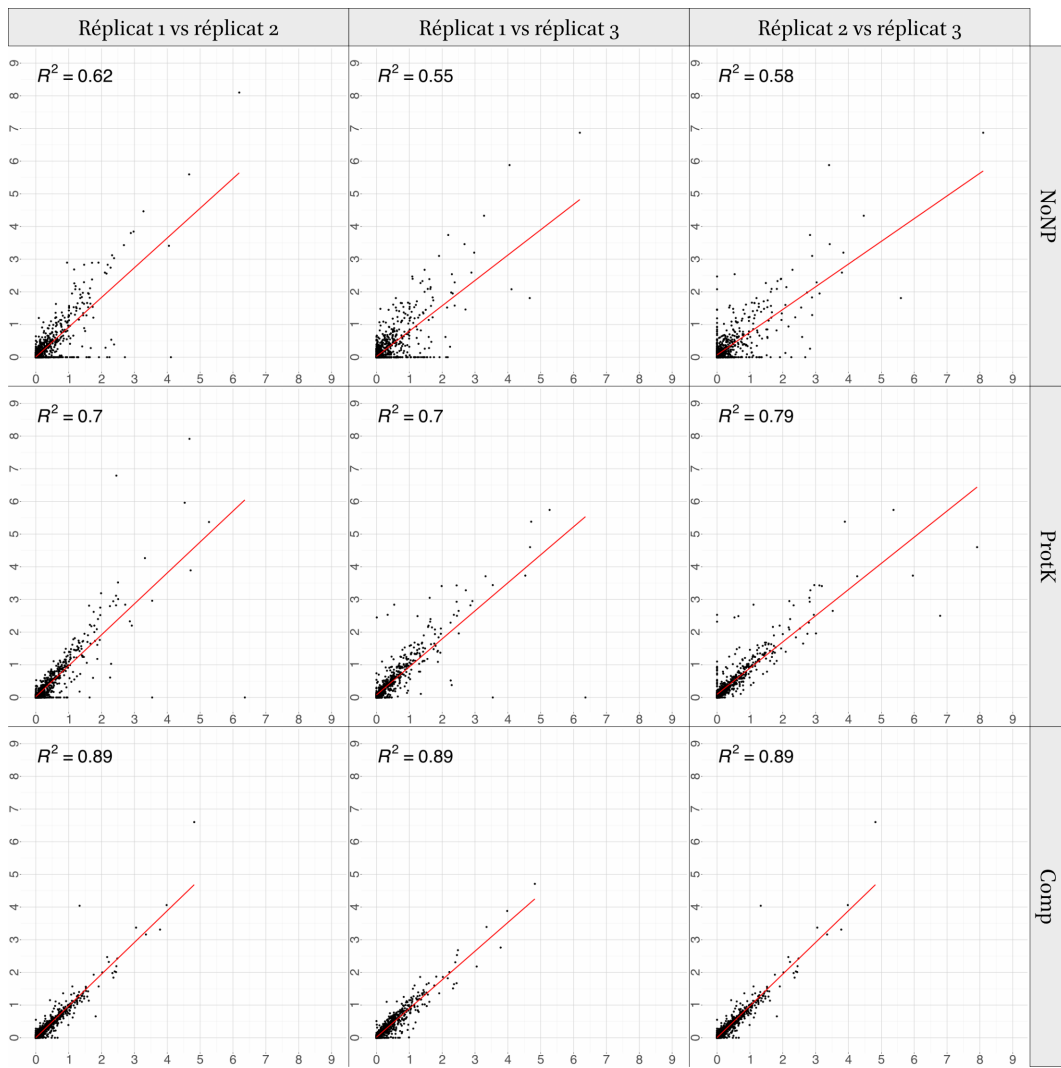
51. Wang, X.D.; Padgett, R.A. Hydroxyl Radical “Footprinting” of RNA: Application to Pre-mRNA Splicing Complexes. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 7795–7799. [[CrossRef](#)]
52. Vary, C.P.; Vournakis, J.N. RNA Structure Analysis Using Methidiumpropyl-EDTA.Fe(II): A Base-Pair-Specific RNA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

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

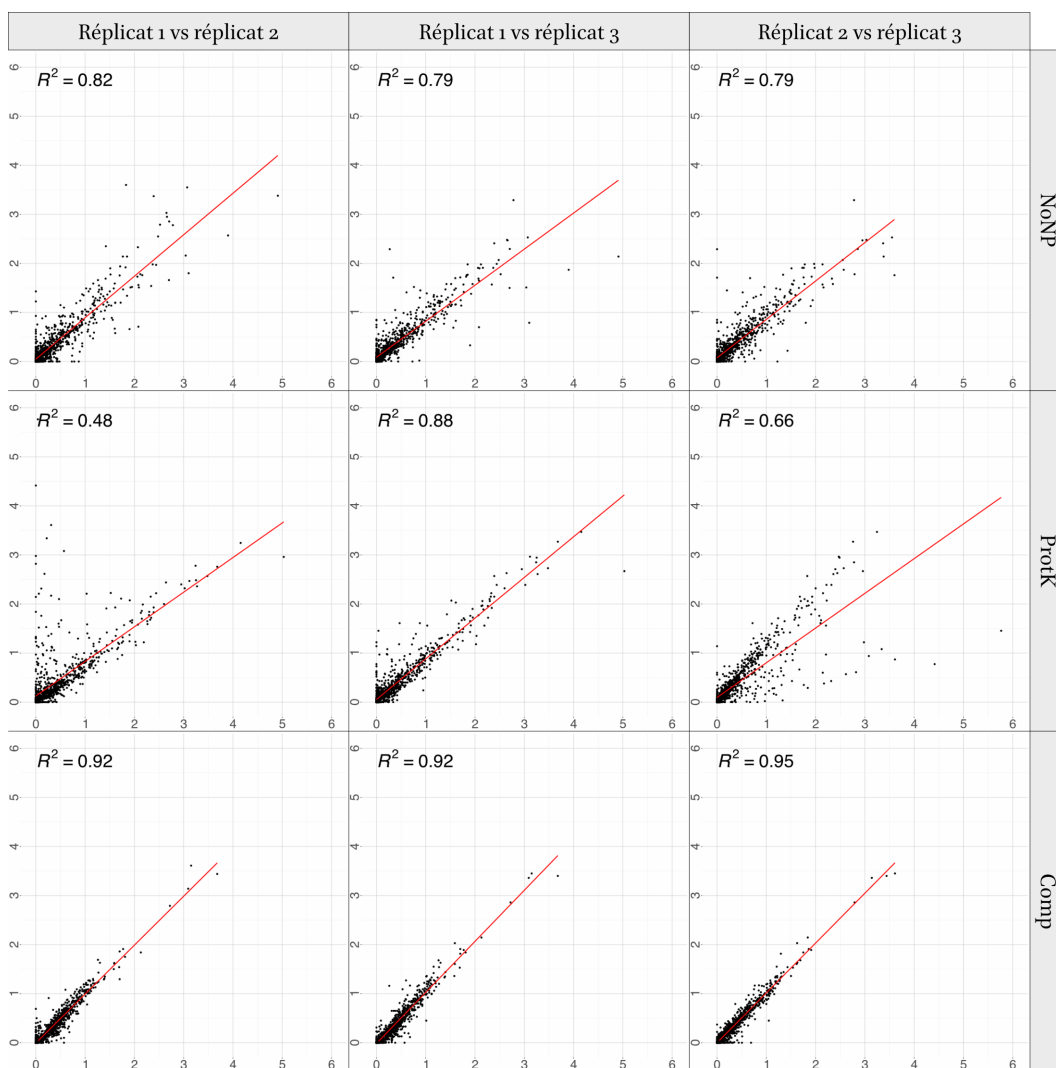
107. Wilkinson, K.A.; Merino, E.J.; Weeks, K.M. Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE): Quantitative RNA Structure Analysis at Single Nucleotide Resolution. *Nat. Protoc.* **2006**, *1*, 1610–1616. [[CrossRef](#)]
108. Wilkinson, K.A.; Merino, E.J.; Weeks, K.M. RNA SHAPE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

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

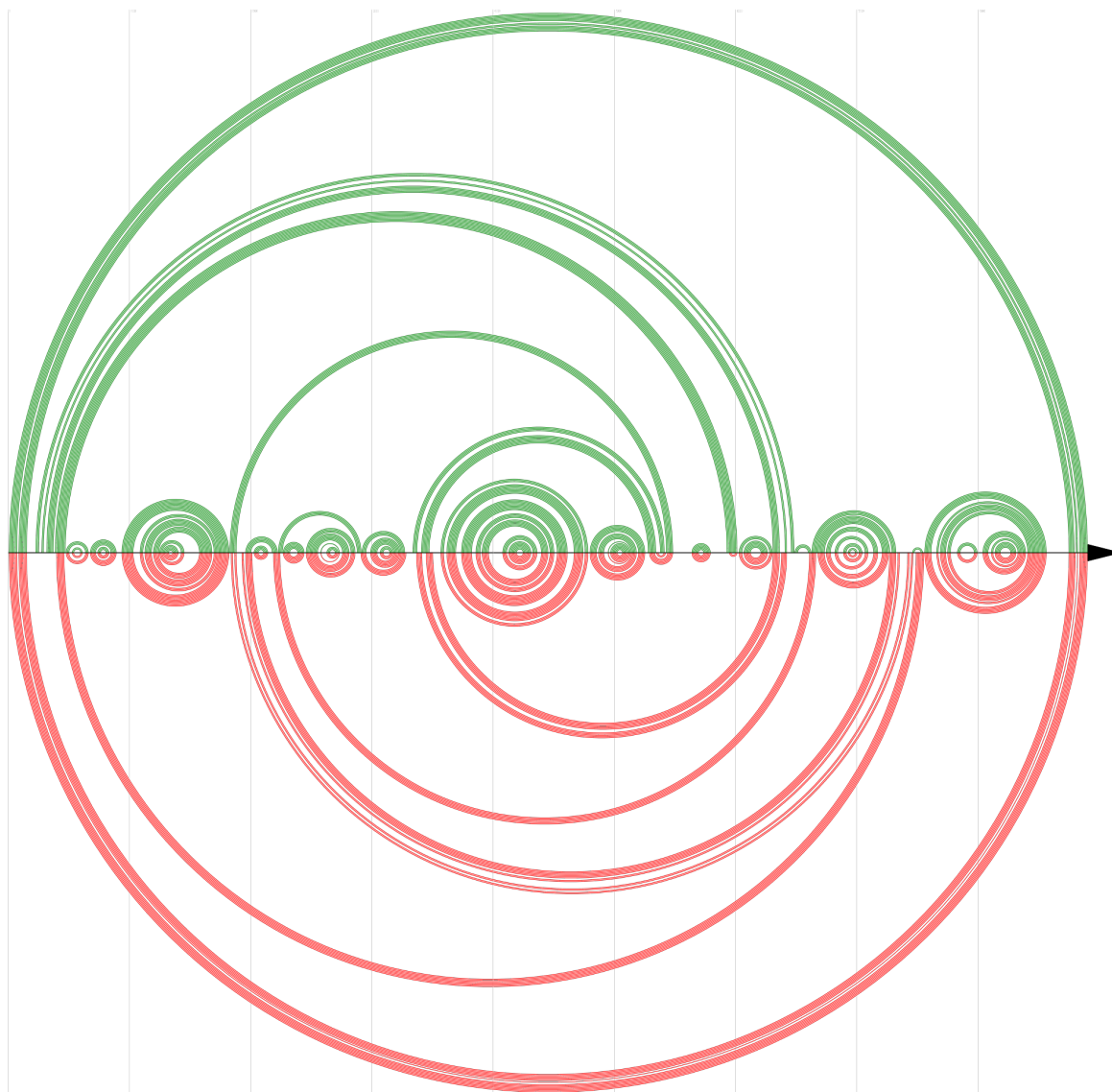
163. Homan, P.J.; Favorov, O.V.; Lavender, C.A.; Kursun, O.; Ge, X.; Busan, S.; Dokholyan, N.V.; Weeks, K.M. Single-Molecule Correlated Chemical Probing of RNA. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13858–13863. [[CrossRef](#)] [[PubMed](#)]
164. Krokhotin, A.; Mustoe, A.M.; Weeks, K.M.; Dokholyan,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



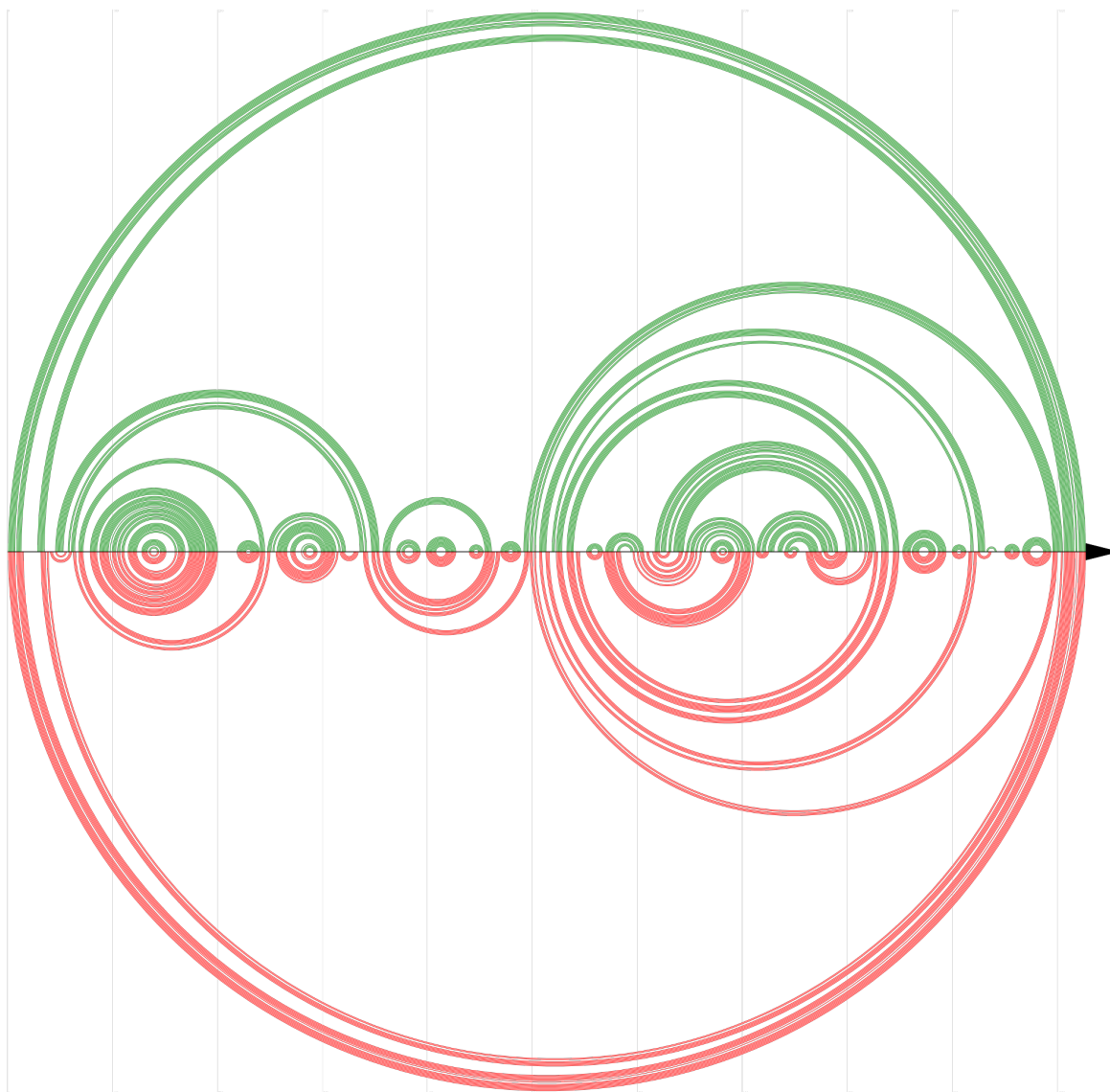
ANNEXE 1. – Représentation des corrélations entre les réplikat NS dans les différentes conditions, calculées avec R.



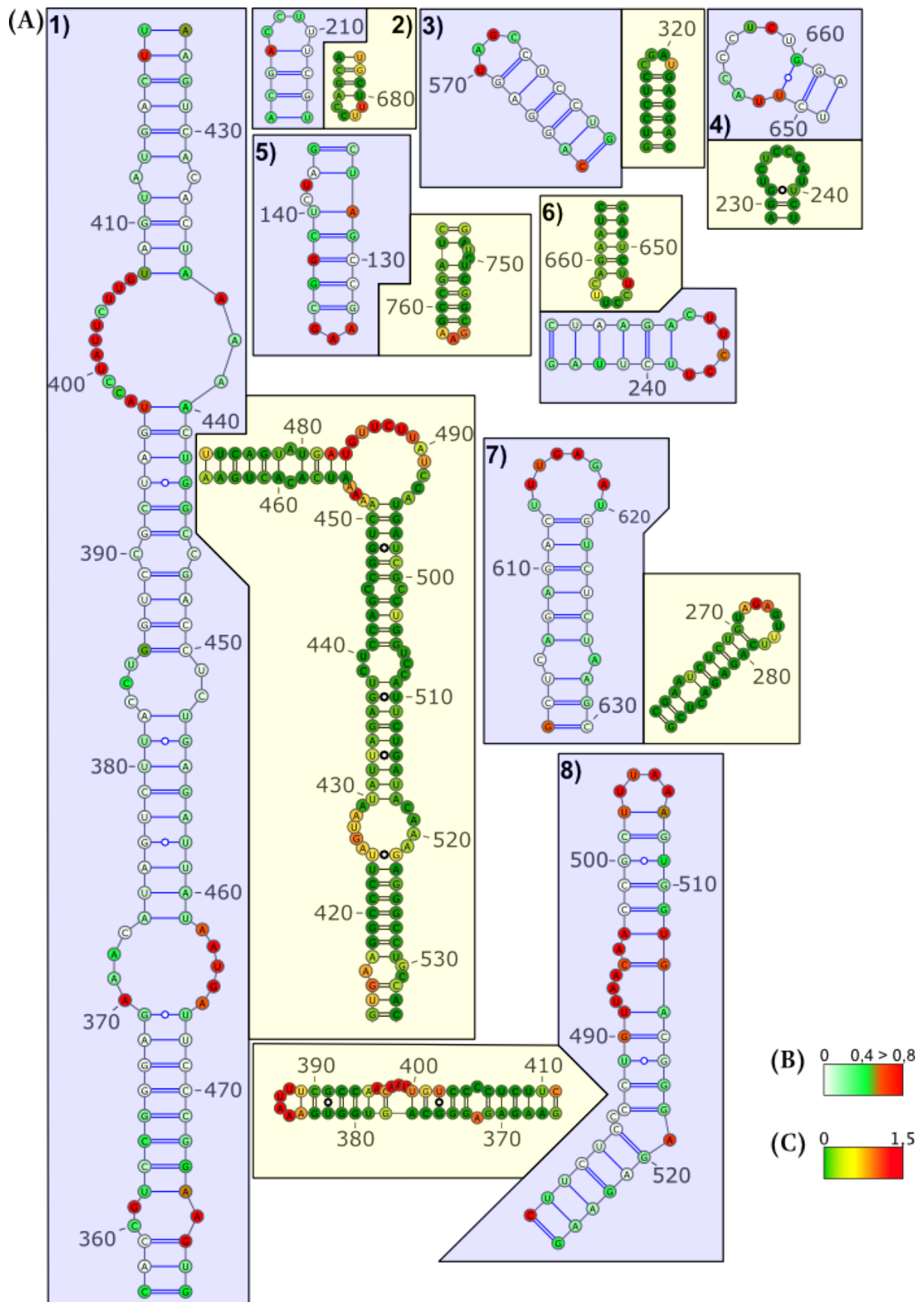
ANNEXE 2. – Représentation des corrélations entre les réplicats M dans les différentes conditions, calculées avec R.

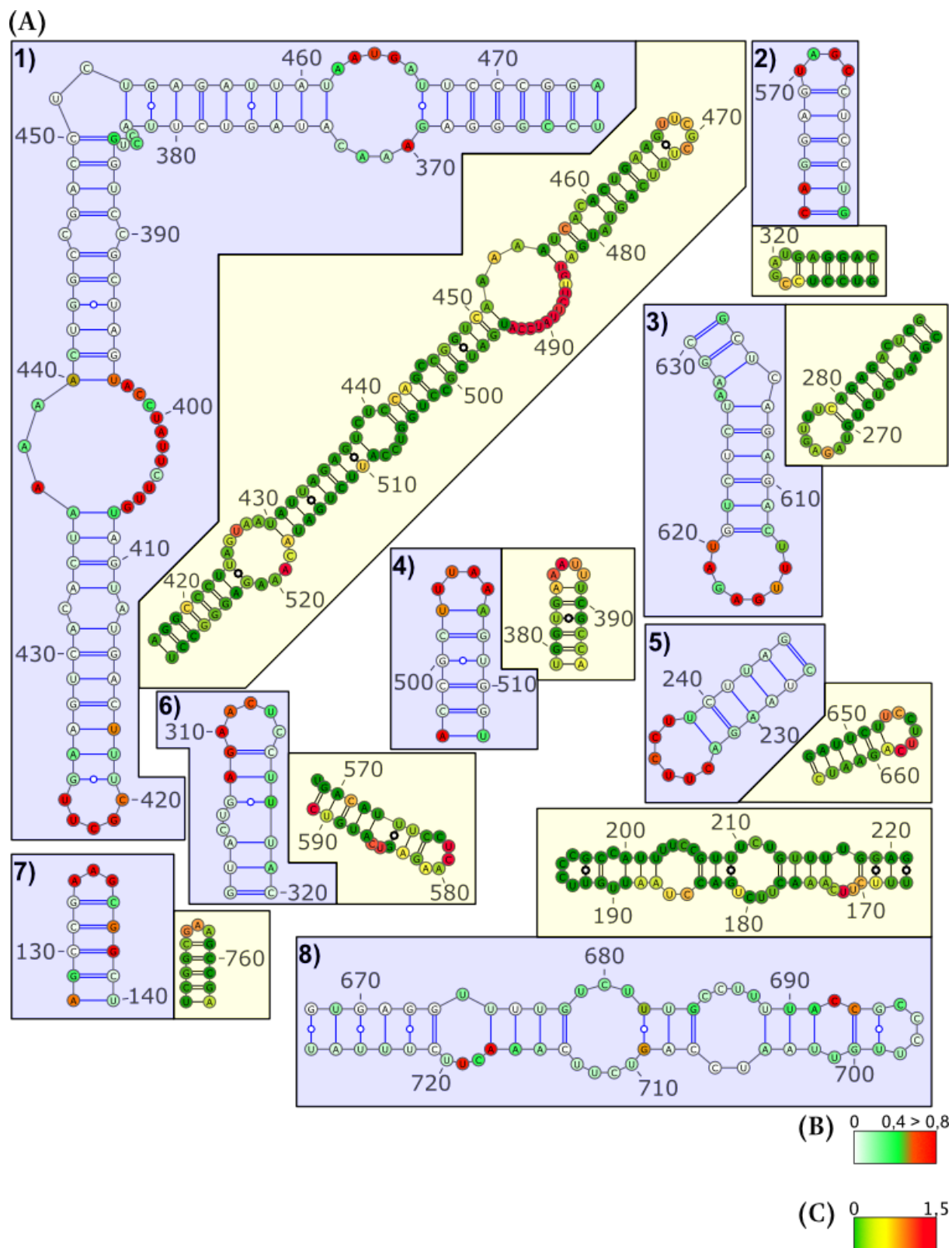


ANNEXE 3. – Représentation en arcs des structures des conditions NoNP et ProtK du segment NS. La condition NoNP est représentée en vert, la condition ProtK en rouge. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

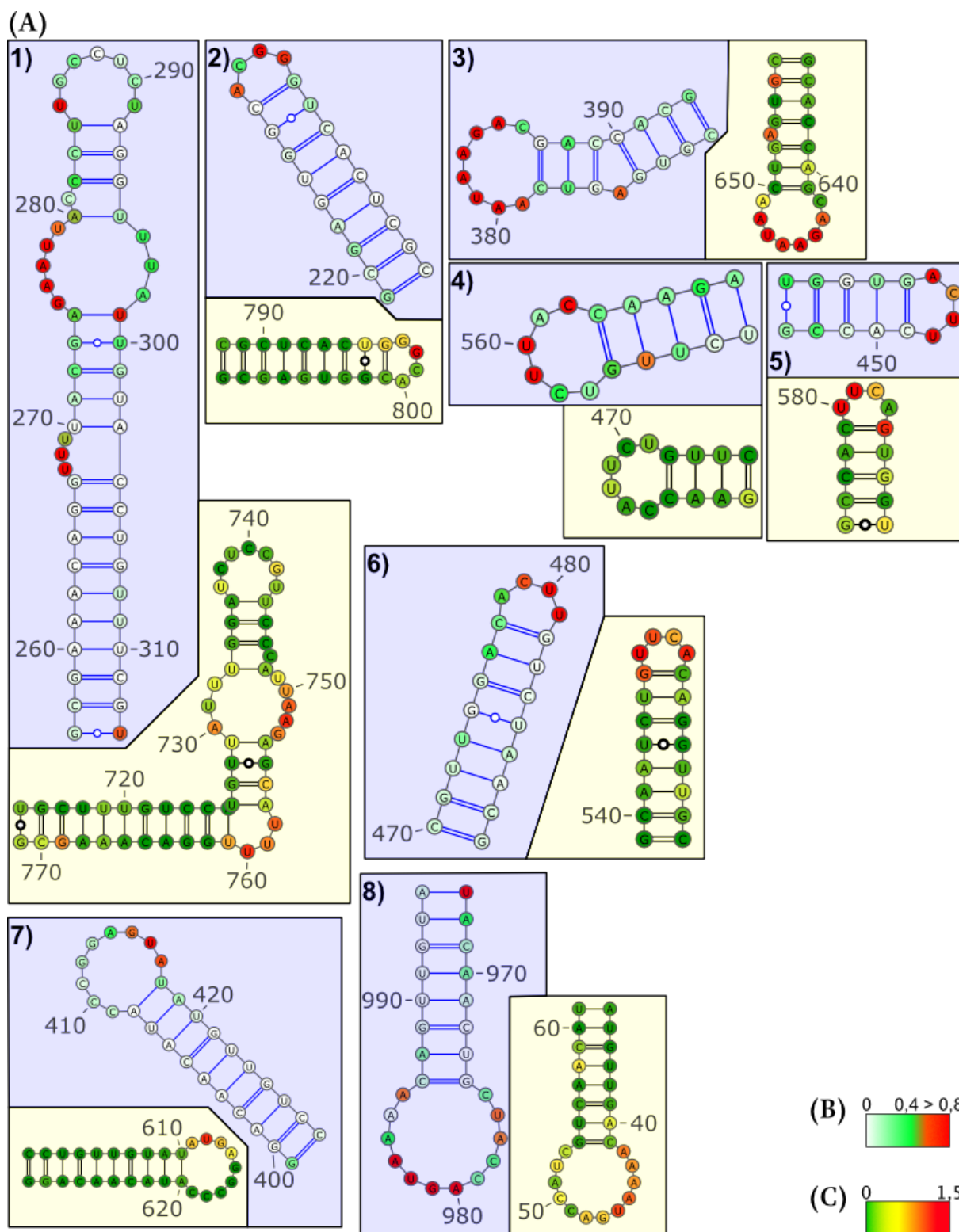


ANNEXE 4. – Représentation en arcs des structures des conditions NoNP et ProtK du segment M. La condition NoNP est représentée en vert, la condition ProtK en rouge. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

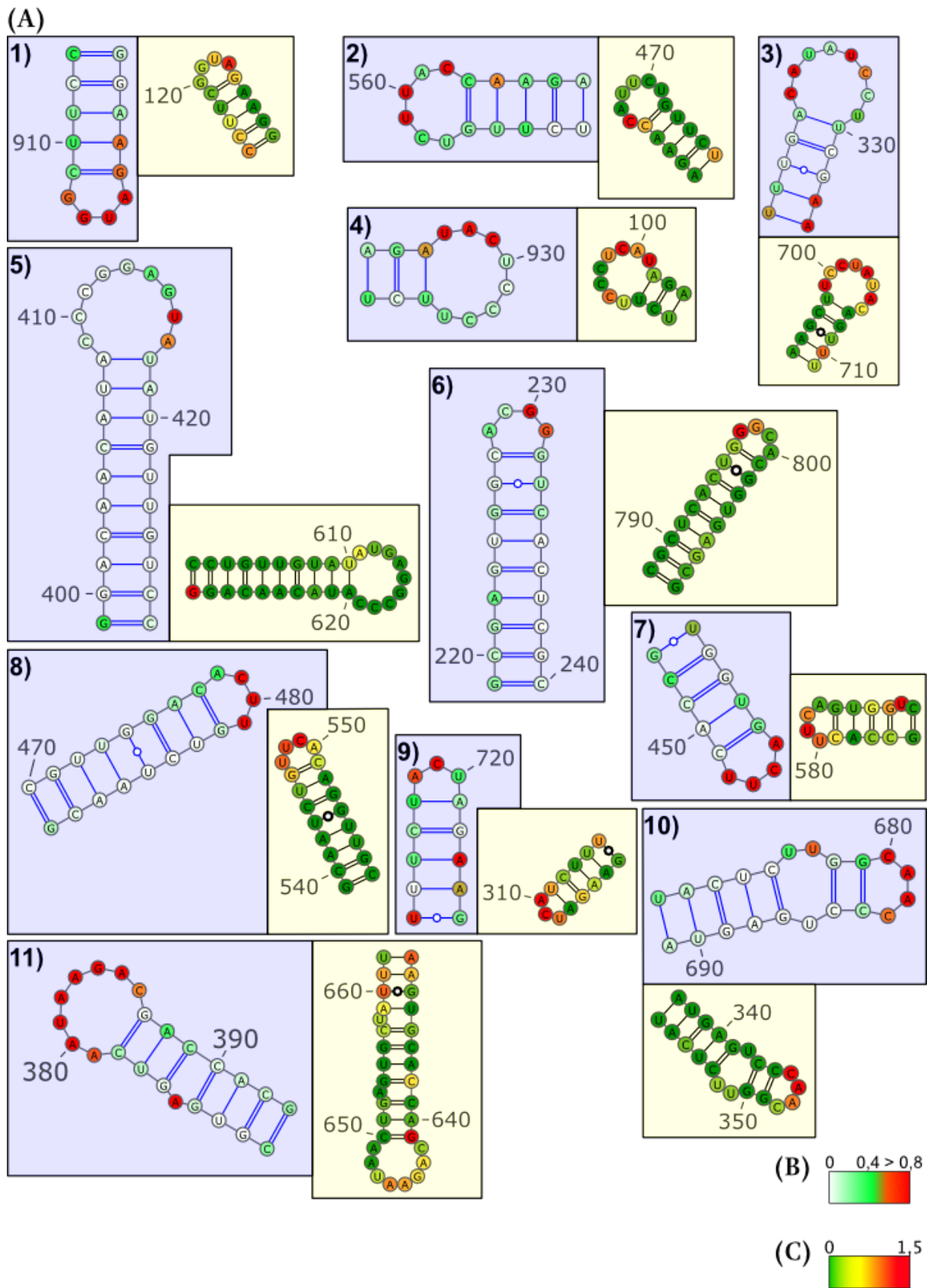




ANNEXE 6. – Comparaison des structures NS ProtK avec celles de D. Ferhadian. Les données obtenues au cours de cette thèse sont représentées dans les cadres bleus, celles obtenues par D. Ferhadian sont représentées dans les cadres jaunes.



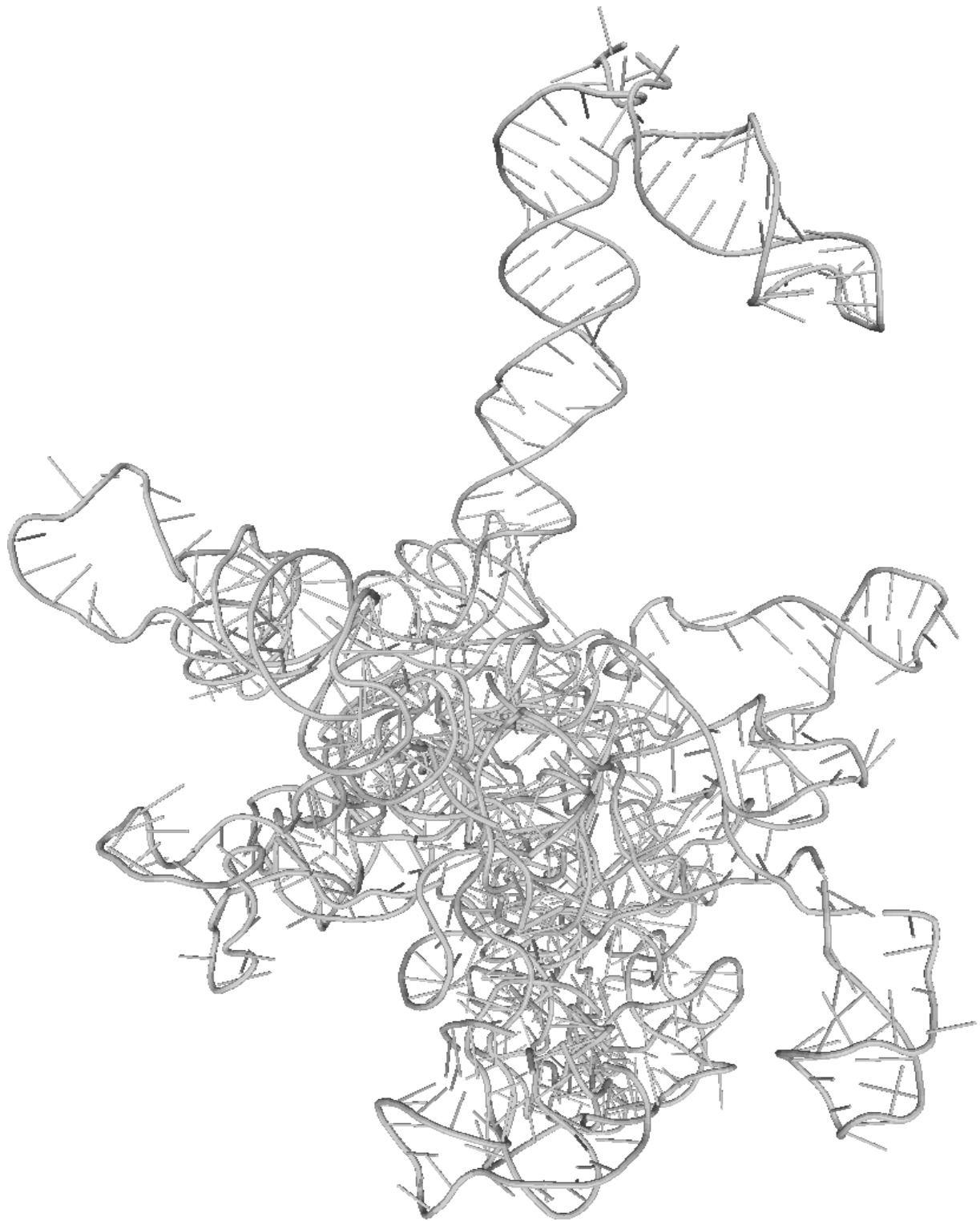
ANNEXE 7. – Comparaison des structures MNoNP avec celles de D. Ferhadian. Les données obtenues au cours de cette thèse sont représentées dans les cadres bleus, celles obtenues par D. Ferhadian sont représentées dans les cadres jaunes.



ANNEXE 8. – Comparaison des structures M ProtK avec celles de D. Ferhadian. Les données obtenues au cours de cette thèse sont représentées dans les cadres bleus, celles obtenues par D. Ferhadian sont représentées dans les cadres jaunes.



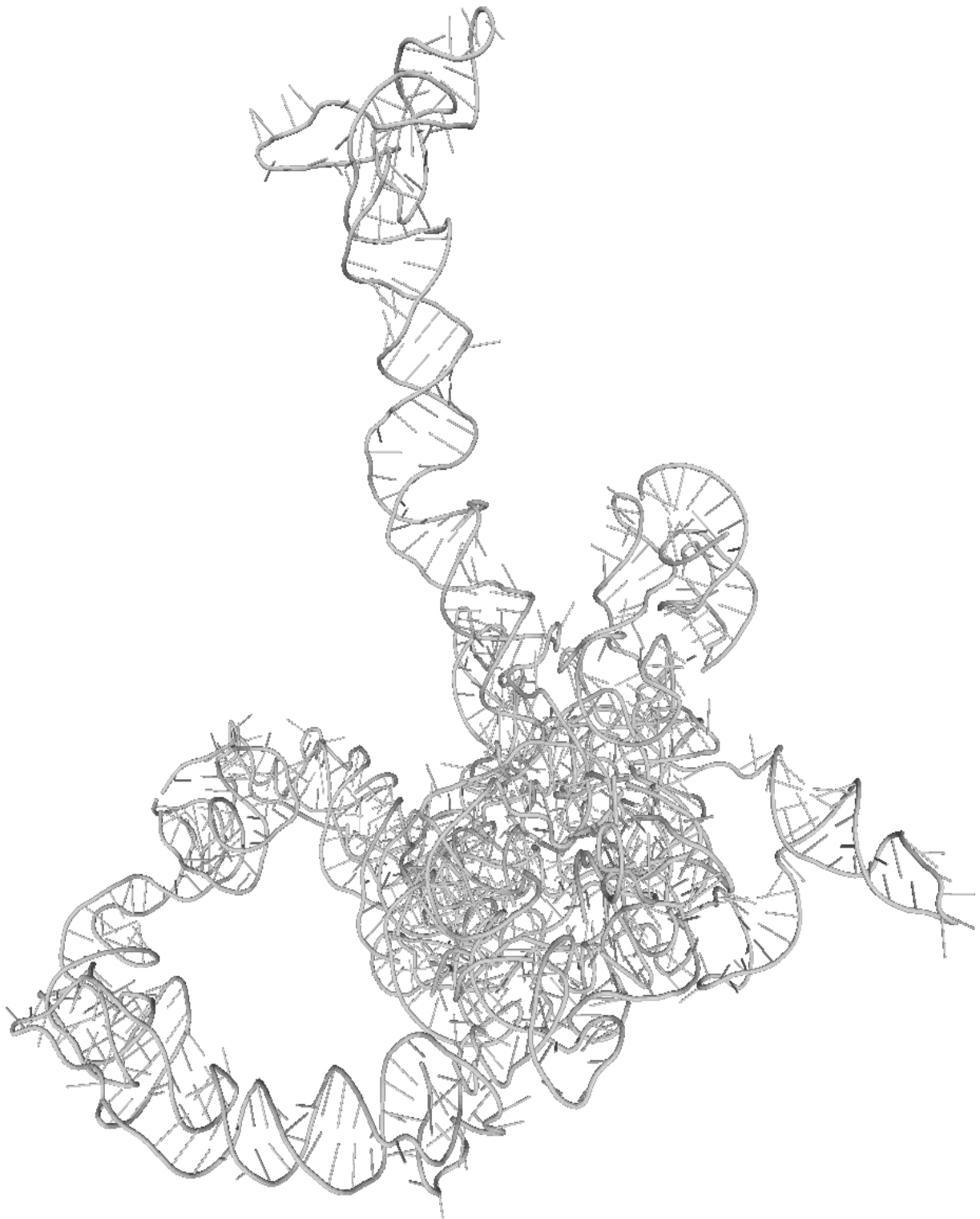
ANNEXE 9. – Prédiction 3D de la structure de l'ARN NS dans la condition « NoNP » à l'aide de 3dRNA. Cette représentation est interactive sous une version récente de Adobe Acrobat DC.



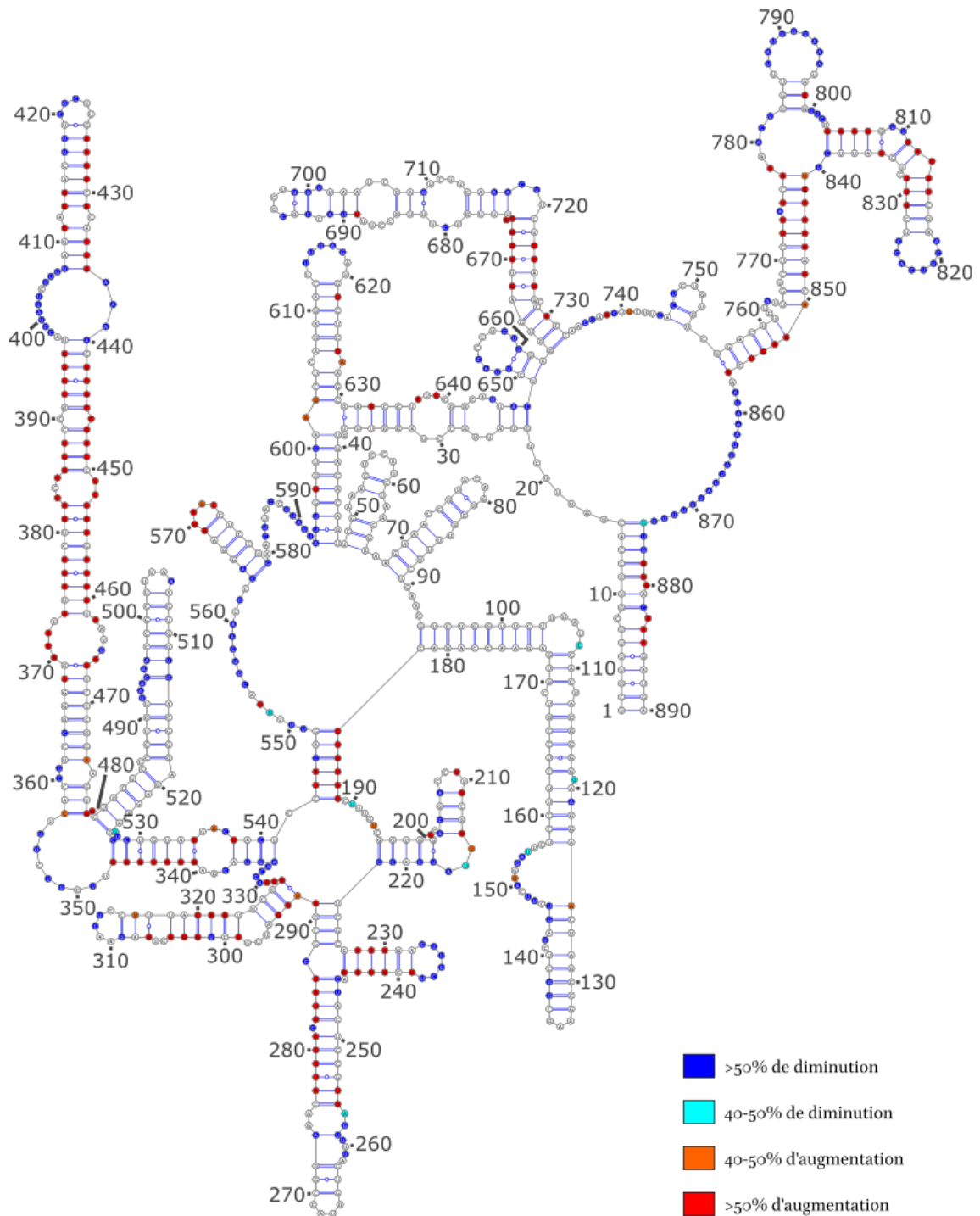
ANNEXE 10. – Prédiction 3D de la structure de l'ARN NS dans la condition « ProtK » à l'aide de 3dRNA. Cette représentation est interactive sous une version récente de Adobe Acrobat DC.



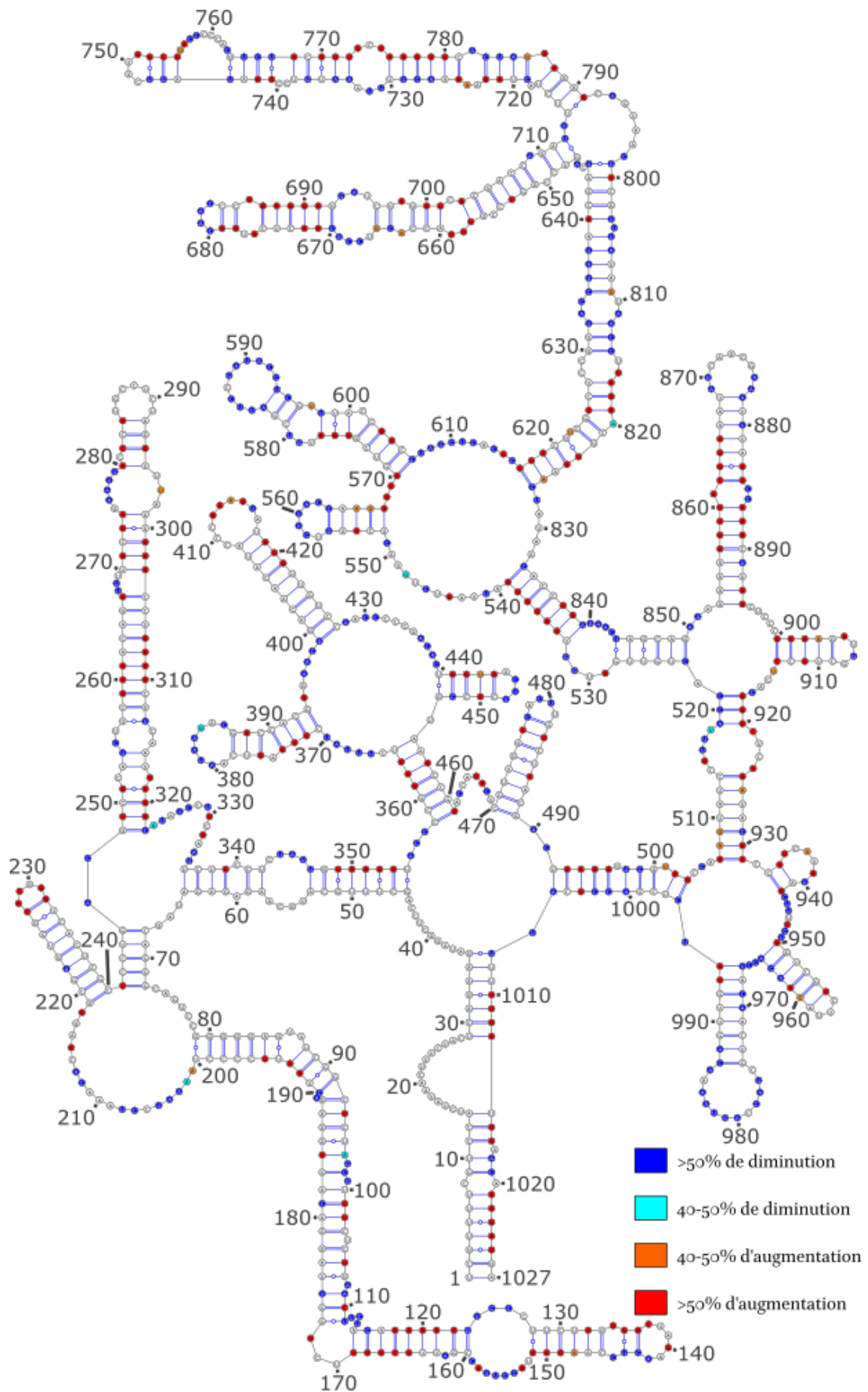
ANNEXE 11. – Prédiction 3D de la structure de l'ARN M dans la condition « NoNP » à l'aide de 3dRNA. Cette représentation est interactive sous une version récente de Adobe Acrobat DC.



ANNEXE 12. – Prédiction 3D de la structure de l'ARN M dans la condition « ProtK » à l'aide de 3dRNA. Cette représentation est interactive sous une version récente de Adobe Acrobat DC.



ANNEXE 13. – Représentation des changements de réactivités de l'ARN NS entre les conditions « ProtK » et « Comp » sur la structure de l'ARN prédite avec les données de la condition « NoNP ». La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).



ANNEXE 14. – Représentation des changements de réactivités de l'ARN M entre les conditions « ProtK » et « Comp » sur la structure de l'ARN prédite avec les données de la condition « NoNP ». La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

TABLEAU A1. – Données calculées par deltaSHAPE pour le segment NS pour la comparaison Δ NoNP-ProtK.

Nuc	Seq	DeltaSHAPE	Z-factor	Std_Score	Smoothed_Data1	Smoothed_Data2	Unsmoothed_Diff	Data1	Data2
564	C	-2.49	0.11	-2.41	-1.70	0.79	-1.24	0.02	1.26
565	A	-3.10	0.21	-3.13	-2.62	0.48	-4.83	-4.01	0.83
566	G	-3.25	0.28	-3.30	-3.41	-0.16	-3.21	-3.87	-0.65
567	G	-1.82	0.17	-1.63	-2.39	-0.57	-1.70	-2.35	-0.65
594	U	-3.68	0.11	-3.81	-2.91	0.77	-4.80	-3.49	1.31
595	U	-4.97	0.26	-5.32	-4.89	0.09	-5.77	-5.79	-0.02
596	G	-3.69	0.19	-3.82	-4.09	-0.40	-4.34	-5.37	-1.03
597	U	-1.89	0.18	-1.72	-2.34	-0.44	-0.95	-1.11	-0.16
653	A	-2.18	0.00	-2.05	-1.45	0.72	-0.38	0.07	0.45
654	C	-3.38	0.25	-3.46	-3.32	0.06	-5.96	-5.12	0.84
655	C	-3.97	0.34	-4.15	-4.16	-0.19	-3.81	-4.93	-1.12
656	C	-2.15	0.28	-2.02	-2.45	-0.30	-2.14	-2.44	-0.30
658	C	-2.29	0.05	-2.18	-1.50	0.79	-1.06	0.28	1.34
659	U	-4.38	0.23	-4.63	-2.40	1.98	-5.31	-4.78	0.53
660	G	-5.23	0.34	-5.62	-4.06	1.17	-6.77	-2.69	4.08
661	G	-4.43	0.29	-4.68	-3.95	0.48	-3.60	-4.70	-1.11
662	A	-2.59	0.27	-2.53	-3.73	-1.15	-2.91	-4.45	-1.54
700	G	-1.35	0.03	-1.08	-1.42	-0.07	-1.72	-2.02	-0.30
701	U	-1.82	0.19	-1.64	-1.98	-0.16	-2.15	-2.22	-0.07
702	U	-1.92	0.26	-1.74	-2.14	-0.23	-1.60	-1.70	-0.10
703	A	-1.57	0.27	-1.33	-1.78	-0.21	-2.00	-2.52	-0.51
704	A	-1.30	0.28	-1.02	-1.56	-0.26	-1.10	-1.13	-0.03
705	U	-2.19	0.13	-2.07	-2.87	-0.68	-0.80	-1.04	-0.23
706	C	-2.39	0.14	-2.29	-3.29	-0.90	-4.68	-6.46	-1.77
707	C	-2.14	0.05	-2.00	-2.96	-0.82	-1.67	-2.38	-0.71

TABLEAU A2. – Données calculées par deltaSHAPE pour le segment M pour la comparaison Δ NoNP-ProtK.

Nuc	Seq	DeltaSHAPE	Z-factor	Std_Score	Smoothed_Data1	Smoothed_Data2	Unsmoothed_Diff	Data1	Data2
151	G	0.24	0.46	1.09	0.44	0.20	0.23	0.32	0.09
152	U	0.35	0.58	1.52	0.55	0.20	0.28	0.66	0.38
153	A	0.26	0.04	1.15	0.67	0.41	0.53	0.66	0.14
165	U	0.26	0.62	1.17	0.45	0.19	0.20	0.18	-0.02
166	C	0.36	0.46	1.59	0.43	0.07	0.20	0.15	-0.06
167	U	0.34	0.30	1.47	0.53	0.19	0.69	0.97	0.28
168	G	0.24	0.01	1.08	0.56	0.32	0.11	0.46	0.35
203	A	-0.69	0.38	-2.79	1.46	2.15	-0.43	1.15	1.57
204	A	-0.68	0.40	-2.73	1.26	1.94	-1.40	1.95	3.35
205	U	-0.65	0.42	-2.61	0.81	1.46	-0.20	0.69	0.90
229	C	0.93	0.43	3.94	1.48	0.55	0.20	0.31	0.11
230	G	0.83	0.34	3.52	1.51	0.68	2.21	3.50	1.29
231	G	0.77	0.30	3.28	1.45	0.68	0.07	0.72	0.65
281	C	0.30	0.08	1.31	0.30	0.00	0.06	0.04	-0.02
282	C	0.28	0.12	1.26	0.21	-0.08	0.56	0.41	-0.15
283	C	0.41	0.40	1.77	0.34	-0.07	0.23	0.17	-0.06
284	U	0.35	0.54	1.53	0.58	0.23	0.43	0.42	0.00
465	A	0.24	0.68	1.06	0.49	0.25	0.33	0.80	0.47
466	U	0.32	0.76	1.39	0.65	0.34	0.31	0.50	0.19
467	A	0.25	0.24	1.10	0.52	0.27	0.30	0.66	0.36
489	G	-0.26	0.44	-1.00	0.47	0.73	-0.10	0.04	0.14
490	A	-0.37	0.45	-1.45	0.79	1.16	-0.70	1.36	2.06
491	C	-0.42	0.51	-1.68	0.98	1.41	-0.30	0.98	1.28
546	C	0.22	0.52	1.00	0.56	0.34	0.29	0.59	0.30
547	U	0.34	0.52	1.51	1.07	0.72	0.27	0.62	0.35
548	G	0.42	0.49	1.84	1.27	0.85	0.47	1.98	1.52
549	U	0.47	0.55	2.04	1.40	0.93	0.53	1.20	0.67
582	G	0.23	0.42	1.02	0.42	0.19	0.04	0.05	0.02
583	A	0.31	0.32	1.36	0.69	0.38	0.58	1.06	0.48
584	U	0.32	0.02	1.40	1.03	0.71	0.31	0.97	0.66
588	G	-0.61	0.66	-2.44	0.49	1.09	-0.45	0.48	0.93
589	A	-0.90	0.68	-3.66	0.75	1.65	-1.40	0.83	2.23
590	U	-0.95	0.63	-3.86	0.79	1.73	-0.85	0.96	1.80
591	A	-0.66	0.16	-2.66	0.85	1.51	-0.59	0.58	1.17
597	U	0.41	0.70	1.79	0.76	0.35	0.35	0.71	0.37
598	U	0.59	0.65	2.53	0.92	0.33	0.77	1.36	0.58
599	U	0.54	0.64	2.32	0.74	0.20	0.65	0.68	0.03
600	A	0.47	0.36	2.03	0.51	0.05	0.20	0.19	-0.01
601	C	0.46	0.39	1.99	0.45	-0.01	0.56	0.68	0.12
602	C	0.37	0.24	1.62	0.36	-0.01	0.62	0.48	-0.14
644	U	0.28	0.55	1.22	0.45	0.17	0.61	1.01	0.40
645	C	0.41	0.60	1.77	0.67	0.26	0.26	0.32	0.06
646	A	0.30	0.44	1.34	0.51	0.20	0.35	0.67	0.32
647	G	0.22	0.29	1.01	0.56	0.33	0.30	0.52	0.22
668	C	0.43	0.24	1.87	1.19	0.76	0.17	0.66	0.48
669	G	0.45	0.27	1.97	1.31	0.85	1.11	2.70	1.60

TABLEAU A2. – Données calculées par deltaSHAPE pour le segment M pour la comparaison Δ NoNP-ProtK.

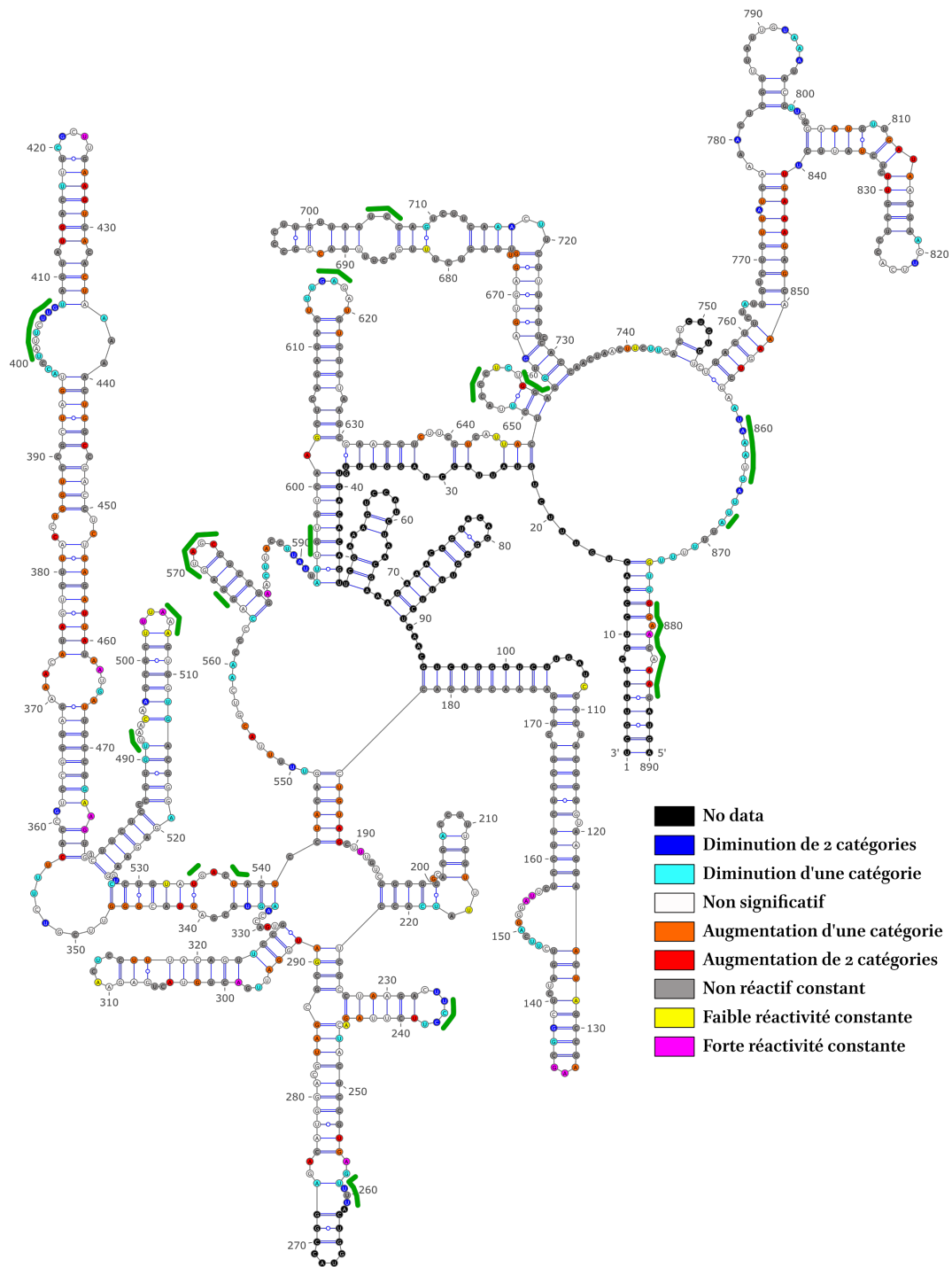
Nuc	Seq	DeltaSHAPE	Z-factor	Std_Score	Smoothed_Data1	Smoothed_Data2	Unsmoothed_Diff	Data1	Data2
670	C	0.35	0.05	1.53	1.12	0.77	0.08	0.56	0.48
708	C	-0.69	0.58	-2.80	0.49	1.18	-0.08	-0.01	0.07
709	A	-0.73	0.59	-2.94	0.61	1.33	-2.00	1.51	3.51
710	G	-0.61	0.49	-2.45	0.82	1.42	-0.10	0.32	0.43
711	A	0.69	0.76	2.94	1.25	0.56	0.28	0.61	0.34
712	U	0.95	0.81	4.05	1.62	0.66	1.90	2.82	0.93
713	U	0.98	0.81	4.17	1.49	0.51	0.69	1.42	0.73
714	U	0.33	0.54	1.43	0.62	0.29	0.36	0.24	-0.12
722	G	0.38	0.59	1.66	0.73	0.35	0.10	0.12	0.02
723	A	0.59	0.71	2.51	1.05	0.46	1.02	1.91	0.89
724	A	0.54	0.67	2.31	1.07	0.54	0.64	1.12	0.48
726	A	-0.37	0.60	-1.46	0.23	0.60	-0.38	0.25	0.63
743	A	-0.27	0.50	-1.03	0.21	0.48	-0.08	0.27	0.36
744	U	-0.27	0.50	-1.03	0.20	0.47	-0.61	0.36	0.97
745	A	-0.39	0.41	-1.55	0.29	0.69	-0.11	-0.02	0.08
748	C	0.37	0.43	1.62	0.88	0.51	0.10	0.26	0.16
749	U	0.42	0.54	1.83	0.82	0.40	0.50	0.95	0.46
750	U	0.43	0.72	1.87	0.78	0.35	0.67	1.25	0.57
751	U	0.45	0.74	1.94	0.75	0.31	0.12	0.13	0.01
782	A	-0.40	0.64	-1.57	0.28	0.68	-0.49	0.42	0.91
783	C	-0.61	0.74	-2.45	0.38	0.99	-0.63	0.47	1.11
784	U	-1.09	0.80	-4.45	0.46	1.55	-0.70	0.24	0.94
785	A	-0.98	0.81	-4.01	0.30	1.29	-1.93	0.68	2.61
786	G	-0.77	0.78	-3.11	0.22	0.99	-0.32	-0.01	0.31
792	C	0.46	0.62	1.97	0.68	0.23	0.60	0.81	0.21
793	A	0.71	0.75	3.04	1.07	0.36	0.77	1.21	0.44
794	G	0.76	0.73	3.23	1.22	0.46	0.76	1.18	0.42
795	U	0.66	0.72	2.84	1.07	0.41	0.74	1.27	0.53
796	A	0.67	0.74	2.87	1.14	0.47	0.49	0.76	0.27
797	A	0.45	0.66	1.94	0.86	0.42	0.78	1.40	0.61
816	U	0.25	0.12	1.11	0.42	0.17	0.22	0.42	0.20
819	A	0.74	0.27	3.17	1.28	0.54	0.46	0.84	0.38
820	A	1.18	0.13	4.98	1.12	-0.06	1.21	2.09	0.88
821	C	1.30	0.09	5.50	0.71	-0.60	1.86	0.42	-1.44
850	A	0.37	0.32	1.64	1.38	1.01	0.27	0.95	0.68
851	A	0.42	0.34	1.81	1.29	0.87	0.71	2.62	1.91
852	A	0.32	0.22	1.40	0.96	0.65	0.26	0.30	0.04
950	U	0.39	0.15	1.70	0.19	-0.20	0.48	0.14	-0.34
951	G	0.48	0.22	2.06	-0.31	-0.79	0.60	-0.41	-1.01
952	U	0.48	0.06	2.07	-0.58	-1.06	0.35	-0.67	-1.02
980	A	0.28	0.45	1.24	0.83	0.55	0.37	0.84	0.48
981	G	0.35	0.32	1.52	1.21	0.87	0.41	1.43	1.02
982	U	0.28	0.10	1.24	1.30	1.02	0.26	1.37	1.11
1017	A	-0.43	0.25	-1.72	0.60	1.03	-0.05	1.39	1.44
1018	A	-0.39	0.08	-1.54	0.70	1.09	-0.91	0.59	1.50

TABLEAU A2. – Données calculées par deltaSHAPE pour le segment M pour la comparaison Δ NoNP-ProtK.

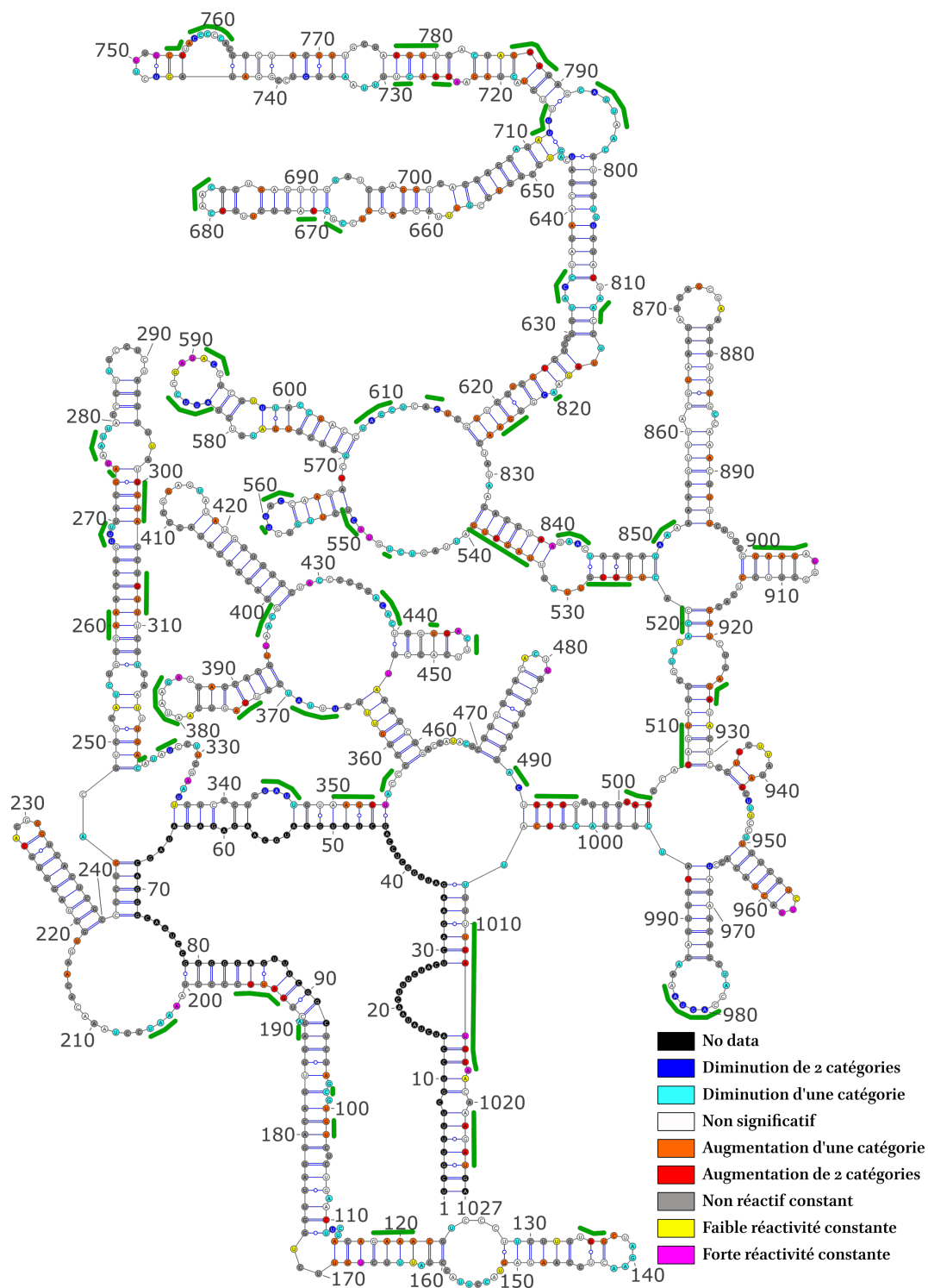
Nuc	Seq	DeltaSHAPE	Z-factor	Std_Score	Smoothed_Data1	Smoothed_Data2	Unsmoothed_Diff	Data1	Data2
1019	C	-0.39	0.16	-1.56	0.32	0.71	-0.21	0.14	0.35

	NoNP				ProtK				Comp			
	NS		M		NS		M		NS		M	
	EQ	BD	EQ	BD	EQ	BD	EQ	BD	EQ	BD	EQ	BD
Minimum	-10	-0,4	-0,9	-2,6	-1,8	-0,3	-1,8	-2,2	-7,8	-0,4	-2,1	-1
Maximum	7,3	7,2	3,5	14,3	5,9	7,6	3,5	12,3	11,6	7,6	3,5	6,9
Médiane	0,05	0,4	0,2	0,4	0,1	0,4	0,2	0,4	0,2	0,3	0,2	0,4
Moyenne	-0,07	0,5	0,3	0,5	0,4	0,5	0,4	0,5	0,3	0,5	0,3	0,5

TABLEAU A3. – Statistiques sur les données « NoNP », « ProtK » et « Comp » obtenues au cours de cette thèse et leurs équivalences au sein des données de B. Dadonaite.



ANNEXE 15. – Représentation des changements de catégories et des régions identifiées à l'aide de deltaSHAPE pour le segment NS entre les conditions « NoNP » et « Comp ». La structure secondaire a été prédite à l'aide des données « NoNP » et les catégories ont été définies selon les valeurs de réactivité : 0 : non réactif; 1 : moyennement réactif (entre 0,4 et 0,8); 2 : fortement réactif (supérieur à 0,8). Les changements de catégories ont été calculés pour les positions pour lesquels un changement de la réactivité d'au moins 40% a été observé. Un changement de catégorie sans un changement suffisant de la réactivité est indiqué en blanc. Les nucléotides pour lesquels les catégories ne changent pas entre les deux conditions sont également indiqués. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).



ANNEXE 16. – Représentation des changements de catégories et des régions identifiées à l'aide de deltaSHAPE pour le segment M entre les conditions « NoNP » et « Comp ». La structure secondaire a été prédite à l'aide des données « NoNP » et les catégories ont été définies selon les valeurs de réactivité : 0 : non réactif; 1 : moyennement réactif (entre 0,4 et 0,8); 2 : fortement réactif (supérieur à 0,8). Les changements de catégories ont été calculés pour les positions pour lesquels un changement de la réactivité d'au moins 40% a été observé. Un changement de catégorie sans un changement suffisant de la réactivité est indiqué en blanc. Les nucléotides pour lesquels les catégories ne changent pas entre les deux conditions sont également indiqués. La numérotation s'effectue du 3' au 5', comme le veut la convention dans le domaine des virus à ARN (-).

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 :

Erwan QUIGNON

soutenue le : 02 septembre 2022

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

Discipline/ Spécialité : Aspects moléculaires et cellulaires de la biologie

**Interaction des segments d'ARN génomique du
virus Influenza A avec la nucléoprotéine virale
NP : impact structural et implications pour
l'encapsidation du génome**

THÈSE dirigée par :

Dr Roland Marquet

IBMC, UPR9002, Strasbourg

RAPPORTEURS :

Dr Sandie Munier

Institut Pasteur, Paris

Dr Romain Volmer

ENVT, Toulouse

Examinatrice :

Pr Maria Dimitrova

IBMP, Strasbourg