

UNIVERSITE TOULOUSE III PAUL SABATIER
FACULTE DES SCIENCES PHARMACEUTIQUES

ANNEE : 2020

THESES 2020/TOU3/2074

THESE

POUR LE DIPLOME D'ETAT DE DOCTEUR EN PHARMACIE

Présentée et soutenue publiquement
par

DUPRE GABRIEL

**Etude de l'évolution des virus influenza aviaries de sous-type H5 vers les formes
hautement pathogènes**

Date de soutenance :

14 Décembre 2020

Directeur de thèse : Dr. Volmer Romain

JURY

Président : Prof. Pasquier, Christophe
1er assesseur : Dr. Chapuy-Regaud, Sabine
2ème assesseur : Prof. Verhaeghe, Pierre



RESUME en français

Les AIV constituent un problème majeur de santé publique et animale, du fait de leur potentiel zoonotique et des conséquences désastreuses qu'ils peuvent avoir au niveau de la filière avicole. Les HPAIV sont caractérisés par la présence d'un site de clivage polybasique de HA qui constitue un déterminant clé de leur pouvoir pathogène. Plusieurs mécanismes génétiques ont été proposés afin d'expliquer l'acquisition de cette séquence chez les souches LPAIV progénitrices de HPAIV de sous-type H5 ou H7. Cette prédisposition génétique restant encore énigmatique, plusieurs questions subsistent : tous les LPAIV de sous-type H5 et H7 ont-ils la même probabilité d'évolution vers un HPAIV ? Si une différence de probabilité existe, dans quelle mesure l'environnement nucléotidique entourant le site de clivage de la HA influence-t-il le chemin évolutif du LPAIV ?

Afin de répondre à ces questions, une analyse phylogénétique de séquences nucléotidiques HA de souches de sous-type H5, combinée à des prédictions de structures secondaires d'ARN a été réalisée. L'analyse bio-informatique nous a permis d'identifier plusieurs groupes de séquences HA de LPAIV H5, catégorisés comme plus ou moins favorables au phénomène d'émergence par insertions ? en fonction de certaines différences observées au niveau de la structure secondaire d'ARN englobant le site de clivage.

Ensuite, un système expérimental de type minigénome, s'appuyant sur une méthode précédemment publiée, a été élaboré dans le but de mesurer expérimentalement le phénomène d'insertion de nucléotide. Ce système, encore en cours d'optimisation devrait à terme permettre la détection d'événements d'insertion rares, possiblement déterminants dans l'évolution des virus influenza aviaires.

Titre et résumé en Anglais : voir au recto de la dernière page de la thèse

DISCIPLINE administrative : Sciences Pharmaceutiques

MOTS-CLES :

Virus influenza aviaires, sous-type H5 et H7, hémagglutinine, émergence, faiblement et hautement pathogène, site de clivage, évolution génétique, insertions, minigénome, structures secondaires d'ARN

INTITULE ET ADRESSE DE L'UFR OU DU LABORATOIRE :

UMR 1225 INRAE-ENVT / Interactions Hôte – Agents Pathogènes / Equipe Virologie

Ecole Nationale Vétérinaire de Toulouse (ENVT)
23 chemin des Capelles - bâtiment 10.1
31076 - Toulouse
France

Directeur de thèse : Dr. Volmer Romain

**PERSONNEL ENSEIGNANT
de la Faculté des Sciences Pharmaceutiques de l'Université Paul Sabatier
au 1er octobre 2020**

Professeurs Emérites

Mme BARRE A.	Biologie Cellulaire
M. BENOIST H.	Immunologie
M. BERNADOU J.	Chimie Thérapeutique
M. CAMPISTRON G.	Physiologie
M. GAIRIN J.E.	Pharmacologie
Mme NEPVEU F.	Chimie analytique
M. ROUGE P.	Biologie Cellulaire
M. SALLES B.	Toxicologie

Professeurs des Universités

Hospitalo-Universitaires

Mme AYYOUB M.	Immunologie
M. CESTAC P.	Pharmacie Clinique
M. CHATELUT E.	Pharmacologie
Mme DE MAS MANSAT V.	Hématologie
M. FAVRE G.	Biochimie
Mme GANDIA P.	Pharmacologie
M. PARINI A.	Physiologie
M. PASQUIER C. (Doyen)	Bactériologie - Virologie
Mme ROQUES C.	Bactériologie - Virologie
Mme ROUSSIN A.	Pharmacologie
Mme SALLERIN B.	Pharmacie Clinique
M. VALENTIN A.	Parasitologie

Universitaires

Mme BERNARDES-GÉNISSON V.	Chimie thérapeutique
Mme BOUTET E.	Toxicologie - Sémiologie
Mme COUDERC B.	Biochimie
M. CUSSAC D. (Vice-Doyen)	Physiologie
M. FABRE N.	Pharmacognosie
Mme GIROD-FULLANA S.	Pharmacie Galénique
M. GUIARD B.	Pharmacologie
M. LETISSE F.	Chimie pharmaceutique
Mme MULLER-STAU MONT C.	Toxicologie - Sémiologie
Mme REYBIER-VUATTOUX K.	Chimie analytique
M. SEGUI B.	Biologie Cellulaire
Mme SIXOU S.	Biochimie
M. SOUCHARD J-P.	Chimie analytique
Mme TABOULET F.	Droit Pharmaceutique
M. VERHAEGHE P.	Chimie Thérapeutique

Maîtres de Conférences des Universités

Hospitalo-Universitaires

M. DELCOURT N.	Biochimie
Mme JUILLARD-CONDAT B.	Droit Pharmaceutique
M. PUISSET F.	Pharmacie Clinique
Mme ROUCH L.	Pharmacie Clinique
Mme ROUZAUD-LABORDE C.	Pharmacie Clinique
Mme SERONIE-VIVIEN S (*)	Biochimie
Mme THOMAS F. (*)	Pharmacologie

Universitaires

Mme ARELLANO C. (*)	Chimie Thérapeutique
Mme AUTHIER H.	Parasitologie
M. BERGE M. (*)	Bactériologie - Virologie
Mme BON C. (*)	Biophysique
M. BOUJILA J. (*)	Chimie analytique
M. BROUILLET F.	Pharmacie Galénique
Mme CABOU C.	Physiologie
Mme CAZALBOU S. (*)	Pharmacie Galénique
Mme CHAPUY-REGAUD S.	Bactériologie - Virologie
Mme COLACIOS C.	Immunologie
Mme COSTE A. (*)	Parasitologie
Mme DERA EVE C. (*)	Chimie Thérapeutique
Mme ECHINARD-DOUIN V.	Physiologie
Mme EL GARAH F.	Chimie Pharmaceutique
Mme EL HAGE S.	Chimie Pharmaceutique
Mme FALLONE F.	Toxicologie
Mme FERNANDEZ-VIDAL A.	Toxicologie
Mme GADEA A.	Pharmacognosie
Mme HALOVA-LAJOIE B.	Chimie Pharmaceutique
Mme JOUANJUS E.	Pharmacologie
Mme LAJOIE-MAZENC I.	Biochimie
Mme LEFEVRE L.	Physiologie
Mme LE LAMER A-C.	Pharmacognosie
M. LE NAOUR A.	Toxicologie
M. LEMARIE A.	Biochimie
M. MARTI G.	Pharmacognosie
Mme MONFERRAN S.	Biochimie
M. SAINTE-MARIE Y.	Physiologie
M. STIGLIANI J-L.	Chimie Pharmaceutique
M. SUDOR J. (*)	Chimie Analytique
Mme TERRISSE A-D.	Hématologie
Mme TOURRETTE-DIALLO A. (*)	Pharmacie Galénique
Mme VANSTEELANDT M.	Pharmacognosie
Mme WHITE-KONING M. (*)	Mathématiques

(*) Titulaire de l'habilitation à diriger des recherches (HDR)

Enseignants non titulaires

Assistants Hospitalo-Universitaires

Mme LARGEAUD L.	Immunologie
M. LE LOUEDEC F.	Pharmacologie
M. MOUMENI A.	Biochimie
M. PAGES A.	Pharmacie Clinique
Mme SALABERT A.S	Biophysique
Mme TRIBAUDEAU L.	Droit Pharmaceutique

Attaché Temporaire d'Enseignement et de Recherche (ATER)

M. François-Xavier TOUBLET	Chimie Thérapeutique
----------------------------	----------------------

Remerciements

Tout d'abord, je remercie Monsieur Christophe Pasquier, Doyen de la Faculté des sciences pharmaceutiques, Professeur des Universités et Patricien hospitalier, d'avoir accepté de présider le jury de ma thèse d'exercice.

Je remercie mon directeur de thèse de Pharmacie et de Science, le Docteur Romain Volmer, Maître de Conférences de l'Ecole Nationale Vétérinaire de Toulouse. Son accueil, l'appui constant qu'il m'apporte et ses qualités scientifiques ne font que renforcer ma volonté de continuer dans cette voie : la recherche en Virologie. C'est un plaisir et un honneur de continuer de travailler ensemble, sur un sujet si passionnant que l'évolution des virus de la grippe aviaire. Je remercie également tous les autres membres de l'équipe de Virologie au sein de l'UMR 1225 – IHAP, aux habitants du « Poulailleur », de belles personnes impliquées à rendre la vie au laboratoire si agréable.

Je remercie le Docteur Pierre Verhaeghe, Professeur des Universités, d'avoir accepté d'être parmi les membres de mon jury de thèse. Je lui suis infiniment reconnaissant pour ses enseignements, son écoute, et son implication à la création et la codirection de la filière recherche au sein de la Faculté de Pharmacie de Toulouse.

Un grand merci au Professeur Elisa Boutet, pour tous ses efforts, également consacrés à la création de la filière recherche et à sa direction. Ses qualités pédagogiques et ses nombreux conseils m'ont été précieux.

Je remercie Madame le Docteur Sabine Chapuy-Regaud, Maître de Conférence et Patricien hospitalier, pour avoir également accepté de prendre part au jury. Sa bienveillance, nos discussions et ses conseils m'ont énormément apporté.

Mes remerciements vont ensuite auprès de ma mère, mon père, mon frère et mes grands-parents. Je tiens à leur exprimer toute ma gratitude et mon amour.

Je remercie du fond du cœur mes chers camarades pharmaciens, pharmaciennes et Ami(e)s, pour tous les moments passés à leurs côtés. J'espère créer avec ces belles personnes des milliers d'autres souvenirs à venir.

Enfin, merci à toi Charlotte, d'être mon second souffle et le partenaire de ma vie.

Merci à tous.

**Etude de l'évolution des virus influenza aviaires de sous-type H5
vers les formes hautement pathogènes**

Liste des abréviations

AIV	<i>Avian Influenza Virus</i> ou virus influenza aviaire
aTIS	<i>Alternative translation initiation site</i> ou site d'initiation alternatif de la traduction
CMV	Cytomégalovirus
cTIS	<i>Canonical translation initiation site</i> ou site d'initiation canonique de la traduction (codon AUG)
DMEM	<i>Dulbecco's Modified Eagle's Medium</i>
Fluc	<i>Firefly-luciferase</i> ou luciférase de luciole
GTR+G+I	<i>General Time Reversible + gamma + Invariant site estimation model</i>
HA	Hémagglutinine
HDV	<i>Hepatitis Delta Virus</i> ou virus de l'hépatite delta
HPAIV	<i>Highly Pathogenic Avian Influenza Virus</i> ou virus influenza aviaire hautement pathogène
IAV	<i>Influenza A Virus</i> ou virus influenza A (genre Alphainfluenzavirus)
ISG	<i>Interferon Stimulating Genes</i>
LPAIV	<i>Low Pathogenic Avian Influenza Virus</i> ou virus influenza aviaire hautement pathogène
ML	<i>Maximum likelihood</i> , maximum de vraisemblance
NA	Neuraminidase
NCR	<i>Non Coding Region</i> ou région non codante
Nluc	Nano-luciférase
NP	Nucléoprotéine
Nt	Nucléotide
ORF	<i>Open Reading Frame</i> ou cadre de lecture
PA	Polymérase
PB1	Protéine basique 1
PB2	Protéine basique 2
PolyA	Polyadénylée
RIG-I	<i>Retinoic acid-inducible gene I</i>
RNPv	Ribonucléoparticule virale
ShimH5	A/whistling swan/Shimane/499/83 (H5N3)
SV40	<i>Simian virus 40</i> ou Virus simien 40 (Polyomavirus)
SVF	Sérum de veau fœtal

Table des matières

Remerciements	1
Liste des abréviations	3
Table des matières	4
Tables des illustrations – Figures	5
Table des illustrations – Tableaux	8
1. Introduction bibliographique	9
1.1. Classification et morphologie des virus Influenza A	10
1.2. Cycle cellulaire	13
1.3. Ecologie virale	21
1.4. Les différents pathotypes d’influenza aviaires	24
1.5. Les structures secondaires d’ARN présentes sur le génome des IAV	26
1.6. Plasticité du génome des IAV et les mécanismes d’évolution des virus Influenza aviaires faiblement pathogènes vers des formes hautement pathogènes	29
2. Partie expérimentale : Etude de l’évolution des virus influenza aviaires de sous-type H5 vers les formes hautement pathogènes	33
2.1. Matériels et méthodes	34
2.1.1. Analyse des séquences nucléotidiques HA et phylogénie	34
2.1.2. Prédiction des structures secondaires d’ARN	34
2.1.3. Vecteurs	35
2.1.4. Linkers	36
2.1.5. Obtention des plasmides pGL4.10-minigénome et clonage des séquences linkers	37
2.1.6. Construction du système rapporteur en pcDNA	38
2.1.7. Cellules	38
2.1.8. Essais luciférase	38
2.2. Résultats	40
2.2.1. Identification des émergences HPAIV	40
2.2.2. Etude des prédictions de structures d’ARN	43
2.2.3. Construction et optimisation du système minigénome	50
2.2.4. Contrôle de la spécificité du système minigénome pour la mesure du phénomène d’insertions	57
2.3. Discussion	59
3. Références bibliographiques	66

Tables des illustrations – Figures

Figure 1 : Organisation interne du virion chez IAV (à gauche); images de microscopie électronique montrant des virions AIV sphériques (A), filamenteux (B) et pléiomorphes (C) (à droite), (d'après Krammer et al., 2018 et Noda, 2012).	11
Figure 2 : Organisation du génome de IAV et les stratégies d'expression des protéines virales (ViralZone).	11
Figure 3 : Schéma de l'entrée de IAV au sein de sa cellule cible (d'après Dou et al. 2018).	13
Figure 4 : Schéma simplifié des différentes étapes du cycle cellulaire de AIV (d'après Nuñez et Ross, 2019).	15
Figure 5 : Métabolisme de l'ARN du virus Influenza. (A) Image de microscopie électronique en coloration négative de RNPv purifiées depuis des particules virales. (B) Schéma de la transcription et la réplication du génome viral de IAV. (C) Implication des étapes du cycle correspondant à la transcription / réplication du génome viral de IAV (d'après Pflug et al, 2017).	17
Figure 6 : Représentation de l'architecture de la polymérase virale de IAV (PDB : 4WSB). Représentation en trois dimensions et dans plusieurs orientations des domaines de la sous-unité PB1 (apparentée au modèle de la main droite) : motifs A-F, la boucle d'amorçage (a), les doigts (b), la paume, le pouce et les doigts (b). Les structures des domaines composant la sous-unités PA (c) et PB2 (d) ont été coloriées. Les extrémités 5' et 3' de l'ARNv sont représentés en noir et en ocre respectivement sur les différentes figures précédentes (a-d). Modèles du complexe polymérase viral à l'état « inactif » (e) et de « pré-initiation de la transcription » (f), montrant le réarrangement conformationnel du domaine de fixation flexible de l'extrémités capée des ARN cellulaire, PB2-627 et le domaine endonucléase de la sous-unité PA (Endo). Les canaux d'entrée/sortie de la matrice, d'entrée des NTP, de sortie de l'ARN néosynthétisé, la boucle d'amorçage près du site actif, les terminaisons 5' et 3' de l'ARNv, la séquence polyU en 5' et l'emplacement des nucléoprotéines (NP) sont illustrées (d'après te Velthuis et Fodor 2016).	18
Figure 7 : Modèles de transcription (synthèse d'ARNm) et de réplication (synthèse d'ARNc et d'ARNv) virale chez IAV. Un modèle de la transcription (a). La polymérase est dans l'état de « pré-initiation de la transcription ». L'ARN cellulaire coiffé de l'hôte est lié par le domaine de fixation de la coiffe de PB2 et clivé par le domaine de l'endonucléase de PA. Le domaine de fixation de la coiffe change de conformation pour permettre l'insertion de l'extrémité 3' de l'ARN cellulaire jouant le rôle au sein du site actif de PB1 via le canal de sortie des produits de synthèse. L'extrémité 3' de la matrice (ARNv) est également insérée au niveau du site actif et les NTP entrent par le canal d'entrée des NTP. La transcription est initiée par l'ajout d'un GTP à l'extrémité 3' de l'ARN cellulaire capé. Durant l'élongation, l'ARNm néosynthétisé et la matrice sont séparés et sortent par leurs canaux de sortie respectifs. Lorsque la matrice est entraînée dans le site actif de PB1, les NP se détachent de l'ARNv, sont transloquées sur la surface de la polymérase et se relient sur l'ARNv lorsqu'il sort par le canal de sortie. La polyadénylation de l'ARNm est réalisée par bégaiement de la polymérase sur une séquence polyU au niveau de l'extrémité 5' de la matrice (ARNv). Modèle de synthèse de l'ARNc (b). La polymérase est présentée dans la même conformation que pour la synthèse de l'ARNm. L'extrémité 3' de l'ARNv s'insère dans le site actif et les NTP entrent par le canal d'entrée des NTP. L'initiation de novo se produit au premier résidu de l'ARNv et le nucléotide initiateur est stabilisé par la boucle d'amorçage. Pendant l'élongation, l'ARNc et la matrice sont séparés et sortent par leurs canaux de sortie respectifs. Les NP transloquent au niveau de la surface de la polymérase pour rejoindre la matrice (ARNv), à mesure qu'elle sort du canal de sortie. L'extrémité 5' du produit ARNc est liée par une deuxième polymérase lorsqu'elle sort du canal de sortie et le reste de l'ARN s'associe avec des NP pour former le RNPv. L'extrémité 5' de l'ARNv est libérée de son site de liaison et est entraînée dans le site actif de la polymérase tandis que l'extrémité 3' se retient à la surface de la polymérase. Modèle de la synthèse de l'ARNv (c). La polymérase est représentée dans l'état « inactif », ce qui correspond à la conformation de la polymérase liée à l'extrémité 5' de l'ARNc. Pour que l'initiation de la réplication ait lieu, la polymérase doit interagir avec une seconde polymérase transactivatrice.	

L'extrémité 3' de l'ARNc s'insère au niveau du site actif et les NTP entrent par leur canal d'entrée. L'initiation de novo se produit aux 4^{ème} et 5^{ème} résidus de la matrice (ARNc) sans la participation de la boucle d'amorçage. Le dinucléotide pppApG qui en résulte est déplacé vers l'extrémité 3' de l'ARNc et l'élongation commence. La terminaison se déroule comme décrit pour la synthèse de l'ARNc. L'extrémité 5' du produit d'ARNv est liée par la polymérase transactivante et le reste de l'ARNv s'associe avec des NP pour former le RNPv (d'après de Velthuis et Fodor 2016). 19

Figure 8 : Schéma représentant les différentes étapes composant l'assemblage et la libération des virions chez IAV : le recrutement des segments et l'empaquetage du génome viral (A), le bourgeonnement de la surface de la membrane plasmique au niveau de radeaux lipidiques (B, C) et la libération des virions néosynthétisés (C) (d'après Rossman et Lamb, 2011)..... 21

Figure 9 : Diversité des échanges virologiques au sein des oiseaux et des espèces de mammifères susceptibles et sensibles à l'infection par IAV. Les différents sous-types concernés par chaque espèce sont indiqués (modifiée, d'après Knipe et Howley, 2007)..... 23

Figure 10 : Différences protéiques au niveau du site de clivage de la HA entre les virus influenza aviaires faiblement (LPAIV, à gauche) et hautement pathogènes (HPAIV, à droite), puis les conséquences cliniques chez le poulet (d'après Horimoto et al., 2005). E, T, K et R représentent des acides aminés dont les deux derniers sont basiques : respectivement les résidus lysine (K) et arginine (R)..... 26

Figure 11 : Mécanismes d'évolution des IAV. (A) Accumulation progressive de mutations au niveau du génome de IAV, conduisant à la génération de nouveaux variants : dérive antigénique, responsable des épidémies samsoniennes chez l'Homme (« antigenic drift »). (B) Echange ou réassortiment des segments entre deux souches IAV dans le cadre d'une coïnfection d'une cellule peut entraîner l'émergence d'une nouvelle souche (nouveau sous-type) : cassure antigénique (« antigenic shift »). Ce dernier mécanisme est observé dans le cas de pandémies (d'après Mostafa et al, 2018)..... 30

Figure 12 : Cartes des vecteur pUC57 sans et avec l'insert correspondant à la construction minigénome. 35

Figure 13 : Cartes des vecteur pGL4.10 sans et avec l'insert correspondant à la construction minigénome. 36

Figure 14 : (A.) Arbre phylogénétique total des séquences nucléotidiques HA H5 (modèle GTR+G+I). (B.) Un grossissement sur les principales émergences par insertions a été représenté (le nom des souches HPAIV sont indiqués en bleu et annotés avec leur n° « d'événement d'émergence » correspondant, répertoriés dans le **Tableau 3**)..... 42

Figure 15 : Alignement des séquences nucléotidiques (ADNc) HA H5 des LPAIV correspondant à la région du site de clivage (indiquée par le symbole *, au-dessus des séquences consensus) de la HA des LPAIV étudiés. La séquence logo de l'alignement de toutes les séquences HA H5 de LPAIV avec le pourcentage d'identité et la fréquence de chaque nt au niveau des positions les plus variantes a été représentée (A.). L'alignement des séquences HA (ARNc) génétiquement proches (B.) et éloignées (C.) d'émergences HPAIV par insertions sont aussi représentées. Les ambiguïtés sont surlignées en couleur. La traduction en protéines est indiquée en dessous de chaque séquence nucléotidique. Les séquences HA (et le nom de la souche correspondante) qui ont été sélectionnées comme représentatives des différents groupes de séquences HA de LPAIV H5 sont encadrées. 49

Figure 16 : Schéma du système minigénome rapporteur des phénomènes d'insertions. (Légende : Le symbole * indique le site de clivage de la HA. Les acides aminés correspondant à la séquence protéique environnant le site de clivage sont représentés en gras et en lettres capitales. Les nucléotides sont représentés en lettre minuscules). 50

Figure 17 : Résultats des essais luciférase après la transfection de cellules CDF1 avec les constructions minigénomes en pUC57 et les plasmides pcDNA3.1(+)-PA, PB1, PB2 et NP (WSN). Les conditions avec le complexe de polymérase incomplet (sans apport de pcDNA3.1(+)-NP, motifs avec les lignes transversales) ont été réalisées. Deux expériences indépendantes ont été réalisées. Les moyennes des valeurs brutes du signal Nluc ainsi que les écarts-types de chaque groupe sont représentés. 52

Figure 18 : Résultats des essais luciférase après la transfection de cellules CDF1 avec les constructions minigénomes en pGL4.10 et les plasmides pcDNA3.1(+)-PA, PB1, PB2 et NP (WSN).

Les conditions avec le complexe de polymérase incomplet (sans apport de pcDNA3.1(+)-NP, motifs avec les lignes transversales) ont été représentées. Trois expériences indépendantes ont été réalisées. Les moyennes des valeurs brutes des signaux Nluc normalisées avec les signaux Fluc (Nluc/Fluc) ainsi que les écarts-types de chaque groupe sont représentés. ** et **** indiquent une p-value < 0.01 et 0.0001 respectivement. &&&& indique une différence significative entre le groupe pGL4.10-minig +NP et pGL4.10-ΔSTARTminig +NP avec une p-value < 0.0001. 53

Figure 19 : Résultats des essais luciférase après la transfection de cellules CDF1 avec les constructions pGL4.10-ΔSTARTminig avec les linkers dérivés de la séquence HA de ShimH5 (B.), et les plasmides pcDNA3.1(+)-PA, PB1, PB2 et NP (WSN). Les conditions avec le complexe de polymérase complet ont été représentées. La condition avec ShimH5-V30 sans NP apparaît avec les motifs en lignes transversales). Quatre expériences indépendantes ont été réalisées. Les moyennes des valeurs brutes des signaux Nluc normalisés avec les signaux Fluc (Nluc/Fluc) relatives au signal obtenu en ShimH5-V30 ainsi que les écarts-types de chaque groupe sont représentés. * : p<0.05, ** : p<0.01, *** : p<0.001, **** : p<0.0001 vs DeltaLinker ; ## : p<0.01 vs ShimH5-V30 ; □ : p<0.05 vs 24a2b-V30; && : p<0.01 vs ShimH5-V29. 54

Figure 20 : Résultats des essais luciférase après la transfection de cellules CDF1 avec la construction pGL4.10-ΔSTARTminig avec et sans les linkers dérivés de la séquence HA de ShimH5 (ShimH5-V30 et V29, ainsi que les plasmides pHW2000-PA, PB1, PB2 et NP (H5N8). Les conditions avec le complexe de polymérase complet et sans NP (motifs avec les lignes transversales) ont été représentées. Une expérience a été réalisée (chaque condition en duplicat). Les moyennes des signaux Nluc normalisés avec les signaux Fluc (Nluc/Fluc) de chaque groupe sont représentés. 56

Figure 21 : Résultats des essais luciférase après la transfection de cellules CDF1 avec les constructions linkersShimH5(sauvage et 24a2b)-ORF_Nluc (ΔATG) clonées en pcDNA3.1(+), avec le promoteur CMV en amont. Deux expériences indépendantes ont été réalisées. Les moyennes des valeurs brutes du signal Nluc ainsi que les écarts-types de chaque groupe sont représentés. * indique une différence significative avec une p-value <0.05 par rapport au groupe ShimH5-V30. 57

Figure 22 : Modèle théorique simplifié représentant la tige-boucle conservée prédite au niveau de la région codant le site de clivage de HA en présence du complexe polymérase virale. Les interactions entre le brins 5' et 3' de la matrice ARN se sont soulignées en jaunes (figure adaptée de Gultyaev et al. 2019)..... 61

Figure 23 : Structures consensus issues de l'analyse comparative de la structure d'ARN prédite au niveau de la région codant le motif de clivage de HA, chez les groupes 1 et 2. La fenêtre d'analyse correspond au cadre suivant : 40 nucléotides (nt) – site de clivage de HA (représentée par une flèche noire) – 40 nt. Le programme LocARNA disponible en ligne (webserver : <http://rna.informatik.uni-freiburg.de/LocARNA/Input.jsp>) a été utilisé. Les groupes correspondant respectivement aux séquences HA (ARNc) des souches progénitrices des émergences HPAIV (sous-type H5) par insertions (groupe 1) et phylogénétiquement éloignées de toutes émergences connues (groupe 2) ont été analysées. 62

Figure 24 : Représentation des ORFs possibles du gène de la Nluc (utilisé comme rapporteur dans le système minigénome). Les codons AUG, CUG, UUG et GUG ont été pris en compte. Les codons alternatifs de l'initiation de la traduction (aTIS) présents sur la première moitié de l'ORF classique de la Nluc, ainsi que le codon AUG interne (cTIS) sont indiqués par les flèches en rouge..... 64

Table des illustrations – Tableaux

Tableau 1 : Listes des protéines virales codées par le génome des IAV et leur(s) fonction(s) associées	12
Tableau 2 : Linkers dérivés de séquences HA de ShimH5 (en gras : ATG initiateur, en bleu : séquences s’hybridant avec les bouts cohésifs du vecteur pGL4.10-minigénome digéré par BsaI, nucléotides soulignés : régions variantes pour la formation des linkers version V30, V29 et V28).....	37
Tableau 3 : Liste des différentes émergences HPAIV (H5) caractérisées sur la base de la séquence nucléotidique HA et des données présentes dans la littérature (les émergences par insertions sont indiquées en rouge).	40
Tableau 4 : Emergences indépendantes par insertions et leurs souches LPAIV partageant le plus d’identité avec l’ancêtre LPAIV commun entre eux, sur la base de la séquence nucléotidique HA. ...	43
Tableau 5 : Prédiction des structures secondaires d’ARNc des séquences LPAIV proches des émergences HPAIV par insertions (données concernant la fenêtre d’analyse de 15 nt en amont et en aval de la région correspondant au site de clivage sont représentées).....	45
Tableau 6 : Prédiction des structures secondaires d’ARNc des séquences LPAIV isolées phylogénétiquement de toutes émergences HPAIV (données concernant la fenêtre d’analyse de 15 nt en amont et en aval de la région correspondant au site de clivage sont représentées).	46

1. Introduction bibliographique

1.1. Classification et morphologie des virus Influenza A

Les virus Influenza A (IAV) appartiennent à la famille des *Orthomyxoviridae*, composée par quatre genres Influenza : A, B, C, D, les Thogotovirus et les Isavirus (ICTV 9th Report (2011)). Une nomenclature précise est utilisée afin d'identifier chaque souche virale. Un isolat est caractérisé par le genre de la souche (A, B, C ou D), l'espèce chez laquelle la souche a été isolée pour la première fois (non précisée s'il s'agit d'un virus infectant l'homme), la localisation géographique de l'échantillon, la référence et l'année de l'isolement. Exclusivement pour les IAV, ces virus sont classés en différents sous-types en fonction de la combinaison des deux glycoprotéines majoritaires de surface du virus : l'hémagglutinine (HA) et la neuraminidase (NA). Les sous-types de HA et de NA apparaissent entre parenthèse sous la forme de HxNy. Au sein des IAV, 18 HA et 11 NA différentes ont été décrites. Les glycoprotéines HA et NA permettent respectivement la reconnaissance des récepteurs de type acide sialique situés sur la cellule cible et l'entrée du virus, puis la libération des virions nouvellement produits lors de la réplication virale. Par ailleurs, HA et NA sont les protéines virales les plus antigéniquement variables et ciblées par la réponse immunitaire à médiation humorale (Kosik and Yewdell 2019).

Les IAV sont des virus enveloppés, dont la membrane lipidique provient de la cellule hôte depuis laquelle ils ont bourgeonné. La particule virale est de forme globalement sphérique (**Figure 1**), voire allongée dans certains cas, avec un diamètre variant entre 80 et 120 nm (Krammer et al. 2018; Noda 2011).

Le génome, situé au cœur du virion, est constitué de 8 segments d'ARN simple brin de polarité négative (sb -, ARN_v) dont la longueur totale est de 13,5 kB (**Figure 2**). Douze à quatorze protéines virales sont codées et produites en fonction des souches. Chaque segment d'ARN est recouvert par des nucléoprotéines (NP) pour former des ribonucléoparticules virales (RNP_v). Le complexe de réplication viral constitué par les protéines PA (*polymerase*), PB1 (*polymerase basic 1*) et PB2 (*polymerase basic 2*) nécessaires à la transcription et réplication du génome viral est lié au niveau des deux extrémités 5' et 3' de chaque segment, qui sont hybridés partiellement (Te Velthuis and Fodor 2016). Plusieurs protéines sont retrouvées enchâssées au niveau de l'enveloppe du virus, à savoir les glycoprotéines HA et NA constituant les principaux déterminants antigéniques puis le canal ionique M2. La protéine M1 quant à elle est retrouvée sur la face interne de la membrane lipidique et constitue la matrice (Krammer et al. 2018; Harris et al. 2006).

D'autres protéines telles que PB2-S1, PB1-F2, PA-X et NEP (*nuclear export protein*, ou protéine d'export nucléaire) sont codées sur des cadres de lectures alternatifs (Figure 2 et Tableau 1).

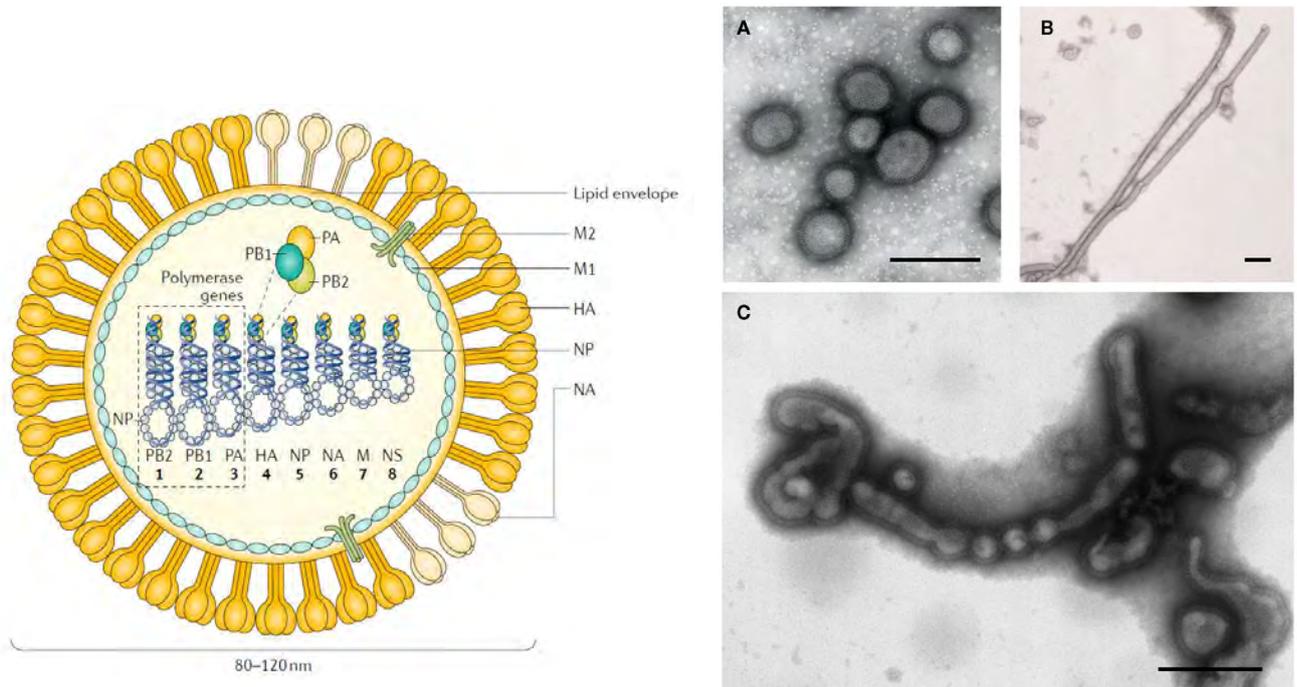


Figure 1 : Organisation interne du virion chez IAV (à gauche); images de microscopie électronique montrant des virions AIV sphériques (A), filamenteux (B) et pléiomorphes (C) (à droite), (d'après Krammer et al., 2018 et Noda, 2012).

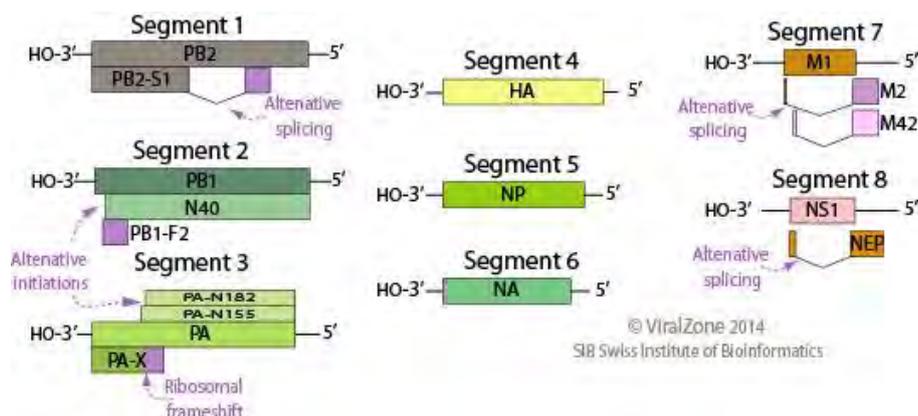


Figure 2 : Organisation du génome de IAV et les stratégies d'expression des protéines virales (ViralZone).

Tableau 1 : Listes des protéines virales codées par le génome des IAV et leur(s) fonction(s) associées

Segment d'ARNv	Protéine(s)	Fonction(s) de la ou des protéines
1	1. PB2 2. PB2-S1	1. Sous-unité du complexe polymérase, reconnaissance de la coiffe 5' des ARNm de l'hôte dans le cadre de la transcription virale (vol de coiffe) (Fodor 2013; Pflug et al. 2017; Ruigrok et al. 2010; Te Velthuis and Fodor 2016) 2. Forme tronquée de PB2 issue d'un épissage alternatif, inhibiteur de RIG-I (Yamayoshi et al. 2015)
2	1. PB1 2. PB1-F2 3. PB1-N40	1. Sous-unité du complexe polymérase en charge de la synthèse d'ARN (Fodor 2013; Pflug et al. 2017; Ruigrok et al. 2010; Te Velthuis and Fodor 2016) 2. Facteur de virulence, induction de la réponse inflammatoire - recrutement des neutrophiles, activation de l'apoptose (Kamal, Alymova, and York 2017) 3. Rôle dans la réplication de certaines souches (Q. Wang et al. 2019)
3	1. PA 2. PA-X 3. PA-N155 4. PA-N182	1. Sous-unité du complexe polymérase virale, permet le clivage de l'ARNm lors de la transcription (Fodor 2013; Pflug et al. 2017; Ruigrok et al. 2010; Te Velthuis and Fodor 2016) 2. Facteur de virulence, régule la réponse immunitaire de l'hôte – répression de gènes cellulaires de l'hôte (via son activité endonucléase) (Hu, Ma, and Liu 2018; Nogales et al. 2018) 3. Rôles dans la réplication (Muramoto et al. 2013) 4. Rôles dans la réplication (Muramoto et al. 2013)
4	HA	Glycoprotéine de surface, rôle dans l'attachement des virions aux récepteurs cellulaires, la fusion, facteur de virulence principale des souches influenza aviaires hautement pathogènes, cible principale des anticorps neutralisant de l'hôte (Abdelwhab, Veits, and Mettenleiter 2013; Gamblin and Skehel 2010; Kosik and Yewdell 2019; Steinhauer 1999; Nao N, Yamagishi J et al. 2017)
5	NP	Liaison aux ARNv et ARNc, import nucléaire des RNPv (signaux NLS), régulation de la transcription/réplication des ARNv (Neumann, Castrucci, and Kawaoka 1997; Pflug et al. 2017; Te Velthuis and Fodor 2016)
6	NA	Glycoprotéine de surface, rôle dans la libération des virions, cible majoritaire des anticorps neutralisants (Kosik and Yewdell 2019; Rossman and Lamb 2011)
7	1. M1 2. M2 3. M42	1. Protéine de matrice, rôle dans l'export des RNPv (Brunotte et al. 2014; Cros and Palese 2003) 2. Canal ionique, rôle dans l'étape d'entrée du virus (fusion, acidification des endosomes) et de la maturation au niveau du trans-Golgi (Pielak and Chou 2011) 3. Forme alternative et fonctionnelle de M2 (épissage alternatif) (Wise et al. 2012)
8	1. NS1 2. NEP (NS2)	1. Facteur de virulence, régulation de l'activation du système IFN (Ayllon and García-Sastre 2015; Nogales et al. 2018) 2. Rôle dans l'export des RNPv (Brunotte et al. 2014; Neumann, Castrucci, and Kawaoka 1997)

1.2. Cycle cellulaire

Le contact du virion avec la cellule cible s'effectue par la mise en jeu de récepteurs de type acide sialique avec la HA virale trimérique (Dou et al. 2018; Gamblin and Skehel 2010). Il existe une spécificité d'hôte selon le type de liaison glycosidique entre l'acide sialique terminal et le galactose adjacent qui détermine le spectre d'hôte des IAV. Ainsi les acides sialiques formant une liaison de type alpha 2-6 glycosidique avec le galactose sont ciblés par des virus IAV de types « humains », tandis que le récepteur cellulaire décrivant une liaison alpha 2-3 est requis pour l'infection AIV des souches d'origine aviaire. Il a été cependant démontré la présence de ces récepteurs aviaires au niveau de certains tissus restreints au niveau de l'arbre respiratoire profonds (bronchioles, alvéoles pulmonaires) chez l'Homme, expliquant une possible susceptibilité aux souches IAV aviaires après une exposition prolongée (Dou et al. 2018; Van Riel et al. 2006).

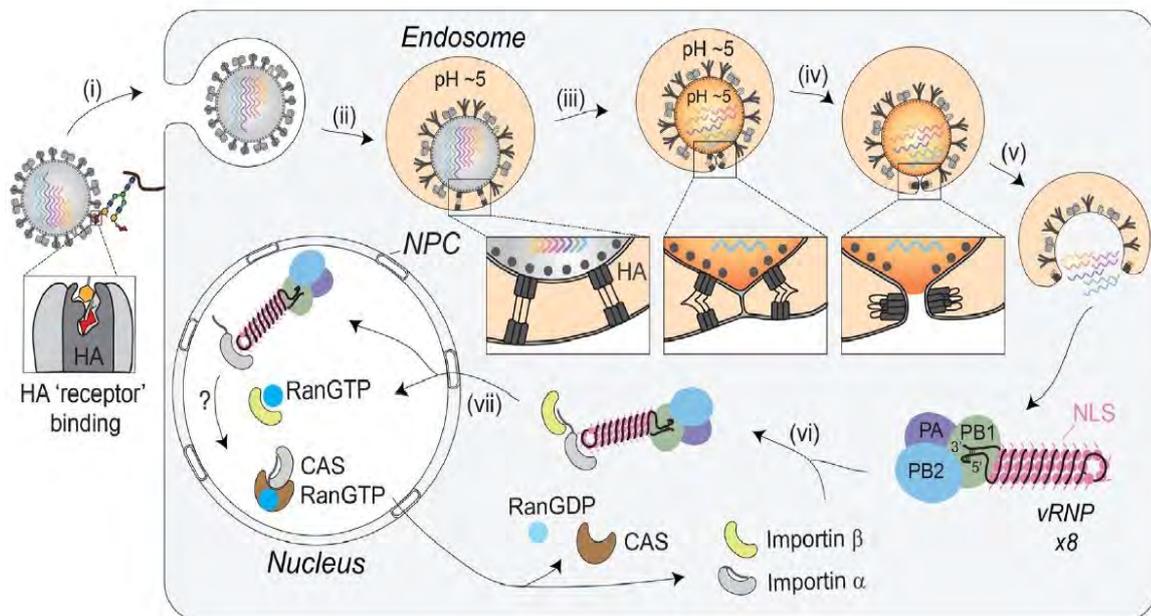


Figure 3 : Schéma de l'entrée de IAV au sein de sa cellule cible (d'après Dou et al. 2018)

Le virus entre dans la cellule par différentes voies, majoritairement par endocytose (Dou et al. 2018; Edinger, Pohl, and Stertz 2014) (**Figure 3** et **Figure 4**). Ensuite, le virion est adressé vers le compartiment endosomal tardif. L'abaissement progressif du pH va provoquer une succession de changements conformationnels de la HA, ayant pour résultat l'exposition du peptide de fusion situé au niveau de l'extrémité N terminale (N-ter) de HA2 (obtenue après maturation protéolytique de HA0 par les protéases cellulaires). Ce dernier va être projeté vers la membrane cible endosomale. Les membranes lipidiques virales et cellulaires vont progressivement se rapprocher puis fusionner, faisant apparaître le diaphragme d'hémi fusion qui va ensuite s'étirer, laissant finalement communiquer les compartiments intra virionique et cytoplasmique (mécanisme de fusion de type I) (Dou et al. 2018).

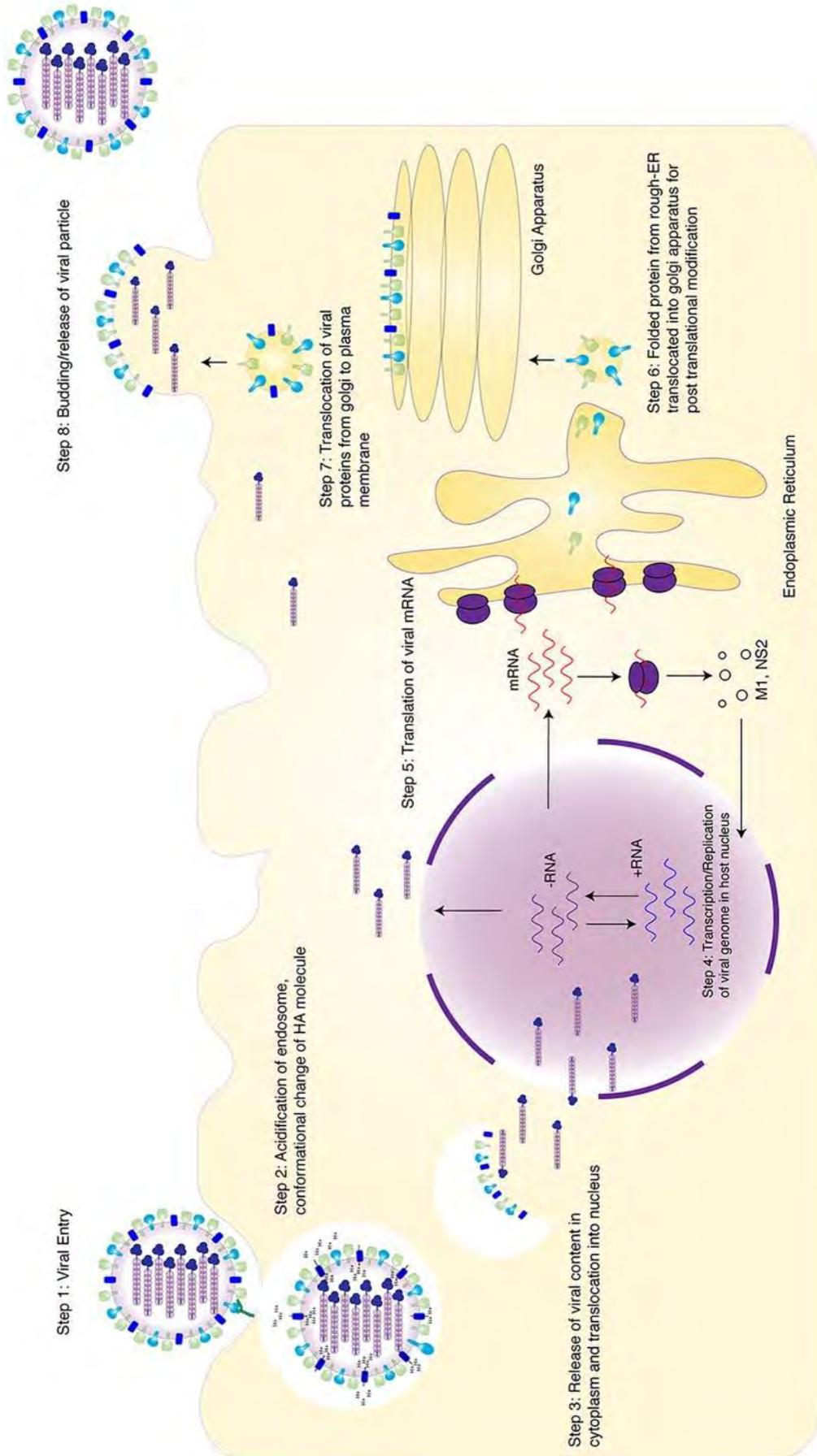


Figure 4 : Schéma simplifié des différentes étapes du cycle cellulaire de AIV (d'après Nuñez et Ross, 2019).

Les ribonucléoprotéines virales (RNPv) sont ensuite transloquées vers le compartiment cytoplasmique. La totalité du métabolisme de l'ARN viral dépendant des fonctions nucléaires, les RNPv vont être ensuite utilisées à la fois pour la transcription et la réplication du génome viral (Te Velthuis and Fodor 2016; Pflug et al. 2017) (**Figure 5**, **Figure 6** et **Figure 7**). Le complexe de réplication en cis de chaque segment ou ARNv, constitué par les sous-unités PB1, PB2 et PA permet la formation des ARN messagers (ARNm) viraux (de polarité positive : +). Un mécanisme particulier de « vol de coiffe » (*cap snatching*), permet la production des ARNm viraux. PB2 lie l'extrémité capée des ARN messagers cellulaires tandis que la matrice RNPv est fixée par la sous-unité PB1. La protéine PA clive l'ARN cellulaire 12 à 15 nucléotides en aval de la coiffe. Le fragment 5' coiffé cellulaire créé est ensuite transféré sur l'ARNv, jouant le rôle d'amorce pour la transcription (Pflug et al. 2017). Pour finir, une queue polyA est ajoutée au niveau de l'extrémité 3' de l'ARN⁺ viral coiffé par bégaiement de l'ARN polymérase virale (encore appelé : *polymerase stuttering*) au niveau d'une séquence répétée en uraciles. Une structure de type épingle à cheveux située à 7 nucléotides en amont de la séquence polyU provoquerait une contrainte au niveau de la progression de la polymérase virale, induisant un retour en arrière sur la matrice et la copie réitérative de la base uracile (Pflug et al. 2017; Luo et al. 1991). La réplication du génome viral est rendue possible via l'utilisation des segments viraux (ARNv) comme matrice et permet ainsi la synthèse des intermédiaires génomiques : ARN complémentaires, de polarité positive (ARNc (+)) (Pflug et al. 2017). Chaque ARNv est ensuite synthétisé à partir des ARNc produits et s'associe avec de nouvelles NP et le complexe de réplication PA-PB1-PB2. La forme transitoire d'ARNc s'assemble aussi avec des NP et le complexe polymérase viral PA-PB1-PB2 (Pflug et al. 2017). Les facteurs régulant le déclenchement de la transcription ou de la réplication du génome viral reste encore inconnus. Les quantités stœchiométriques des différents acteurs moléculaires nécessaires à la transcription / réplication et la localisation du complexe réplicatif complet dans le noyau pourraient expliquer le choix de tels chemins métaboliques.

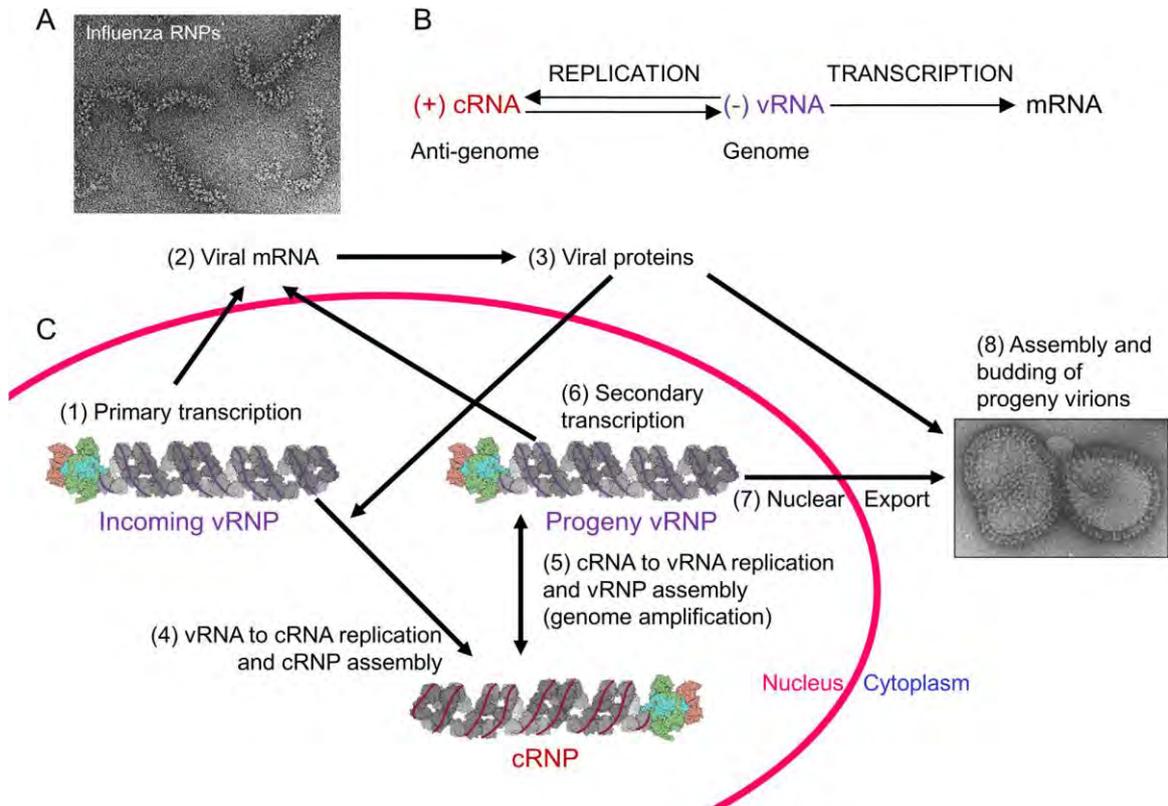


Figure 5 : Métabolisme de l'ARN du virus Influenza. (A) Image de microscopie électronique en coloration négative de RNPv purifiées depuis des particules virales. (B) Schéma de la transcription et la réplication du génome viral de IAV. (C) Implication des étapes du cycle correspondant à la transcription / réplication du génome viral de IAV (d'après Pflug et al, 2017).

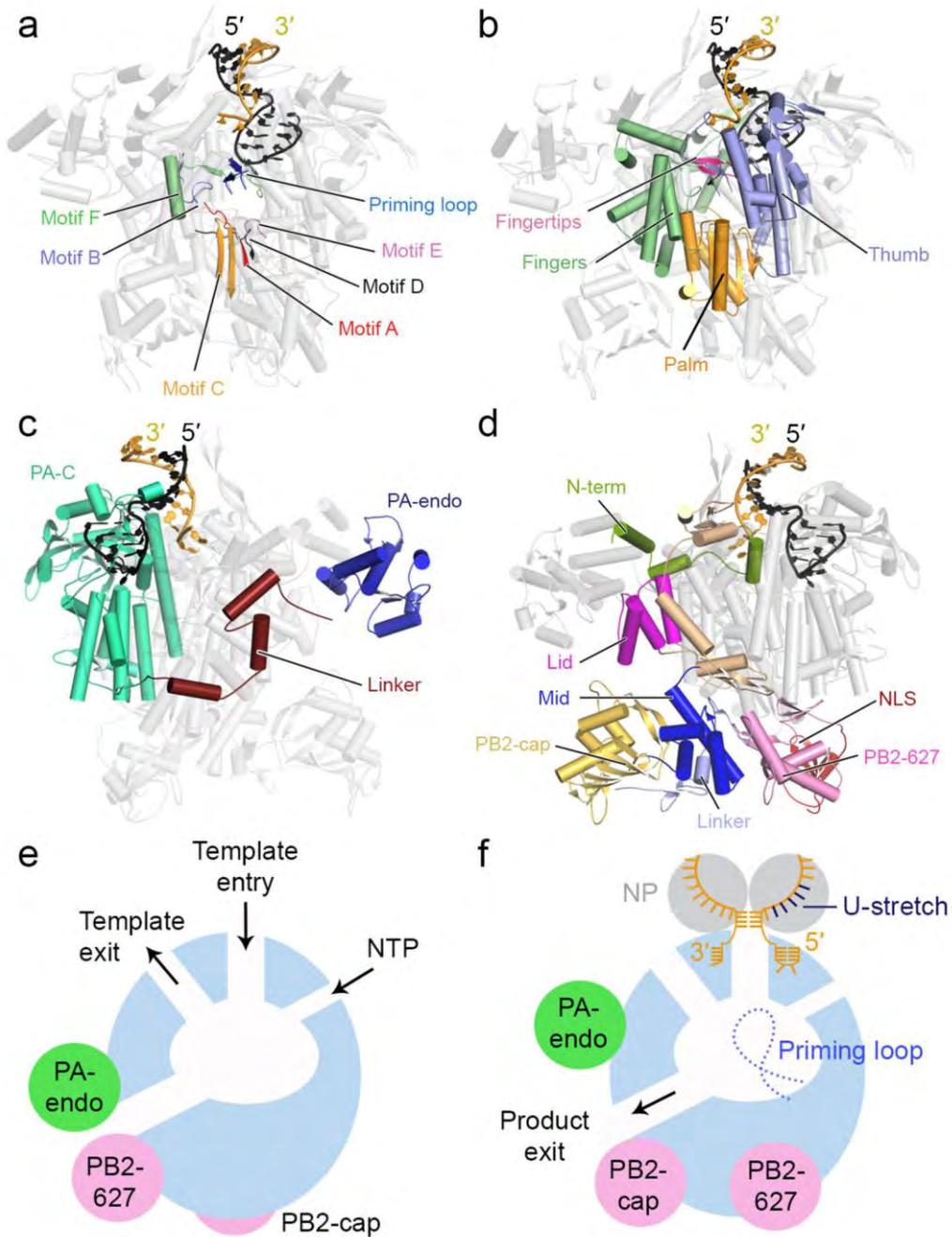


Figure 6 : Représentation de l'architecture de la polymérase virale de IAV (PDB : 4WSB). Représentation en trois dimensions et dans plusieurs orientations des domaines de la sous-unité PB1 (apparentée au modèle de la main droite) : motifs A-F, la boucle d'amorçage (a), les doigts (b), la paume, le pouce et les doigts (b). Les structures des domaines composant la sous-unités PA (c) et PB2 (d) ont été coloriées. Les extrémités 5' et 3' de l'ARNv sont représentés en noir et en ocre respectivement sur les différentes figures précédentes (a-d). Modèles du complexe polymérase viral à l'état « inactif » (e) et de « pré-initiation de la transcription » (f), montrant le réarrangement conformationnel du domaine de fixation flexible de l'extrémités capée des ARN cellulaire, PB2-627 et le domaine endonucléase de la sous-unité PA (Endo). Les canaux d'entrée/sortie de la matrice, d'entrée des NTP, de sortie de l'ARN néosynthétisé, la boucle d'amorçage près du site actif, les terminaisons 5' et 3' de l'ARNv, la séquence polyU en 5' et l'emplacement des nucléoprotéines (NP) sont illustrées (d'après te Velthuis et Fodor 2016).

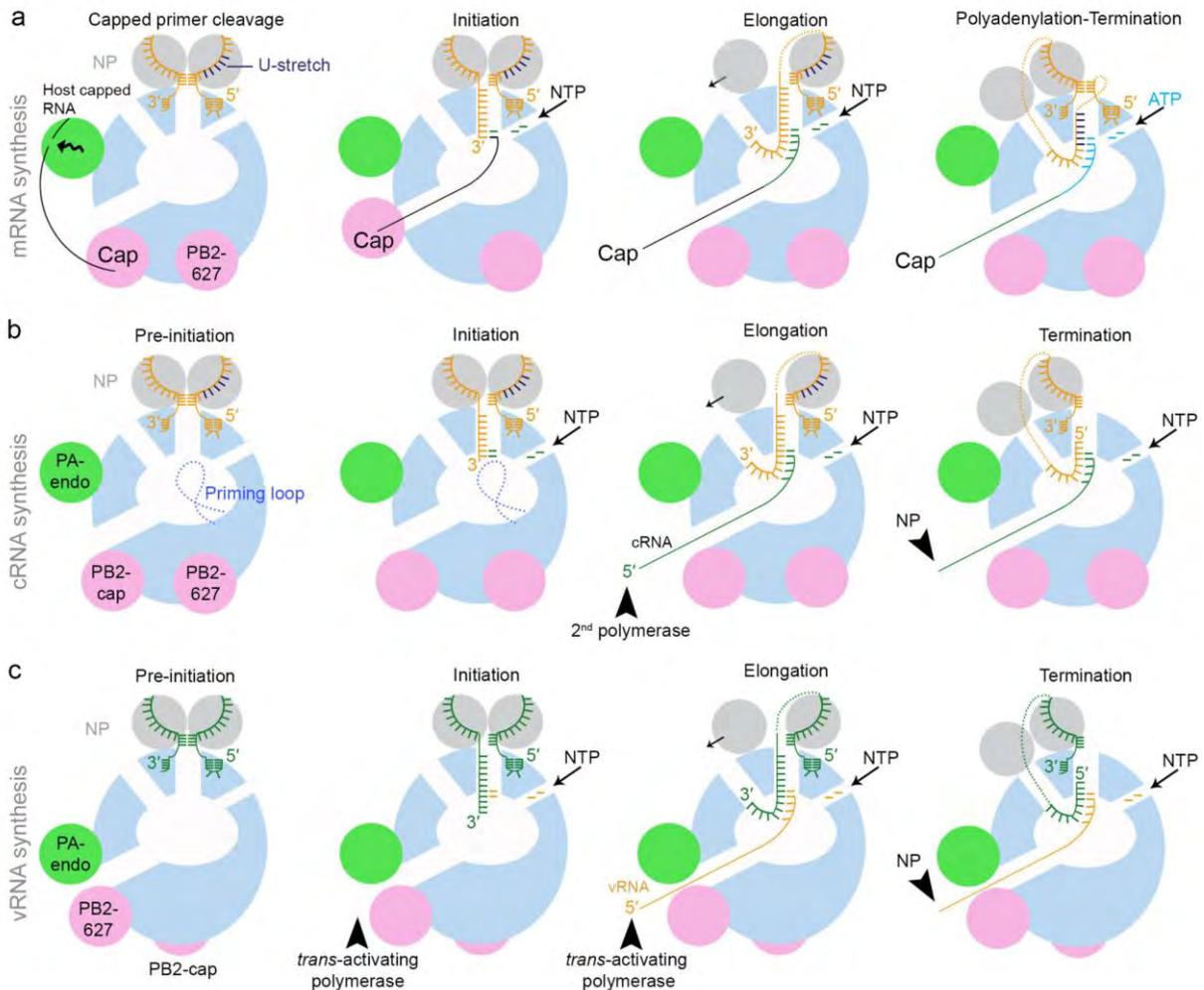


Figure 7 : Modèles de transcription (synthèse d'ARNm) et de répllication (synthèse d'ARNc et d'ARNv) virale chez IAV. Un modèle de la transcription (a). La polymérase est dans l'état de « pré-initiation de la transcription ». L'ARN cellulaire coiffé de l'hôte est lié par le domaine de fixation de la coiffe de PB2 et clivé par le domaine de l'endonucléase de PA. Le domaine de fixation de la coiffe change de conformation pour permettre l'insertion de l'extrémité 3' de l'ARN cellulaire jouant le rôle au sein du site actif de PB1 via le canal de sortie des produits de synthèse. L'extrémité 3' de la matrice (ARNv) est également insérée au niveau du site actif et les NTP entrent par le canal d'entrée des NTP. La transcription est initiée par l'ajout d'un GTP à l'extrémité 3' de l'ARN cellulaire capé. Durant l'élongation, l'ARNm néosynthétisé et la matrice sont séparés et sortent par leurs canaux de sortie respectifs. Lorsque la matrice est entraînée dans le site actif de PB1, les NP se détachent de l'ARNv, sont transloquées sur la surface de la polymérase et se refixent sur l'ARNv lorsqu'il sort par le canal de sortie. La polyadénylation de l'ARNm est réalisée par bégaiement de la polymérase sur une séquence polyU au niveau de l'extrémité 5' de la matrice (ARNv). **Modèle de synthèse de l'ARNc (b).** La polymérase est présentée dans la même conformation que pour la synthèse de l'ARNm. L'extrémité 3' de l'ARNv s'insère dans le site actif et les NTP entrent par le canal d'entrée des NTP. L'initiation de novo se produit au premier résidu de l'ARNv et le nucléotide initiateur est stabilisé par la boucle d'amorçage. Pendant l'élongation, l'ARNc et la matrice sont séparés et sortent par leurs canaux de sortie respectifs. Les NP transloquent au niveau de la surface de la polymérase pour rejoindre la matrice (ARNv), à mesure qu'elle sort du canal de sortie. L'extrémité 5' du produit ARNc est liée par une deuxième polymérase lorsqu'elle sort du canal de sortie et le reste de l'ARN s'associe avec des NP pour former le RNPv. L'extrémité 5' de l'ARNv est libérée de son site de liaison et est entraînée dans le site actif de la polymérase tandis que l'extrémité 3' se retient à la surface de la polymérase. **Modèle de la synthèse de l'ARNv (c).** La polymérase est représentée dans l'état « inactif », ce qui correspond à la conformation de la polymérase liée à l'extrémité 5' de l'ARNc. Pour que l'initiation de la répllication ait lieu, la polymérase doit interagir avec une seconde polymérase transactivatrice. L'extrémité 3' de l'ARNc s'insère au niveau du site actif et les NTP entrent par leur canal d'entrée. L'initiation de novo se produit aux 4^{ième} et 5^{ième} résidus de la matrice (ARNc) sans la participation de la boucle d'amorçage. Le dinucléotide pppApG qui en résulte est déplacé vers l'extrémité 3' de l'ARNc et l'élongation commence. La terminaison se déroule comme décrit pour la synthèse de l'ARNc. L'extrémité 5' du produit d'ARNv est liée par la polymérase transactivante et le reste de l'ARNv s'associe avec des NP pour former le RNPv (d'après te Velthuis et Fodor 2016).

Les ARNm et RNPv vont être ensuite exportés hors du noyau. L'action conjointe de la protéine NEP et de la protéine de matrice M1 permet l'export des RNPv néosynthétisées. M1 recouvrant les signaux NLS des NP (Dou et al. 2018), NEP interagit avec la machinerie d'export de la cellule et provoque alors la sortie des RNPv hors du noyau. Au niveau cytoplasmique les ARNm viraux sont pris en charge par les ribosomes cellulaires qui vont ensuite produire les protéines virales.

Les ARNm viraux codant pour HA, NA et le canal ionique M2 sont adressés vers le réticulum endoplasmique (RE) et traduits. Les protéines synthétisées sont alors enchâssées au niveau de la membrane lipidique du RE. Ces dernières sont ensuite adressées vers la membrane plasmique de la cellule en utilisant le système de sécrétion de la cellule et vont s'accumuler au niveau de radeaux lipidiques (Dou et al. 2018; Nayak, Hui, and Barman 2004). Les RNPv exportées hors du noyau sont transportées vers les radeaux lipidiques via le cytosquelette d'actine et le réseau vésiculaires issue du réticulum endoplasmique (De Castro Martin et al. 2017). Les huit RNPv constituant le génome complet du virus vont être adressées au niveau de ces sites de formation de nouveaux virions. Des interactions fines au niveau de zones moins denses en NP entre les segments d'ARN vont permettre l'empaquetage correct du génome viral complet. Certaines structures secondaires d'ARN seraient impliquées (Dadonaité et al. 2019; Dawson, Lazniewski, and Plewczynski 2018; Williams et al. 2018; Ferhadian et al. 2018). Après la déformation de la membrane plasmique induite par la composition biochimique des radeaux lipidiques, les virions néosynthétisés sortent de la cellule par bourgeonnement et clivage entre la HA virale et les acides sialiques membranaires cellulaires par la NA (**Figure 8**), libérant ainsi le virus (Dou et al. 2018; Rossman and Lamb 2011).

Les virions néosynthétisés deviennent infectieux après l'action de protéases cellulaires qui vont cliver ou maturer la protéine HA0, entraînant la formation du dimère HA1 / HA2 fonctionnel (Gamblin and Skehel 2010).

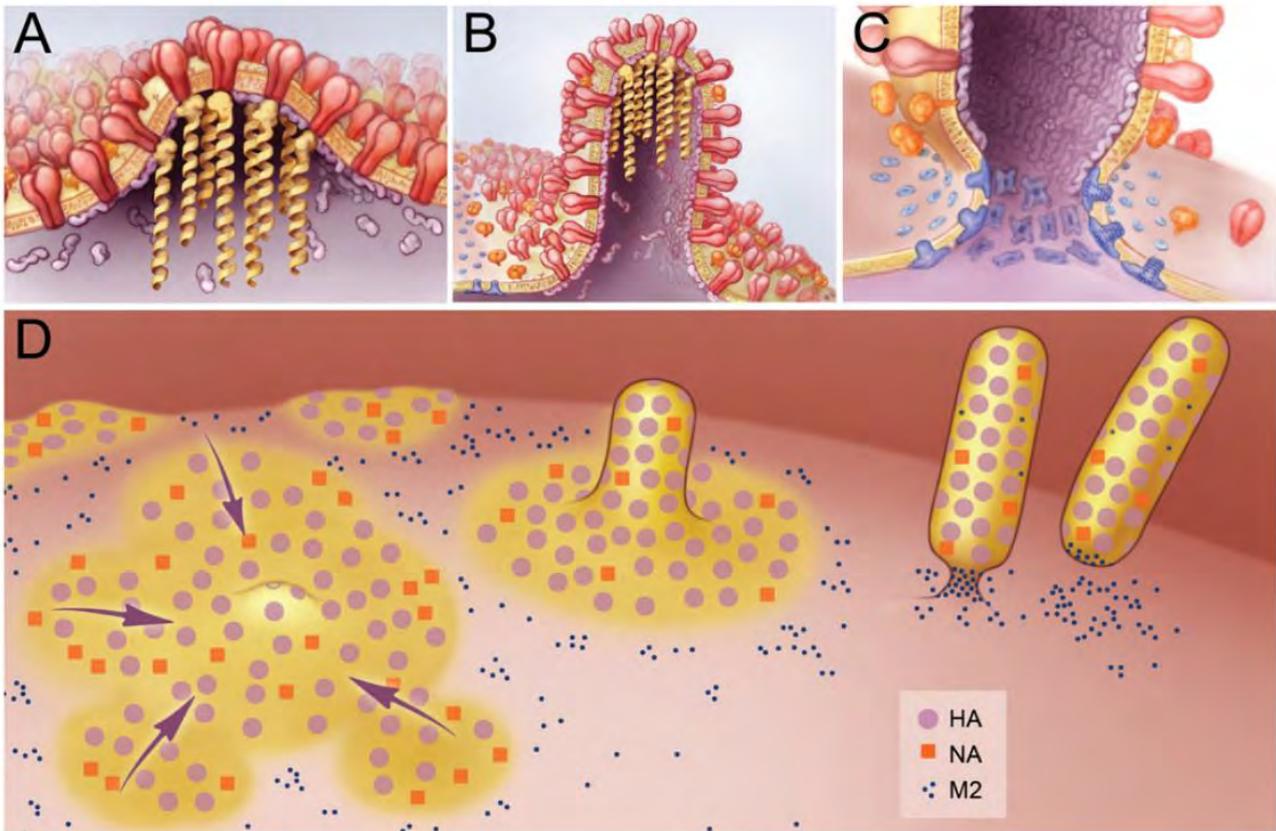


Figure 8 : Schéma représentant les différentes étapes composant l'assemblage et la libération des virions chez IAV : le recrutement des segments et l'emballage du génome viral (A), le bourgeonnement de la surface de la membrane plasmique au niveau de radeaux lipidiques (B, C) et la libération des virions néosynthétisés (C) (d'après Rossman et Lamb, 2011).

1.3. Ecologie virale

Les IAV infectent une très grande diversité d'espèces animales. Ils ont été détectés chez plus d'une centaine d'espèces d'oiseaux répartis dans 13 ordres différents (Munster and Fouchier 2009; Vandegrift et al. 2010; Ohuchi et al. 1989; Webster et al. 1992). Les oiseaux sauvages sont infectés par la majorité des IAV (**Figure 9**). Une multitude de virus issus de la combinaison entre les sous-types HA1-16 et NA 1-9 ont été isolés chez ces hôtes et sont dénommés virus influenza aviaries (AIV) (Mostafa et al. 2018).

Les oiseaux aquatiques de l'ordre des Anseriformes (incluant les canards, les oies et les cygnes) suivis par les Charadriiformes (les mouettes, les goélands, les sternes, les bécasseaux, les pluviers) constituent les réservoirs naturels des AIV. D'autres oiseaux aquatiques comme les espèces appartenant aux ordres de Ciconiiformes, Gaviiformes, Gruiformes, Pelecaniformes, Podicipediformes et Procellariiformes sont moins fréquemment infectés. D'autres oiseaux

terrestres sont occasionnellement infectés par des IAV, tels que les espèces des Ordres des Columbiformes, Falconiformes, Passeriformes, Piciformes et Piscittaciformes. Cela suggère que ces derniers jouent un rôle de « *spill-over* » de la zoonose (impliqués dans la propagation secondaire du virus chez d'autres espèces : effet de « débordement ») après un contact avec des volailles ou des oiseaux aquatiques infectés (Olsen et al. 2006; Wahlgren 2011; Vandegrift et al. 2010; França, Stallknecht, and Howerth 2013).

Chez les Ansériformes, les canards barboteurs ou *Anas spp.* (classés au sein de la famille des *Anatidae*) sont les plus fréquemment infectés par les AIV, en particulier l'espèce *Anas platyrhynchos* (canard col-vert). La quasi-totalité des 16 HA et 9 NA décrites ont été retrouvées chez *Anas spp.* Une circulation plus intense de ces virus (prévalence virologique entre 10 et 60%) est observée avant et pendant l'automne en Eurasie et en Amérique du Nord chez *Anas platyrhynchos*. Une plus faible prévalence d'individus infectés est observée durant le printemps et l'été (moins de 10%)(Vandegrift et al. 2010; Spackman 2009).

Chez les oiseaux classés au sein des Charadriiformes, composés par les oiseaux de rivage, les goélands et les sternes, le schéma épidémiologique des AIV diffère par rapport aux Ansériformes. Deux sous-types de HA : HA13 et H16, ont été isolés presque exclusivement chez les Charadriiformes. De plus, les oiseaux de rivage révèlent un pic de prévalence d'infection durant le printemps et l'été (Vandegrift et al. 2010; Spackman 2009; Olsen et al. 2006).

Certaines caractéristiques liées aux niches écologiques des oiseaux aquatiques sauvages et aux virus influenza aviaires (AIV) permettent d'expliquer une circulation importante et efficace de ces virus au sein de cette population. Une très forte densité et variété d'espèces sont retrouvées dans les lieux de rassemblement d'un grand nombre d'espèces d'oiseaux sauvages (Olsen et al. 2006). De plus, le virus est excrété en de très grandes quantités via l'appareil respiratoire et surtout au niveau des selles. Ces facteurs favorisent les infections interindividuelles et la maintenance de ces virus au sein des populations d'oiseaux. Un brassage génétique important des AIV chez les oiseaux est alors observé (Mostafa et al. 2018; Olsen et al. 2006; Wahlgren 2011). D'autres part, des caractéristiques virologiques et environnementales favorisent la propagation et la persistance de ces virus dans le biotope de ses hôtes naturels. En effet, les IAV peuvent persister plusieurs semaines dans un environnement humide et froid (Brown et al. 2009; Stallknecht et al. 1990).

Compte tenu du spectre d'hôte naturel des IAV, la distribution et la transmission de ces virus sont dépendantes des mouvements de populations d'espèces d'oiseaux, qui, pour une large

partie d'entre-elles sont migratrices. Différents couloirs majeurs de migrations ont pu être décrits (Olsen et al. 2006).

D'autres espèces animales autres que les oiseaux sont aussi sujets à l'infection aux IAV, notamment les mammifères marins, les porcs, les furets, les humains, les chevaux, les chats et les chiens. De plus, les chauves-souris constituent un autre réservoir de AIV à prendre en compte (Wahlgren 2011; Webster et al. 1992).

Les volailles (oiseaux domestiqués utilisés pour la production de viande ou d'œufs destinés à la consommation humaine, pour la reproduction de ce type d'oiseaux, ou pour la production de gibier de repeuplement ou d'autres produits commerciaux, constitués par des Gallinacées (dindes, poulets, cailles, poules, pintades, faisans) et des palmipèdes (canards domestiques, oies)) sont infectés par contacts directs ou indirects (transmission via l'environnement : sol / eau contaminé ou fomites) avec des oiseaux sauvages contaminés (Wahlgren 2011; Vandegrift et al. 2010; Webster et al. 1992). L'infection démarre soit au niveau des cellules épithéliales de l'appareil respiratoire supérieur ou du tube digestif avant d'aller se propager au niveau de l'appareil respiratoire inférieur (**Figure 10**). En comparaison aux Ansériformes et Charadriiformes, les dindes, poulets et cailles sont généralement plus sensibles à l'infection aux AIV.

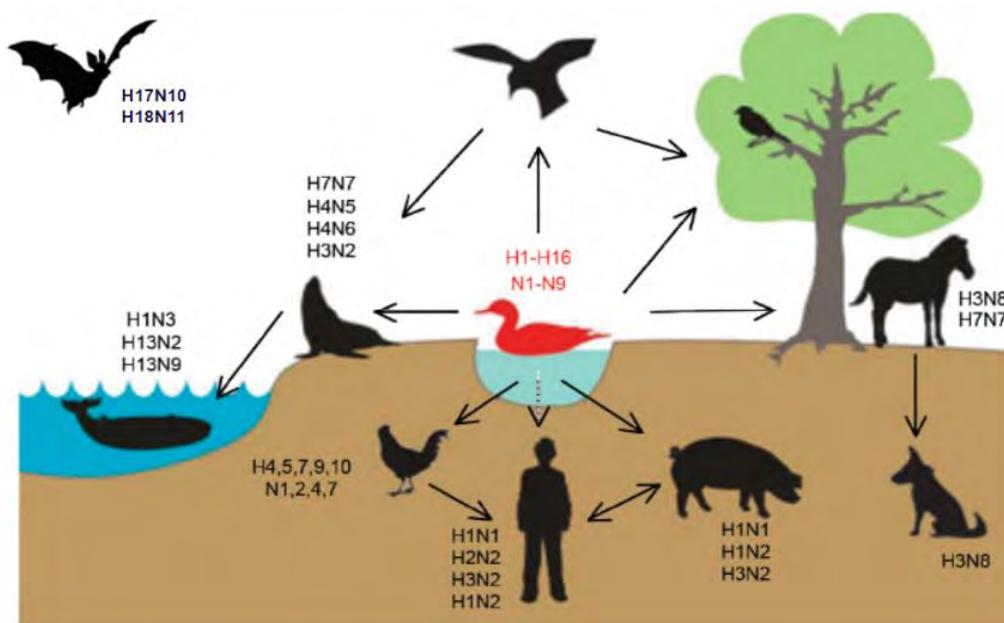


Figure 9 : Diversité des échanges virologiques au sein des oiseaux et des espèces de mammifères susceptibles et sensibles à l'infection par IAV. Les différents sous-types concernés par chaque espèce sont indiqués (modifiée, d'après Knipe et Howley, 2007).

1.4. Les différents pathotypes d'influenza aviaires

Les AIV peuvent être classés en deux catégories en fonction de leur pathogénicité chez les volailles : les virus influenza aviaires faiblement pathogènes (LPAIV : *low pathogenic avian influenza virus*) et hautement pathogènes (HPAIV : *highly pathogenic avian influenza virus*). Généralement, les LPAIV causent peu ou pas de symptômes, que ce soit chez les oiseaux sauvages aquatiques (Anseriformes et Charadriiformes), captifs ou d'élevage. Les LPAIV se répliquent principalement au niveau des cellules épithéliales de l'appareil respiratoire supérieur et du tube digestif des oiseaux. De très grandes quantités de virus sont excrétées via les fèces, expliquant une contamination importante de l'environnement et une transmission principalement oro-fécale chez ces hôtes (Olsen et al. 2006; Wahlgren 2011). Malgré une clinique majoritairement inapparente chez les oiseaux aquatiques, certaines souches LPAIV peuvent perturber le comportement migratoire chez les canards (Vandegrift et al. 2010).

Dans le cadre d'une infection avec une souche LPAIV chez le canard et le poulet, l'infection sera restreinte au niveau de l'appareil respiratoire et du tube digestif. La HA va être clivée ou maturée par des protéases extracellulaires de type trypsine reconnaissant un motif de clivage monobasique. Ces enzymes sont exclusivement exprimées au niveau de l'appareil respiratoire et du tube digestif de l'hôte. Le tableau clinique d'une infection est peu marqué. Classiquement, 5% en moyenne de mortalité sont observés chez ces individus et des signes de morbidité sont identifiés dans 50% des cas. Classiquement, le poulet est plus susceptible à l'infection avec un LPAIV que le canard. Les symptômes les plus retrouvés sont des signes ORL comme des éternuements, la toux, des râles, des sécrétions oculaires, ainsi que de la diarrhée (Steinhauer 1999).

Lors d'une infection avec une souche HPAIV chez un Galliforme, un tableau clinique plus sévère est souvent observé. Des hémorragies, des œdèmes, des nécroses et le dysfonctionnement de la majorité des organes sont observés (Steinhauer 1999). Des symptômes respiratoires et/ou neurologiques majeurs caractérisent le tableau clinique, dont l'issue est souvent fatale. L'infection est associée à une sécrétion importante de cytokines pro inflammatoires responsable en grande partie des signes cliniques (Kuribayashi et al. 2013). Chez le poulet et la dinde les HPAIV causent une infection systémique létale pouvant provoquer 100 % de mortalité au cours des 48 heures suivant l'infection (David E. Swayne 2007; D E Swayne and Pantin-Jackwood 2006).

Dans la majorité des cas, une infection avec une souche HPAIV induit peu de mortalité et de morbidité chez les hôtes appartenant aux ordres des Anseriformes et des Charadriiformes

(Vandegrift et al. 2010). L'isolement d'une souche HPAIV au sein des Charadriiformes reste rare.

On note très fréquemment une différence de sensibilité à l'infection avec une souche HPAIV entre l'hôte canard et poulet. Cette différence semble être multifactorielle, impliquant notamment la différence de tropisme des virus et la réponse immunitaire de l'hôte (Steinhauer 1999; Evseev and Magor 2019). La réponse immunitaire innée caractérisée par la sécrétion d'interférons bêta (IFN- α et β) et la synthèse des ISGs (*Interferon Stimulating Genes*) est beaucoup plus précoce, efficace et régulée chez le canard. L'absence de la protéine RIG-I et une réponse inflammatoire massive chez le poulet expliqueraient en partie sa vulnérabilité accrue face à l'infection (Evseev and Magor 2019).

Les souches LPAIV ont un potentiel d'évolution vers des formes dites HPAIV hautement pathogènes. Cette évolution a été observée naturellement chez les sous-types H5 et H7. Elle est caractérisée par l'acquisition d'acides aminés basiques au niveau du site de clivage de HA (Abdelwhab, Veits, and Mettenleiter 2013). L'obtention d'un tel motif polybasique va considérablement changer le tropisme du virus (**Figure 10**). La maturation protéolytique de HA va pouvoir être effectuée par d'autres protéases golgiennes de type furine dont l'expression est ubiquitaire. Le virus va alors pouvoir infecter d'autres types cellulaires, tels que les cellules endothéliales et provoquer une extension de l'infection au niveau de la circulation sanguine et lymphatique, expliquant le tableau clinique sévère avec une infection généralisée dans le cas d'un HPAIV (Steinhauer 1999).

L'OIE prévoit une définition pour la classification des souches H5 et H7 influenza aviaire suivante (Article 10.4.1 du Code Sanitaire pour les animaux terrestres 2011) :

« Les virus responsables de l'influenza aviaire à déclaration obligatoire peuvent être classés en deux catégories : virus hautement pathogènes et virus faiblement pathogènes :

Les virus responsables de l'influenza aviaire à déclaration obligatoire hautement pathogènes ont un indice de pathogénicité par voie intraveineuse supérieur à 1,2 chez les poulets âgés de six semaines, ou entraînent une mortalité d'au moins 75 % chez les poulets âgés de quatre à huit semaines infectés par voie intraveineuse. Les virus de sous-types H5 et H7 dont l'indice de pathogénicité par voie intraveineuse n'est pas supérieur à 1,2 ou qui entraînent une mortalité inférieure à 75 % lors d'un test de létalité par voie intraveineuse, doivent être séquencés pour déterminer si de multiples acides aminés basiques sont présents sur le site de clivage de la molécule d'hémagglutinine (HA0) ; si la séquence d'acides aminés est similaire à celle observée chez d'autres virus responsables de l'influenza aviaire à déclaration obligatoire hautement

pathogènes isolés précédemment, il doit être considéré qu'il s'agit du virus responsable de la forme hautement pathogène de la maladie. Les virus responsables de l'influenza aviaire à déclaration obligatoire faiblement pathogènes comprennent tous les virus influenza de type A appartenant aux sous-types H5 et H7, qui ne sont pas des virus responsables de la forme hautement pathogène de la maladie. »

L'indice de pathogénicité représente le score clinique moyen quotidien pour chaque animal observé à l'issue une période de 10 jours après inoculation intraveineuse chez 10 poulets âgés de quatre à huit semaines. Un score de zéro est attribué à un animal non malade, un pour un animal malade, deux pour un animal très malade ou paralysé et trois pour un animal mort.

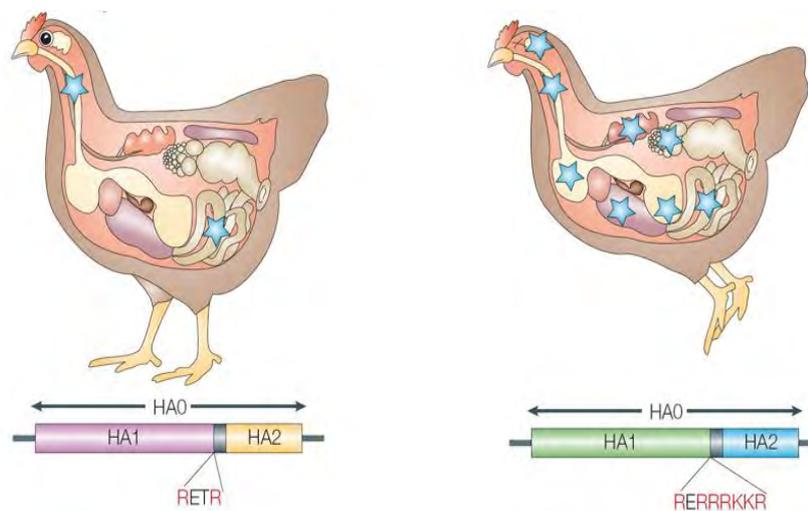


Figure 10 : Différences protéiques au niveau du site de clivage de la HA entre les virus influenza aviaries faiblement (LPAIV, à gauche) et hautement pathogènes (HPAIV, à droite), puis les conséquences cliniques chez le poulet (d'après Horimoto et al., 2005). E, T, K et R représentent des acides aminés dont les deux derniers sont basiques : respectivement les résidus lysine (K) et arginine (R).

1.5. Les structures secondaires d'ARN présentes sur le génome des IAV

Depuis plusieurs décennies, la structure de l'ARN est progressivement décrite comme jouant un rôle majeur dans la régulation de plusieurs fonctions cellulaires chez beaucoup d'organismes dont les virus. La réplication virale, l'empaquetage du génome, l'édition de l'ARN et son épissage sont finement régulés par ces structures (Mehedi et al. 2013; Kolakofsky 2016; Lim and Brown 2018; Zhang and Landick 2016).

Ces découvertes ont été rendues possibles via le développement de plusieurs méthodes : des modèles de prédiction de structures d'ARN ainsi que des techniques de détection directe de ces motifs. Deux grands types de méthodes peuvent être utilisées pour la prédiction de structures secondaires d'ARN. La première est fondée sur des concepts thermodynamiques comme la minimisation de l'énergie libre. Plusieurs structures sont calculées et celles étant les plus stables sont retenues comme les plus probables d'exister dans la nature (SantaLucia 1998). L'autre méthode de prédiction est basée sur l'alignement multiple de séquences relativement homologues et l'analyse de covariation (Parsch, Braverman, and Stephan 2000). Une substitution primaire suivie d'une mutation compensatrice au niveau de bases non adjacentes, correspondant potentiellement à un point de coévolution est déterminée. Ce procédé donne des informations sur des zones de contrainte au niveau d'une séquence nucléotidique donnée (ex : zone d'appariement sur la tige d'une structure de type épingle à cheveux).

Ensuite, des technologies telles que la cartographie chimique de l'ARN ont permis d'identifier directement la présence de motifs d'ARN chez plusieurs virus dans des conditions *in vitro* et cellulaires mimant l'infection virale (Williams et al. 2018; Lin and Miles 2019; Flynn et al. 2016; Merino et al. 2005).

Chez les AIV, de multiples motifs au niveau de leur génome, tant sur l'ARN_v, l'ARN_c que sur l'ARN_m ont été observées et jouent des rôles variés au niveau du cycle viral. Les principaux motifs sont décrits dans cette revue en détail (Ferhadian et al. 2018).

Au niveau des huit segments composant le génome viral des IAV, les séquences terminales 5' et 3' (de 13 et 12 nucléotides) sont très conservées chez toutes les souches du même genre. En effet, ces dernières vont remplir le rôle de promoteurs et réguler la réplication du génome viral ainsi que la transcription. C'est à ce niveau précis où le complexe polymérase viral va se fixer et former les RNP_v, composés aussi de la molécule d'ARN viral imparfaitement recouverte de NP. Au cours de trente dernières années, plusieurs modèles de topologie du promoteur viral ont été suggérés. L'ARN_v nu décrit au niveau de ses promoteurs une forme de manche (modèle *panhandle*) dans lequel les bases sont majoritairement appariées jusqu'aux positions 15 et 14 (depuis les extrémités 5' et 3' respectivement). En revanche la fixation du complexe polymérase au niveau du promoteur entraîne un changement de celui-ci. Une tige-boucle apparaît en 5' terminal entre les nucléotides 1-10 appelée « crochet » (modèle *hook*). L'extrémité 3' terminale libre simple brin va pouvoir entrer au sein de la sous-unité PB1 et arriver au niveau du site catalytique. Plusieurs études ont montré que la région duplex du promoteur est primordiale, de telle façon que des mutations induites à ce niveau par mutagenèse dirigée perturbe

l'empaquetage correct des segments et/ou la réplication (Ferhadian et al. 2018; Pflug et al. 2014; J. Wang et al. 2017).

Comme dit précédemment, des structures secondaires régulent aussi la transcription virale chez les IAV. Il a été montré que la structure en épingle à cheveu au niveau 5' du promoteur est essentielle à l'activité endonucléase de PA lors du mécanisme de vol de coiffe pour la formation de l'ARNm (Leahy et al. 2001). De plus, une autre structure de type tige-boucle secondairement induite au niveau 3' terminale de chaque segment participerait aussi à cette activité endonucléasique (Leahy, Dobbyn, and Brownlee 2001). La formation de la queue polyA est aussi sous la dépendance de structures secondaires d'ARN au niveau du promoteur. Lors du passage de la polymérase virale sur le brin matrice, cette dernière va rencontrer une structure en tige boucle qui provoquerait un mécanisme de bégaïement de la polymérase, entraînant la copie réitérative d'uraciles présents entre les positions 15 et 17 depuis l'extrémités 3' terminal (De Vlugt, Sikora, and Pelchat 2018; Pflug et al. 2017; 2014; Luo et al. 1991).

D'autre part, le promoteur est impliqué aussi dans la reconnaissance des IAV par le système immunitaire innée. La protéine RIG-I est majoritairement activée en présence de ARNv, induisant la production d'IFN. La structure en manche (*panhandle*) du promoteur et l'extrémité 5'triphosphate sont tous deux des signaux agonistes nécessaires pour l'activation de RIG-I (M.-K. Lee et al. 2016; Ferhadian et al. 2018).

Outre la région duplex proche du promoteur en 5' et 3' terminales, d'autres zones au niveau interne des gènes viraux sont impliquées dans des interactions entre les segments, permettant un empaquetage viral efficace. Une étude réalisée sur le segment NP a montré la présence de structure de type pseudo-nœud (Gulyaev et al. 2014). D'autres structures de type épingle à cheveux ont été prédites sur le segment M, dont la perturbation par mutagenèse dirigée entraîne une diminution de la capacité répliquative chez la souche A/Puerto Rico/8/34 (PR8, H1N1), suggérant un rôle dans l'empaquetage du génome (Kobayashi et al. 2016). Ces interactions inter-segments semblent être sous-type-spécifiques et pourraient en partie expliquer les mécanismes de réassortiments préférentiels entre plusieurs souches. De telles interactions ont été décrites entre deux virus de sous-type H3N2 et H5N2 (Gavazzi et al. 2013). De plus, ces interactions semblent se réaliser au niveau de régions particulières. Chez H5N2, les segments NS et PB1 interagissent au niveau de deux structures tiges boucles, à l'instar du mécanisme de *kissing-loops* observé chez les rétrovirus. Quand bien même ces mécanismes semblent plutôt spécifiques en termes de sous-types et de segments concernés, plusieurs types d'interactions impliquant des structures d'ARN différentes peuvent avoir lieu (Dadonaite et al. 2019). Cela dénote une capacité d'adaptation certaine des IAV pour l'empaquetage efficace de leur génome et les réassortiments lors de coïnfection.

D'autre part, certaines régions structurées au niveau du segment HA semble être soumises à une pression de sélection positive et joueraient potentiellement un rôle dans l'évolution. Des études bioinformatiques combinant des prédictions de structures secondaires d'ARN et des analyses de covariation indiquent qu'au niveau de la région codant le site de clivage de HA des structures de types épingle à cheveux seraient présentes. Des analyses de cartographie de l'ARN montrent que ces zones sont plutôt faiblement recouvertes de NP, corroborant ainsi avec les données de prédictions de structures. Cependant, des études supplémentaires mettant directement en évidence la présence de ces structures prédites sont nécessaires afin de déterminer leur impact réel dans l'évolution des IAV (Williams et al. 2018; Gulyaev et al. 2016; Dawson, Lazniewski, and Plewczynski 2018; Dadonaite et al. 2019; Ferhadian et al. 2018).

La régulation du métabolisme de l'ARNm notamment l'épissage est aussi soumis à la présence de structures de secondaires. De tels mécanismes ont été décrits sur les segments NS et M, ayant pour finalité la production d'ARNm alternatifs codant pour NS1 - NS2 (ou NEP) et M1 - M2 respectivement (Ferhadian et al. 2018; Jiang et al. 2016).

1.6. Plasticité du génome des IAV et les mécanismes d'évolution des virus Influenza aviaires faiblement pathogènes vers des formes hautement pathogènes

Les virus influenza A peuvent échanger leurs segments génomiques lors de la coïnfection d'une même cellule et ainsi former de nouveaux virus par réassortiment (Figure 11). Ces « cassures antigéniques » sont responsables de pandémies chez les *Orthomyxoviridae*. Plusieurs vagues pandémiques causées par des virus Influenza A ont marqué l'histoire : la grippe espagnole (1918-1920, causée par une souche H1N1), asiatique (1957-1958, H2N2), de Hong Kong (1968-1970, H3N2) et porcine (2009-2011, H1N1), présentant ce pathogène comme un problème majeur en santé humaine (Bjørnstad and Viboud 2016; Monne et al. 2014; Nelson and Holmes 2007).

L'ARN polymérase virale ARN dépendante de IAV est peu fidèle. Un taux de mutations pouvant atteindre jusqu'à 10^{-4} substitutions/nucléotide/matrice répliquée a été calculé chez certaines souches (Pauly, Procaro, and Lauring 2017). Cette forte variabilité entraîne la formation de populations de génomes différents, appelées quasi-espèces (Bragstad, Nielsen, and Fomsgaard 2008; Barbezange et al. 2018). Les épidémies saisonnières touchant la population humaine sont principalement à l'origine d'acquisitions de mutations (substitutions)

au niveau des glycoprotéines virales permettant d'échapper à la pression du système immunitaire et responsables d'un « glissement antigénique » progressif (Nelson and Holmes 2007) (Figure 11).

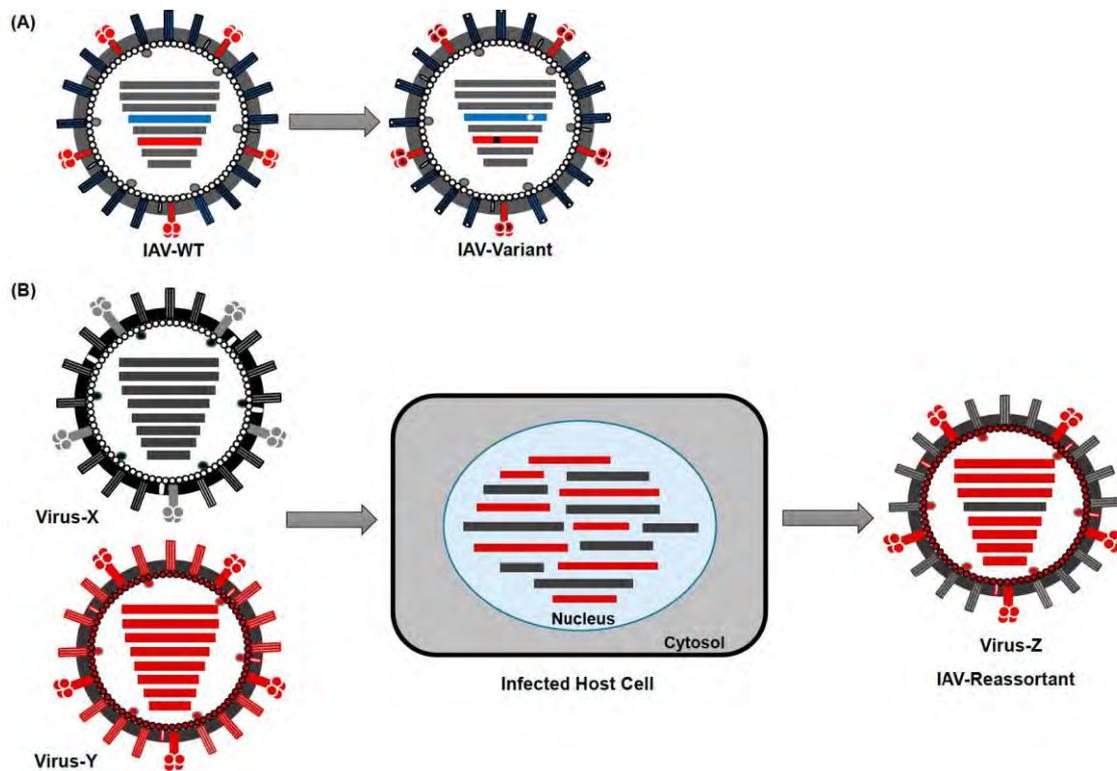


Figure 11 : Mécanismes d'évolution des IAV. (A) Accumulation progressives de mutations au niveau du génome de IAV, conduisant à la génération de nouveaux variants : dérive antigénique, responsable des épidémies samsonières chez l'Homme (« antigenic drift »). (B) Echange ou réassortiment des segments entre deux souches IAV dans le cadre d'une coïnfection d'une cellule peut entraîner l'émergence d'une nouvelle souche (nouveau sous-type) : cassure antigénique (« antigenic shift »). Ce dernier mécanisme est observé dans le cas de pandémies (d'après Mostafa et al, 2018).

Comme énoncé précédemment, nous pouvons distinguer deux grands types de virus influenza chez les oiseaux : les virus influenza aviaire faiblement (LPAIV) et hautement pathogènes (HPAIV), dont ces derniers sont naturellement issus de virus influenza aviaires faiblement pathogènes (LPAIV) qui ont évolué, aboutissant à l'apparition d'un motif protéique polybasique au niveau du site de clivage de HA (Abdelwhab, Veits, and Mettenleiter 2013). Les HPAIV apparus naturellement appartiennent aux sous-type H5 et H7. De manière intéressante, cette restriction ne semble pas être dépendante de caractéristiques spécifiques au gènes HA de sous-type H5 et H7. Grâce aux techniques de génétique inverse, des virus de sous-

type non H5 et non H7 ont pu acquérir un phénotype hautement pathogènes semblable au HPAIV naturel via le remplacement de l'environnement de leur site de clivage de HA par des séquences polybasiques provenant de souches HPAIV (Veits et al. 2012). Mise à part un exemple d'émergence de site polybasique par insertion ayant eu lieu chez un virus de sous-type H4, le fait que l'évolution de virus LPAIV vers des formes hautement pathogènes ne soit observé que chez les sous-types H5 et H7 reste encore énigmatique. Des facteurs génétiques présents au niveau de l'environnement du site de clivage de HA seulement chez les virus de sous-type H5 et H7 pourraient expliquer la capacité de ces derniers à évoluer vers des formes hautement pathogènes chez la volaille.

Différents mécanismes génétiques ont été proposés comme étant responsables de l'apparition du motif polybasique au niveau du site de clivage de HA : la recombinaison génétique avec des gènes cellulaires ou viraux, des erreurs de la polymérase virale (substitutions) et des insertions de nucléotides (Abdelwhab, Veits, and Mettenleiter 2013; Abolnik 2017; Lowen 2017; Nao N, Yamagishi J et al. 2017; Pasick et al. 2005).

Les phénomènes de glissement ou de bégaiement des polymérases virales (*polymerase slippage* ou *stuttering*) participeraient à l'évolution des génomes viraux en provoquant l'insertion de nucléotides (Abolnik 2017; Kolakofsky 2016; Nao N, Yamagishi J et al. 2017). Dans le cas de matrices à ARN, des régions riches en adénosines (A) ou uridines (U) sont favorables à ce phénomène caractérisé par une pause de la polymérase lors de la transcription et/ou de la réplication du génome (Koscielniak et al. 2018; Olsper et al. 2015; Le Mercier and Kolakofsky 2019; Mehedi et al. 2013). Ceci a pour conséquence des insertions, notamment par duplication de codons situés en amont de la zone modifiée du génome. Les régions riches en A (ou en U) sont caractérisées par des structures secondaires d'ARN de type épingles à cheveux, pouvant conduire à un ralentissement de la progression de la polymérase lors de l'élongation. Un lien a été suggéré entre la présence de structures secondaires d'ARN et des événements d'insertions chez les virus Influenza de sous-type H5 et H7 (Abolnik 2017; Nao N, Yamagishi J et al. 2017). Ce mécanisme pourrait expliquer une grande partie des émergences connues jusqu'à ce jour, comme la souche HPAIV H5N2 qui est apparue au Mexique en 1995 (Garcia et al. 1996; Perdue et al. 1997).

Des recombinaisons entre des gènes cellulaires et le segment HA ont été décrites chez un nombre conséquent de AIV de sous-type H7 (Pasick et al. 2005; Abdelwhab, Veits, and Mettenleiter 2013). Du fait de la présence d'une structure de type tige boucle conservée au niveau du site de clivage de HA, un mécanisme de changement de matrice au niveau de cette structure a été proposé comme favorisant la recombinaison, à l'image de ce qui peut être

observé au sein des rétrovirus (Moumen et al. 2003; Simon-Lorier et al. 2010). Cependant le mécanisme exact et les facteurs expliquant pourquoi des recombinaisons ont été observées uniquement chez les virus de sous-types H7 restent encore inconnus.

Actuellement, les élevages de volailles dans lesquels un LPAIV de type H5 ou H7 est détecté subissent un abattage total, selon le principe de précaution visant à prévenir le risque potentiel d'évolution des LPAIV H5 et H7 en HPAIV. Un classement des LPAIV H5 et H7 en fonction de leur risque d'évolution vers un HPAIV pourrait être une donnée prise en compte par les autorités de santé publique animale, par exemple en ne préconisant l'abattage total que lorsque la souche identifiée présente un risque élevé d'évolution vers un HPAIV.

C'est pourquoi la détermination de facteurs impactant l'évolution des LPAIV est primordial dans la gestion du risque d'émergence de souches HPAIV

Dans ce travail, nous nous sommes focalisés sur les AIV de sous-type H5. L'objectif du projet a été de développer de nouvelles méthodes afin de répondre aux questions suivantes :

- i. Est-ce que toutes les souches LPAIV H5 ont la même probabilité d'évoluer vers des HPAIV ?
- ii. Si cette probabilité est différente selon les souches, l'environnement nucléotidique du site de clivage de la HA influence-t-il l'émergence de HPAIV par le biais de structures secondaires d'ARN favorables aux insertions ?

Tout d'abord, une analyse phylogénétique du segment HA H5 couplée à une étude des structures secondaires d'ARN de la séquence correspondant à la région du site de clivage de la HA a été réalisée. Ensuite, nous avons essayé de développer un système expérimental permettant d'analyser la fréquence avec laquelle des nucléotides sont insérés dans la séquence nucléotidique entourant le site de clivage de HA. Il s'appuie sur une méthode qui utilise un système de type minigénome dans lequel le gène codant la luciférase est inséré en aval d'une séquence nucléotidique correspondant à celle entourant le site de clivage de HA (séquence appelé « linker »), et en décalage de cadre de lecture par rapport au codon de départ (Nao N, Yamagishi J et al. 2017). Ainsi, la luciférase ne s'exprimera que si son cadre de lecture est restauré par l'insertion d'un ou deux nucléotides dans la séquence nucléotidique entourant le clivage de HA. Notre objectif a été de créer et d'optimiser ce système afin de pouvoir tester plusieurs fragments de séquence nucléotidique HA H5. A terme, ce système pourrait aboutir en une classification du risque d'évolution des LPAIV H5 et/ou H7 vers les HPAIV en fonction de la séquence nucléotidique de HA.

2. Partie expérimentale : Etude de l'évolution des virus influenza aviaries de sous-type H5 vers les formes hautement pathogènes

2.1. Matériels et méthodes

2.1.1. Analyse des séquences nucléotidiques HA et phylogénie

Les séquences cDNA correspondant aux segments 4 de AIV de sous-type H5, enregistrées sur la base de données FludB (<https://www.fludb.org/brc/home.spg?decorator=influenza> , date de consultation : 10/01/2019) ont été utilisées afin d'analyser leurs évolutions relatives. Plusieurs critères ont été retenus afin de sélectionner les séquences HA H5 à analyser :

- i. Les séquences de sous-type H5 et issues d'hôtes aviaires ont été incluses dans l'analyse.
- ii. Les séquences HA mutantes provenant de laboratoires ont été écartées de l'étude.
- iii. Les séquences HA HPAIV dérivant de la souche A/Goose/Guangdong/1/96 (H5N1) ont été écartées de l'analyse. Effectivement, ces séquences descendent de la même souche hautement pathogène et biaiseraient donc l'étude. Seule une séquence HA correspondant à cet événement d'émergence a été retenue.

Au total, 1101 séquences nucléotidiques HA H5 ont été téléchargées puis alignées. Un arbre phylogénétique a été construit avec le logiciel MEGA 6 à partir des séquences HA H5, selon une approche de maximum de vraisemblance. Le critère d'information d'Akaike nous a permis de déterminer le modèle de substitution le plus adapté pour notre analyse phylogénétique : GTR+G+I. Une analyse de la robustesse de l'arbre par « bootstrap » a été réalisée (1000 répétitions). Plusieurs séquences HA de LPAIV ont ensuite été sélectionnées selon leurs positions relatives sur l'arbre par rapport aux émergences HPAIV connues.

2.1.2. Prédiction des structures secondaires d'ARN

Les programmes mFold (<http://unafold.RNA.albany.edu/?q=mfold/RNA-Folding-Form>, version 3.6) et RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>, ViennaRNA Package 2.0) ont été utilisés pour prédire les structures secondaires des ARNc(+) et ARNv(-), correspondants aux régions avoisinant le site de clivage de la HA. Plusieurs fenêtres d'analyse ont été utilisées : 15 et 30 nucléotides (nt) en amont et en aval du site correspondant à la séquence de clivage de HA.

2.1.3. Vecteurs

L'insert correspondant à la construction minigénome a été synthétisé et cloné dans le plasmide de clonage pUC57 par Genscript®. La construction minigénome est constituée de plusieurs éléments (**Figure 12**) :

- les séquences 3' et 5' NCR du segment 4 de la souche A/chicken/Germany-MV/R10048/2016 (H5N8), flanquant le site de clonage des linkers suivi par l'ORF de la Nano-luciférase,
- le site de clonage des linkers comprenant les sites de restrictions BsaI,
- l'ORF codant pour la Nano-luciférase (Nluc) délété de son ATG initiateur (Nluc(Δ ATG)),
- la séquence promotrice de l'ARN polymérase I de poulet, positionnée en aval du 5'NCR (segment 4, H5N8) et en antisens par rapport à l'ORF de la Nluc,
- la séquence régulatrice auto-clivable du virus de l'hépatite delta (HDV), située en amont de la construction minigénome, avant la séquence 3' NCR,
- la séquence du promoteur constitutif de la souris du phosphoglycérate kinase 1.

Le plasmide pUC57 comprend un gène de résistance à la kanamycine permettant une production sous milieu de sélection. La construction minigénome a été transféré dans le plasmide pGL4.10 comportant les éléments suivants (**Figure 13**) : deux séquences polyA entourant la zone de clonage de la construction minigénome suivie par l'ORF codant pour la *Firefly luciférase* (Fluc, luciférase de luciole), permettant la normalisation du signal Nluc et le gène de résistance à l'ampicilline.

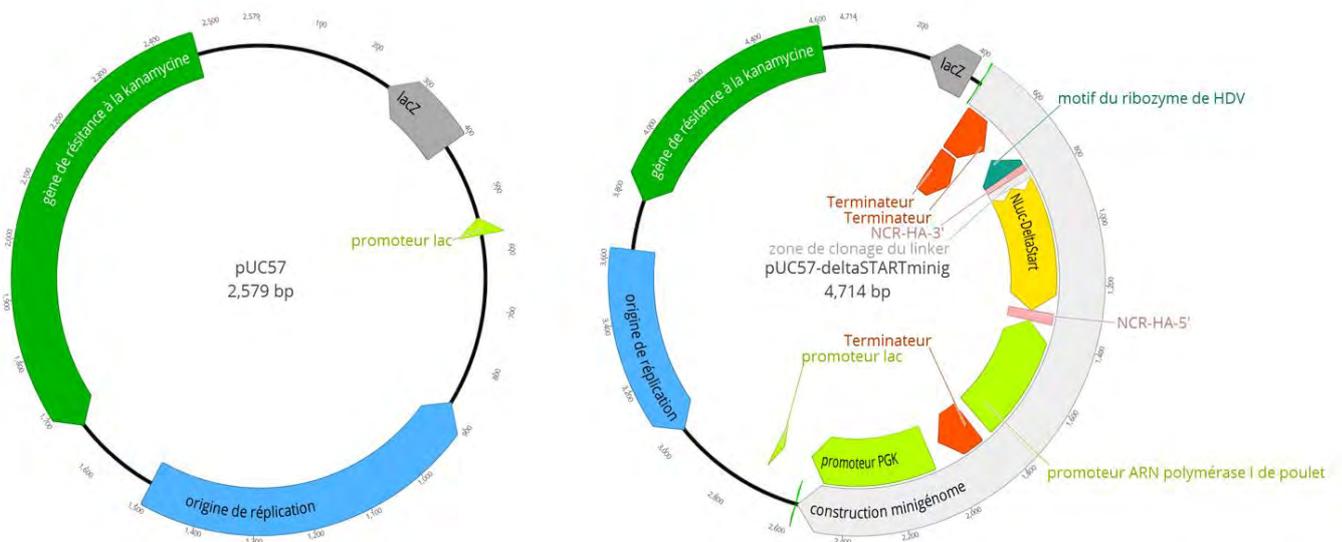


Figure 12 : Cartes des vecteur pUC57 sans et avec l'insert correspondant à la construction minigénome.

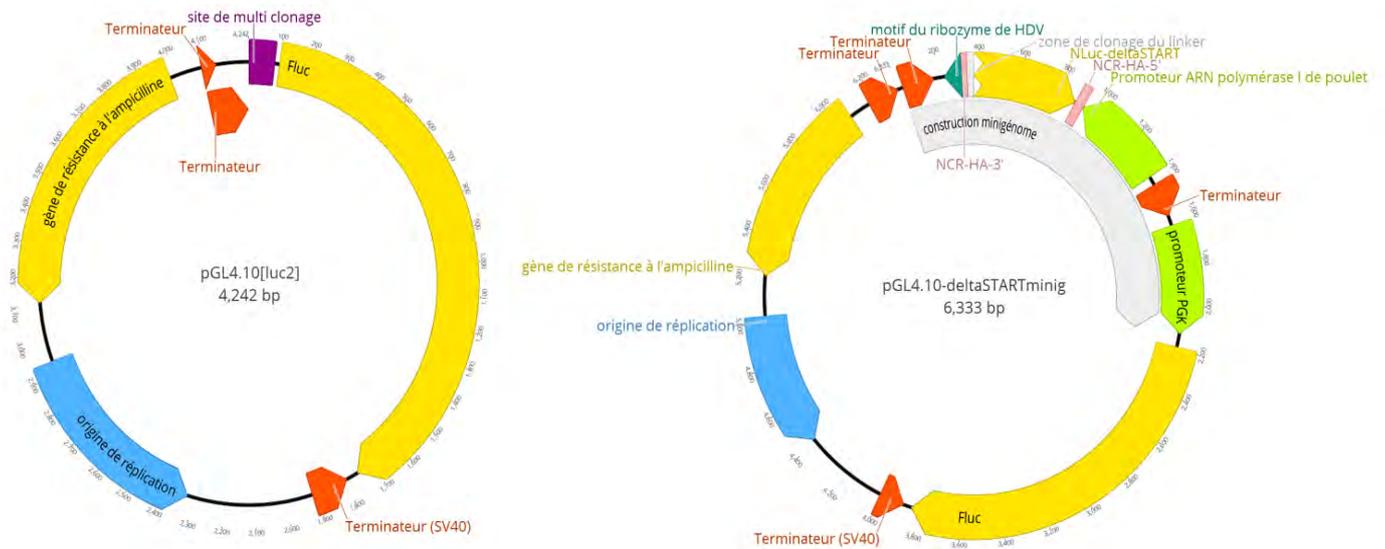


Figure 13 : Cartes des vecteur pGL4.10 sans et avec l'insert correspondant à la construction minigénome.

2.1.4. Linkers

Les séquences des linkers dérivent des séquences nucléotidiques des souches H5N3 ShimH5 et ShimH5_24a2b (Nao N, Yamagishi J et al. 2017) et correspondent à la région +15 / -15 nucléotides autour du site de clivage de la HA (linkerShimH5, **Tableau 2**). Ces linkers sont également composés d'un codon ATG en amont de la région dérivant de la séquence virale. Une zone variante de 3, 2 ou un désoxy-cytidines (C) placée en aval de la séquence virale permet les réalisations respectives des linkers versions V30 (30 nt), V29 (29 nt) et V28 (28 nt). Pour chaque linker, un oligonucléotide sens (ss) et un anti-sens (as) ont été synthétisés séparément pour la formation ultérieure d'un dimère correspondant à un ADN double-brin prêt pour ligation. Les extrémités 5' des oligonucléotides sens et antisens possèdent les séquences coïncidentes aux extrémités du vecteur receveur pGL4.10-minigénome (pGL4.10- Δ STARTminig) après digestion par BsaI, soit respectivement [TCAA] et [AGAC].

Les dimères ont été réalisés par hybridation. Pour cela une première étape de phosphorylation de leurs extrémités 5' a été effectuée sur cent micromolaires de chaque oligonucléotide par action de la phosphonucléotide kinase (Thermo Fisher®) pendant 1h30 à 37°C.

Après inactivation de la phosphonucléotide kinase (20 min à 65°C), cinquante micromolaires de chaque oligonucléotide ont été mélangés pour une concentration finale de 10 μ M et ont subis une première étape de dénaturation 95°C pendant 5 min, suivie d'un refroidissement progressif de 1h30 jusqu'à température ambiante.

Tableau 2 : Linkers dérivés de séquences HA de ShimH5 (en gras : ATG initiateur, en bleu : séquences s'hybridant avec les bouts cohésifs du vecteur pGL4.10-minigénome digéré par BsaI, nucléotides soulignés : régions variantes pour la formation des linkers version V30, V29 et V28).

Nom du linker	sens	Séquence (5' → 3')	Nombre de bases
linkerShimH5_V30	ss	TCAAAATG CCAAAGAGAAACAAGAGGTCTGTTTGGACCCGT	41
	as	AGACACGGGTCCAAACAGACCTCTGTTTCTCTTTGGCATT	41
linkerShimH5_V29	ss	TCAAAATG CCAAAGAGAAACAAGAGGTCTGTTTGGACCCGT	40
	as	AGACACGGGTCCAAACAGACCTCTGTTTCTCTTTGGCATT	40
linkerShimH5_V28	ss	TCAAAATG CCAAAGAGAAACAAGAGGTCTGTTTGGACCGT	39
	as	AGACACGTCCAAACAGACCTCTGTTTCTCTTTGGCATT	39
linkerShimH5_24a2b_V30	ss	TCAAAATG CCAAAGAAAAAAAAAGAGGTCTGTTTGGACCCGT	41
	as	AGACACGGGTCCAAACAGACCTCTTTTTTTCTTTGGCATT	41
linkerShimH5_24a2b_V29	ss	TCAAAATG CCAAAGAAAAAAAAAGAGGTCTGTTTGGACCCGT	40
	as	AGACACGGGTCCAAACAGACCTCTTTTTTTCTTTGGCATT	40
linkerShimH5_24a2b_V28	ss	TCAAAATG CCAAAGAAAAAAAAAGAGGTCTGTTTGGACCGT	39
	as	AGACACGTCCAAACAGACCTCTTTTTTTCTTTGGCATT	39

2.1.5. Obtention des plasmides pGL4.10-minigénome et clonage des séquences linkers

Le minigénome cloné dans le plasmide pUC57 a été transféré dans le plasmide d'expression pGL4.10 (Promega®) (Figure 13). Les plasmides pGL4.10 vide et pUC57-minigénome ont été en parallèle digérés par les enzymes des restriction KpnI et BglII. Le minigénome libéré a été purifié sur gel d'agarose (0.8%, kit NucleoSpin® Gel and PCR Clean-up de Macherey Nagel®) et ligué en pGL4.10 à l'aide de la ligase T4 (Thermo Fisher®) selon les recommandations du fabricant à température ambiante pendant 30 min. Des cellules chimio-compétentes (Stellar™ Competent Cells, génotype : F⁻, endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Φ80d lacZΔ M15, Δ (lacZYA - argF) U169, Δ (mrr - hsdRMS - mcrBC), ΔmcrA, λ⁻) ont été transformées avec les produits de ligation précédents par choc thermique selon le protocole PT5055-2 (Clontech Laboratories, Inc. A Takara Bio Company). Après étalement sur gélose Ampiciline⁺ (50 µg/mL) plusieurs clones ont été amplifiés et purifiés par Minipreps (kit QIAprep Kit, Quiagen®). Les séquences des minigénomes ont ensuite été vérifiées par séquençage (Eurofins Genomics) à l'aide de l'oligonucléotide TFA-167 (CAGGAAACAGCTATGACC). Un clone validé a ensuite été amplifié à l'aide d'une culture bactérienne en milieu LB (Luria Broth) et purifié par Midiprep *endonuclease-free* (NucleoBond® Xtra Midi EF, Macherey-Nagel®) afin de donner le vecteur pGL4.10-ΔSTARTminig. Ce dernier a été digéré par BsaI afin de former des extrémités cohésives compatibles avec les extrémités des linkers en vue de leur ligation. Le plasmide linéarisé a ensuite été purifié sur gel d'agarose (0,8 %) et déphosphorylé au niveau de ses extrémités 5' via l'action de la phosphatase alcaline (SAP, Promega®) selon les recommandations du

fabricant. Les linkers ont ensuite été ligués à l'aide de la ligase T4 (Thermo Fisher®). Des cellules compétentes ont été transformées par les produits de ligation et plusieurs clones ont été isolés, amplifiés puis vérifiés par séquençage (Eurofins Genomics) à l'aide de l'oligonucléotide TFA-167. L'amplification bactérienne en milieu LB et la purification des clones validés par séquençage ont ainsi permis l'obtention des constructions :

- pGL4.10- Δ STARTminig-linkerShimH5_V30, V29 et _V28
- pGL4.10- Δ STARTminig-linkerShimH5_24a2b_V30, V29 et _V28

2.1.6. Construction du système rapporteur en pcDNA

Les constructions minigénomes ont été amplifiées par PCR à partir des vecteurs pGL4.10- Δ STARTminig-linkerShimH5_V30, V29 et _V28 et pGL4.10- Δ STARTminig-linker2ShimH5_24a2b_V29 et _V28. Pour cela, l'ADN polymérase de haute-fidélité Platinum™ SuperFi™ (Invitrogen) a été utilisée en association avec les amorces sens TFA-207 (ATCAGGTACCCAGCAGGGGTTCACTCTGTCA, site KpnI souligné) et antisens TFA-176 (GCTTGGATCCTCACAATTTACGCCAGAATGCGT, site BamHI souligné) s'hybridant respectivement au niveau des séquences NCR 3' et 5'. Les amplicons ont été digérés par les enzymes de restriction KpnI et BamHI, purifiés sur gel d'agarose (0,8 %) puis ligués en pCDNA3.1(+) (cf protocole de ligation ci-dessus), préalablement digéré par KpnI et BamHI.

2.1.7. Cellules

Les cellules de la lignée C-DF1 (UMNSAH/DF-1, ATCC® CRL-12203™), des fibroblastes immortalisés issues de tissus embryonnaires de poulet, ont été cultivés dans du milieu *Dulbecco's Modified Eagle's* (DMEM) avec 10% de sérum de veau fœtal (SVF), 100 U/ml de pénicilline, 0,1 mg/ml de streptomycine, et mises en incubation à 37°C avec 5 % de CO₂

2.1.8. Essais luciférases

Des cellules C-DF1 cultivées en plaques 24 puits ont été co-transfectées avec le système minigénome et les plasmides codant pour les sous-unités du complexe polymérase virale, selon le protocole du kit Lipofectamine® LTX & PLUS™ Reagent (Invitrogen™). Plusieurs quantités d'ADN ont été testées (données pour 1x10⁶ cellules transfectées) : 2 µg - 0,1 µg pour les constructions minigénomes, 2 µg pour le plasmide exprimant la GFP (pMAX-GFP) et 500

ng pour les plasmides permettant l'expression de chaque sous-unité des complexes polymérase viraux. Vingt-quatre heures après transfection, les cellules ont été lysées avec l'utilisation du tampon *Passive Lysis Buffer* (Promega®). Après 15 min de lyse, 40 µL de chaque échantillon ont été déposés sur une plaque 96 puits (Greiner bio-one).

Les lectures des activités enzymatiques des luciférase Nluc et Fluc ont été réalisées respectivement avec les kits Nano-Glo et One-Glo de chez Promega® sur un appareil Clariostar (BMG Labtech ; paramètres optiques : filtre 470-480 nm pour les lectures Nluc, absence de filtre pour les lectures Fluc, temps d'acquisition : 10 s). Le signal Nluc a été normalisé par le signal Fluc et les ratios Nluc/Fluc ont été calculés afin de donner les valeurs absolues normalisées du signal pour chaque puit. Les valeurs relatives ont ensuite été calculées en plaçant à 100% le signal correspondant aux conditions expérimentales avec les constructions pGL4.10-linkerShimH5_V30-Nluc et pcDNA3.1(+)-linkerShimH5_V30-Nluc.

Les analyses statistiques et les graphiques ont été générés avec le logiciel GraphPad Prism 7. La moyenne des valeurs absolues de chaque groupe et leurs écarts type ont été calculés. Les significativités des résultats ont été déterminées à l'aide de tests de Student (risque α placé à 0,05).

2.2. Résultats

2.2.1. Identification des émergences HPAIV

Une bibliothèque de séquences nucléotidiques HA H5 a été constituée à partir des informations disponibles sur la banque de données FludB. Seules les séquences HA H5 issues d'hôtes aviaires non mutantes ont été récupérées (4612 séquences). Afin d'étudier l'événement même de l'émergence, nous avons volontairement écarté de l'analyse les 3511 séquences dérivant de la souche H5N1 HPAIV apparue en Chine en 1996 (A/Goose/Guangdong/1/1996). En effet, intégrer ces séquences HA surreprésenterait le nombre de souches HPAIV, correspondant aux descendants d'une seule émergence et pourrait constituer un biais dans notre étude. *In fine*, 1101 séquences HA H5 ont été analysées. D'après l'alignement des séquences et les données issues de la littérature ((Abdelwhab, Veits, and Mettenleiter 2013), http://www.offlu.net/fileadmin/home/en/resource-centre/pdf/Influenza_A_Cleavage_Sites.pdf) les émergences HPAIV ont pu être identifiées (Tableau 3). Deux grands types de phénomènes permettent d'expliquer les événements d'émergences HPAIV par apparition d'un site de clivage polybasique de la HA : les insertions de nucléotides et les substitutions.

Tableau 3 : Liste des différentes émergences HPAIV (H5) caractérisées sur la base de la séquence nucléotidique HA et des données présentes dans la littérature (les émergences par insertions sont indiquées en rouge).

Evènement d'émergence (n°)	Identification de la souche HPAIV	Type d'évolution de séquence correspondant au site de clivage de la HA
1	A/duck/Ireland/113/1983 (H5N8)	Insertions
2	A/turkey/England/50-92/1991 (H5N1)	Insertions
3	A/Goose/Guangdong/1/1996 (H5N1)	Insertions
4	A/chicken/Italy/8/1998 (H5N2)	Insertions
5	A/ostrich/South Africa/N227/2004 (H5N2)	Insertions
6	A/chicken/France/150169a/2015 (H5N1)	Substitutions
7	A/duck/France/150236/2015 (H5N9)	Substitutions
8	A/chicken/France/160013g/2016 (H5N2)	Substitutions
9	A/emu/Texas/39442/1993 (H5)	Insertions
10	A/turkey/Ontario/7732/1966 (H5N9)	Substitutions
11	A/chicken/Pennsylvania/1/1983 (H5N2)	Substitutions (N-deglycosylation)
12	A/chicken/Puebla/8623-607/1994 (H5N2)	Insertions
13	A/chicken/Queretaro/14588-19/1995 (H5N2)	Insertions
14	A/chicken/Taiwan/CH-A1029/2010 (H5N2)	Substitutions
15	A/chicken/Taiwan/A703-1/2008 (H5N2)	Substitutions
16	A/chicken/Taiwan/7350/2013 (H5N2)	Substitutions
17	A/chicken/Taiwan/A1997/2012 (H5N2)	Substitutions
18	A/chicken/Mexico/1433-5/2005 (H5N2)	Substitutions
19	A/chicken/Hidalgo/7637-05/2005 (H5N2)	Substitutions
20	A/chicken/Michoacan/4240-06/2006 (H5N2)	Substitutions
21	A/chicken/Texas/298313-2/2004 (H5N2)	Substitutions

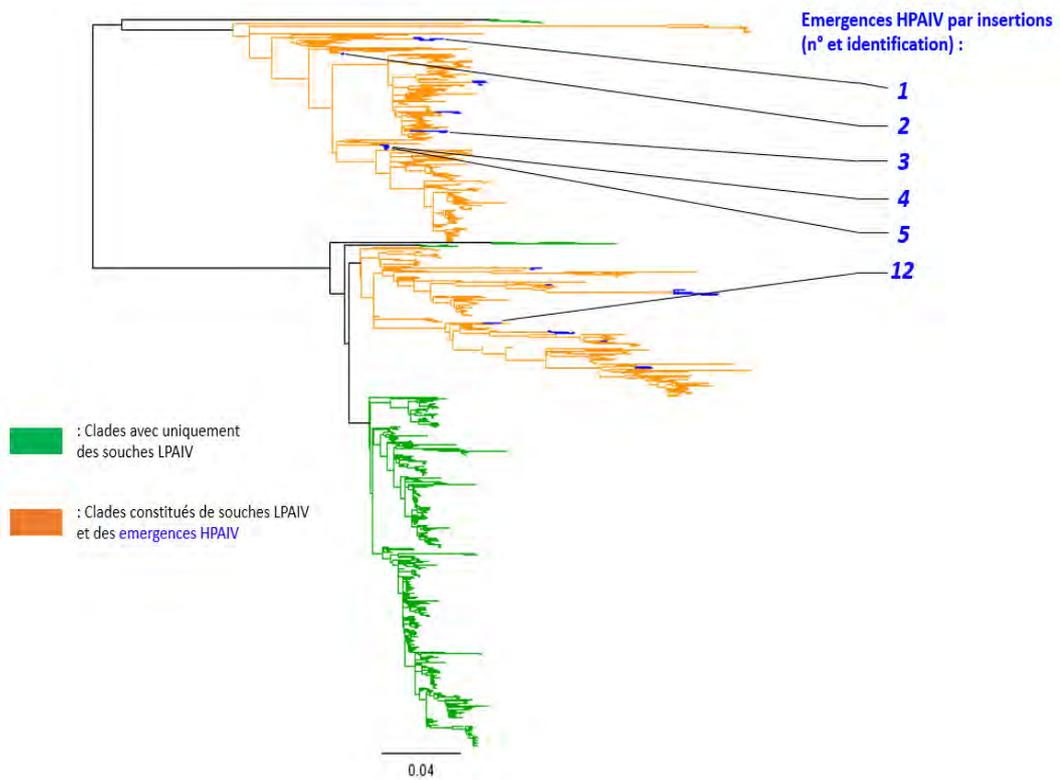
Nos premiers objectifs étaient de répondre aux questions suivantes : Les phénomènes d'émergence sont-ils rassemblés phylogénétiquement sur des mêmes clades ? Inversement, existe-t-il des clades constitués uniquement par des séquences HA de souches LPAIV, parmi lesquels aucun événement d'émergence n'a été décrit ?

Afin de répondre à ces questions, les séquences HA des souches correspondant aux événements d'émergence ont été annotées et un arbre phylogénétique du segment 4 a été construit avec notre bibliothèque complète de séquences HA H5 (**Figure 14**). D'après nos données nous pouvons remarquer que les émergences par insertions et substitutions se dispersent sur différents clades. Il semblerait donc qu'il n'y ait pas de rassemblement des émergences sur certains clusters particuliers, que ce soit pour les phénomènes d'insertions ou de mutations. En revanche, on peut distinguer des clades dépourvus de toute émergence et uniquement constitués de séquences de souches LPAIV. Il serait intéressant de pouvoir déterminer si ces séquences HA de LPAIV sont vraiment défavorables au phénomène d'émergence.

Dans la suite de l'étude, nous nous sommes focalisés uniquement sur les émergences par insertion car le système expérimental minigénome utilisé par la suite ne permet d'analyser que les phénomènes d'insertion.

Pour détecter la présence éventuelle de déterminants communs entre les séquences nucléotidiques HA de LPAIV progéniteurs de HPAIV, nous avons cherché à déterminer la séquence nucléotidique HA de l'ancêtre commun le plus proche de chaque émergence par insertions. Le LPAIV partageant le plus d'identité, en termes de séquence nucléotidique HA, pour chaque événement d'émergence indépendant (par insertions) a été identifié grâce aux données de la littérature et notre arbre phylogénétique (**Figure 14** et **Tableau 4**).

A.



B.



Figure 14 : (A) Arbre phylogénétique total des séquences nucléotidiques HA H5 (modèle GTR+G+I). (B) Un grossissement sur les principales émergences par insertions a été représenté (le nom des souches HPAIV sont indiqués en bleu et annotés avec leur n° « d'événement d'émergence » correspondant, répertoriés dans le **Tableau 3**).

L'émergence HPAIV H5N2 ayant eu lieu au Mexique en 1994 (A/chicken/Puebla/ 8623-607/1994) a bien été décrite dans la littérature et est retrouvée sur notre arbre (García *et al.*, 1996 ; Perdue *et al.*, 1997). Après une circulation intense de souches LPAIV chez des poulets, des premiers symptômes modérés sont apparus chez ces animaux en 1994, avant d'acquérir un caractère fortement létal début 1995. Un motif polybasique du site de clivage a été identifié chez la souche de pouvoir pathogène modéré et chez la souche hautement pathogène. Un mécanisme d'insertion de nucléotides par glissement de la polymérase virale et duplication de codon a été suggéré (Abdelwhab, Veits, and Mettenleiter 2013; Garcia et al. 1996; Perdue et al. 1997). La souche LPAIV progénitrice serait vraisemblablement l'ancêtre commun partagée avec le LPAIV A/chicken/Mexico/31381-7/1994 (H5N2) et la souche émergente en question.

Tableau 4 : Emergences indépendantes par insertions et leurs souches LPAIV partageant le plus d'identité avec l'ancêtre LPAIV commun entre eux, sur la base de la séquence nucléotidique HA.

Evènement d'émergence (n°)	Souche HPAIV par insertions	Identification de la souche LPAIV la plus proche de l'ancêtre commun partagée avec la souche HPAIV
1	A/duck/Ireland/113/1983 (H5N8)	A/turkey/Italy/1980 (H5N2)
2	A/turkey/England/50-92/1991 (H5N1)	A/finch/England/2051/1991 (H5N2)
3	A/Goose/Guangdong/1/1996 (H5N1)	A/Anas platyrhynchos/Altai/1285/1991 (H5N3)
4	A/chicken/Italy/8/1998 (H5N2)	A/mallard/Italy/208/2000 (H5N3)
5	A/ostrich/South Africa/N227/2004 (H5N2)	A/mallard/Bavaria/1/2005 (H5N2)
12	A/chicken/Puebla/8623-607/1994 (H5N2)	A/chicken/Mexico/31381-7/1994 (H5N2)

Les événements d'émergence HPAIV H5 indépendants ont pu être retrouvés et positionnés dans un arbre phylogénétique, sur la base de la séquence nucléotidique de HA. Grâce à cet arbre, la souche LPAIV possédant le plus d'identité (sur la base de la séquence nucléotidiques HA) avec l'ancêtre commun de chaque émergence par insertion a été identifié (taxon sœur le plus proche de chaque émergence HPAIV) (Tableau 4).

2.2.2. Etude des prédictions de structures d'ARN

Une prédisposition génétique des souches LPAIV des sous-type H5 et H7 à l'émergence par phénomène d'insertion a été proposé par Nao et ses collègues. A partir d'une banque de séquences nucléotidiques HA issue de LPAIV (isolées chez le canard), ils ont montré qu'il existe une structure conservée de type épingle à cheveu située au voisinage du site de clivage de la HA. Cette structure semble être plus conservée au sein des sous-types H5 et H7 (Nao N,

Yamagishi J et al. 2017). Néanmoins, leur analyse a été réalisée à partir d'une banque de données mélangeant les séquences HA génétiquement proches et éloignées des émergences. Nous avons donc analysé l'environnement nucléotidique du site de clivage de la séquence HA en tenant compte de l'histoire évolutive et de la proximité phylogénétique avec les événements d'émergence, afin de répondre à la question suivante : existe-il un point commun entre toutes les souches LPAIV responsables d'émergences HPAIV par insertion, notamment en termes de structures secondaires d'ARN prédites au niveau de la région correspondant au site de clivage de la HA ? Des prédictions de structures secondaires d'ARN ont été réalisées au niveau de la région correspondant au site de clivage des séquences HA de LPAIV considérées comme précurseurs au phénomène d'émergence. Deux fenêtres d'analyse ont été utilisées lors des prédictions de structures : 15 et 30 nucléotides (nt) de part et d'autre du site de clivage.

Toutes les séquences des souches LPAIV potentiellement précurseurs de HPAIV possèdent des structures secondaires d'ARNc similaires de type épingle à cheveux (**Tableau 5**), décrites comme favorables au phénomène d'insertion par *polymerase slippage* (Abdelwhab, Veits, and Mettenleiter 2013; Nao N, Yamagishi J et al. 2017). Les séquences HA des LPAIV A/chicken/Mexico/31381-7/1994 (H5N2) et A/finch/England/2051/1991 (H5N2) partagent le plus d'identité avec les deux émergences d'HPAIV ayant eu lieu au Mexique (A/chicken/Puebla/8623-607/1994 (H5N2)) et en Angleterre (A/turkey/England/50-92/1991 (H5N1)) respectivement. Les ARNc de ces deux séquences HA de LPAIV possèdent chacune une structure secondaire prédite de type épingle à cheveux, flanquant la région codant le site de clivage de la HA. La boucle principale de la structure en épingle est constituée de 11 nt libres. Une boucle interne, plus petite, est aussi observée pour chacune de ces séquences.

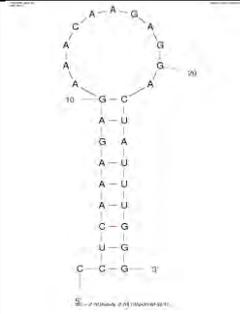
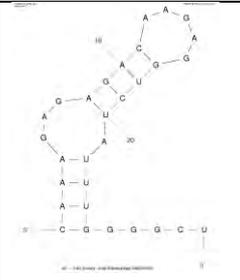
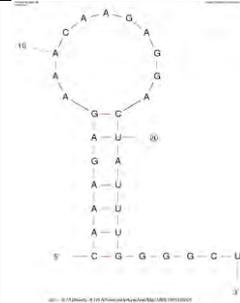
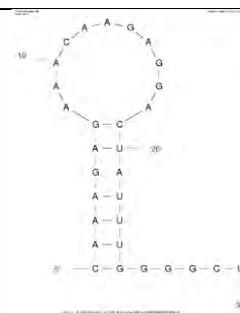
Inversement, les prédictions faites pour les séquences ARNc de HA des LPAIV isolées phylogénétiquement de toutes émergences connues, comme celles des souches A/shearwater/Australia/751/1975 (H5N3) et A/duck/Taiwan/4360/2013 (H5N2) décrivent des structures secondaires d'ARN très différentes en comparaison avec les souches mexicaines et anglaises précédentes (**Tableau 6**). Des structures en épingle à cheveux sont retrouvées, mais la boucle principale est beaucoup moins importante en termes de taille. D'autres séquences HA de souches LPAIV isolées d'émergences connues décrivent des structures d'ARN sans épingle à cheveux flanquant la région correspondant au site de clivage de la HA.

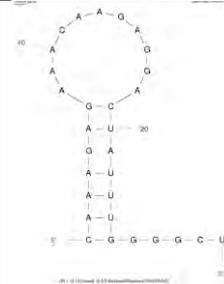
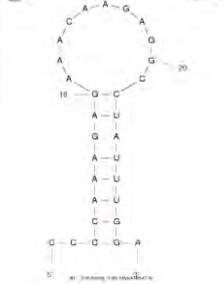
Un score de 0 à 2 a été attribué pour chaque souche de LPAIV en fonction des structures d'ARN prédites au niveau de la séquence HA. Un point est attribué par critère si ce dernier est vrai :

- i. Critère 1 : Présence d'une structure de type épingle à cheveu flanquant la séquence correspondant au site de clivage de la HA

- ii. Critère 2 : La boucle principale de la structure épingle à cheveu est composée de plus de 9 nt non appariés

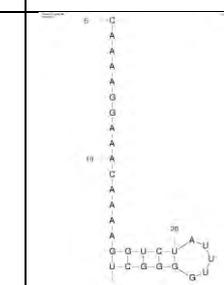
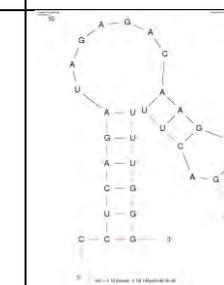
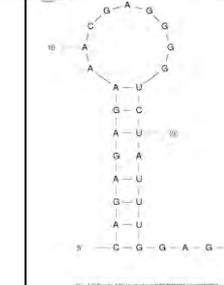
Tableau 5 : Prédiction des structures secondaires d'ARNc des séquences LPAIV proches des émergences HPAIV par insertions (données concernant la fenêtre d'analyse de 15 nt en amont et en aval de la région correspondant au site de clivage sont représentées).

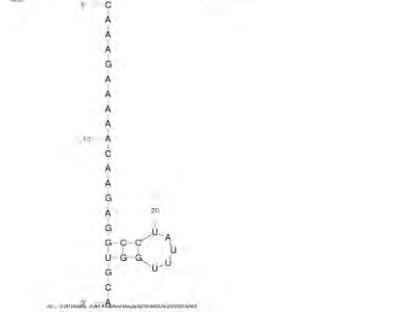
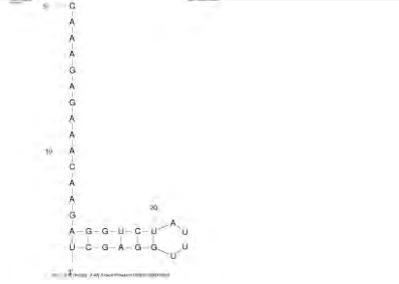
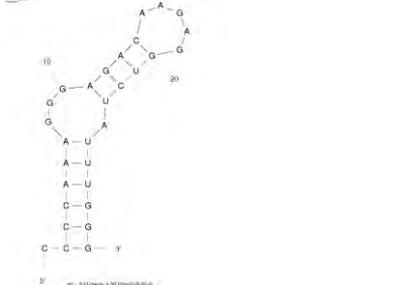
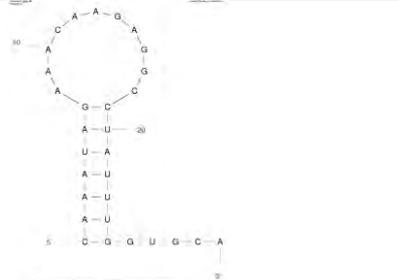
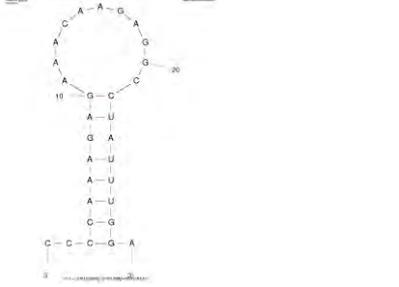
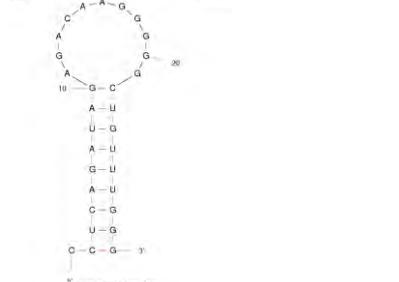
Evènement d'émergence (n°)	Souche LPAIV et sa séquence ARNc comprenant la région correspondant au site de clivage de la HA	Structure secondaire d'ARNc prédite	Score S
1	A/finch/England/2051/1991 (H5N2) CAA AGA GAA ACA AGA GGA CUA UUU GGG GCU		2
2	A/turkey/Italy/1980 (H5N2) CAA AGA GAG ACA AGA GGU CUA UUU GGG GCU		1
3	A/Anas platyrhynchos/Altai/1285/1991 (H5N3) CAA AGA GAA ACA AGA GGA CUA UUU GGG GCU		2
4	A/mallard/Italy/208/2000 (H5N3) CAA AGA GAA ACA AGA GGA CUA UUU GGG GCU		2

5	A/mallard/Bavaria/1/2005 (H5N2) CAA AGA GAA ACA AGA GGA CUA UUU GGG GCU		2
12	A/chicken/Mexico/31381-7/1994 (H5N2) CAA AGA GAA ACA AGA GGC CUA UUU GGA GCA		2

- Les nucléotides (nt) représentés **en gras** codent en protéines pour les acides aminés GLF constants chez tous les segments HA H5 en aval du site du clivage de la HA.

Tableau 6 : Prédiction des structures secondaires d'ARNc des séquences LPAIV isolées phylogénétiquement de toutes émergences HPAIV (données concernant la fenêtre d'analyse de 15 nt en amont et en aval de la région correspondant au site de clivage sont représentées).

Souche LPAIV et sa séquence ARNc comprenant la région correspondant au site de clivage de la HA	Structure secondaire d'ARNc prédite	Score S
A/duck/Tsukuba/168/2005 (H5N2) CAA AAG GAA ACA AAA GGU CUA UUU GGG GCU		0
A/duck/Taiwan/4360/2013 (H5N2) CAG AUA GAG ACA AGG GGA CUU UUU GGG GCA		1
A/mallard duck/NZ/NZ0700316-53c/2007 (H5N2) CAG AGA GAA ACG AGG GGU CUA UUU GGA GCA		1

<p>A/mallard/Maryland/16-042534-2/2016 (H5N2)</p> <p>CAA AGA AAA ACA AGA GGC CUA UUU GGU GCA</p>		0
<p>A/duck/Potsdam/1402-6/1986 (H5N2)</p> <p>CAA AGA GAA ACA AGA GGU CUA UUU GGA GCU</p>		0
<p>A/shearwater/Australia/751/1975 (H5N3)</p> <p>CAA AGG GAG ACA AGA GGU CUA UUU GGG GCU</p>		1
<p><i>*A/thick-billed_murre/Greenland/9045-2K/2014 (H5N1)</i></p> <p>CAA AUA GAA ACA AGA GGC CUA UUU GGU GCA</p>		2
<p><i>*A/mallard/Maryland/792/2002 (H5N2)</i></p> <p>CAA AGA GAA ACA AGA GGC CUA UUU GGA GCA</p>		2
<p><i>*A/duck/Taiwan/DV518/2006 (H5N2)</i></p> <p>CAG AUA GAG ACA AGG GGG CUG UUU GGG GCA</p>		2

- Les nucléotides (nt) représentés **en gras** codent en protéines pour la séquence GLF constantes chez tous les segments HA H5, en aval du site du clivage de la HA. * : LPAIV non décrits comme étant des précurseurs potentiels de HPAIV (isolées de toutes émergences) mais possédant des structures secondaires d'ARN prédites comme favorables aux insertions

De manière intéressante, certaines souches LPAIV comme *A/duck/Taiwan/DV518/2006* (H5N2) et *A/mallard/Maryland/792/2002* (H5N2), non décrites comme étant des précurseurs potentiels de HPAIV (isolées de toutes émergences) possèdent aussi des structures secondaires d'ARN prédites comme favorables aux insertions (**Tableau 6**). Des structures conservées, identiques aux souches mexicaines et anglaises précédentes, sont aussi retrouvées. La capacité de ces dernières séquences de HA (issues des souches LPAIV suivantes : *A/thick-billed_murre/Greenland/9045-2K/2014* (H5N1), *A/mallard/Maryland/792/2002* (H5N2) et *A/duck/Taiwan/DV518/2006* (H5N2)) à être favorables au phénomène d'émergence reste une question ouverte. Des prédictions de structures secondaires d'ARN ont été aussi réalisées sur l'ARNv. Des résultats similaires à ceux observés avec l'ARNc des séquences étudiées ont été retrouvés (résultats non montrés). Par ailleurs, des résultats équivalents ont été observés avec l'utilisation d'une plus grande fenêtre d'analyse (30 nt en amont et en aval du site de clivage).

Des structures secondaires d'ARN de type épingle à cheveux situées au voisinage de la séquence correspondant au site de clivage de la HA semblent être conservées parmi les LPAIV progéniteurs d'émergences. D'après les données phylogénétiques et de prédictions des structures d'ARN, trois groupes de séquences nucléotidiques de HA provenant de LPAIV H5 ont pu être déterminés :

- i. Séquences HA de LPAIV **proches** phylogénétiquement d'émergences avec des structures secondaires d'ARN considérées comme potentiellement **favorables** aux insertions
- ii. Séquences HA de LPAIV **isolées** phylogénétiquement d'émergences avec des structures secondaires d'ARN considérées comme potentiellement **non favorables** aux insertions
- iii. Séquences HA de LPAIV **isolées** phylogénétiquement d'émergences avec des structures secondaires d'ARN considérées comme potentiellement **favorables** aux insertions.

Les fragments de séquences nucléotidiques de HA des souches LPAIV étudiées (correspondant au cadre de 30 nt en amont et en aval du site de clivage de la HA), ainsi que la totalité des séquences HA de LPAIV H5 ont été alignés (**Figure 15**). Pour la suite du projet, deux souches LPAIV ont été sélectionnées pour chaque groupe de séquence HA constituées (cf données

phylogénétiques et de prédictions de structures ci-dessus) comme étant représentatives de la variabilité nucléotidique observée au niveau de la région d'intérêt (15 nt en amont et en aval du site de clivage de la HA). Les séquences candidates représentant le groupe de séquences LPAIV proche d'émergence dont les structures prédites sont favorables aux insertions sont les séquences HA de A/chicken/Mexico/31381-7/1994 (H5N2) et A/finch/England/2051/1991 (H5N2). Concernant le groupe de LPAIV isolés d'émergences connues et dont les structures d'ARN sont non-favorables, les séquences HA issues de A/shearwater/Australia/751/1975 (H5N3) et A/duckF/Taiwan/4360/2013 (H5N2) ont été sélectionnées. Enfin, les séquences HA des souches A/duck/Taiwan/DV518/2006 (H5N2) et A/mallard/Maryland/792/2002 (H5N2) ont été retenues afin de pouvoir vérifier leurs potentiels favorisant ou non aux insertions, et ce, malgré l'absence de parenté avec une souche émergente d'HPAIV.

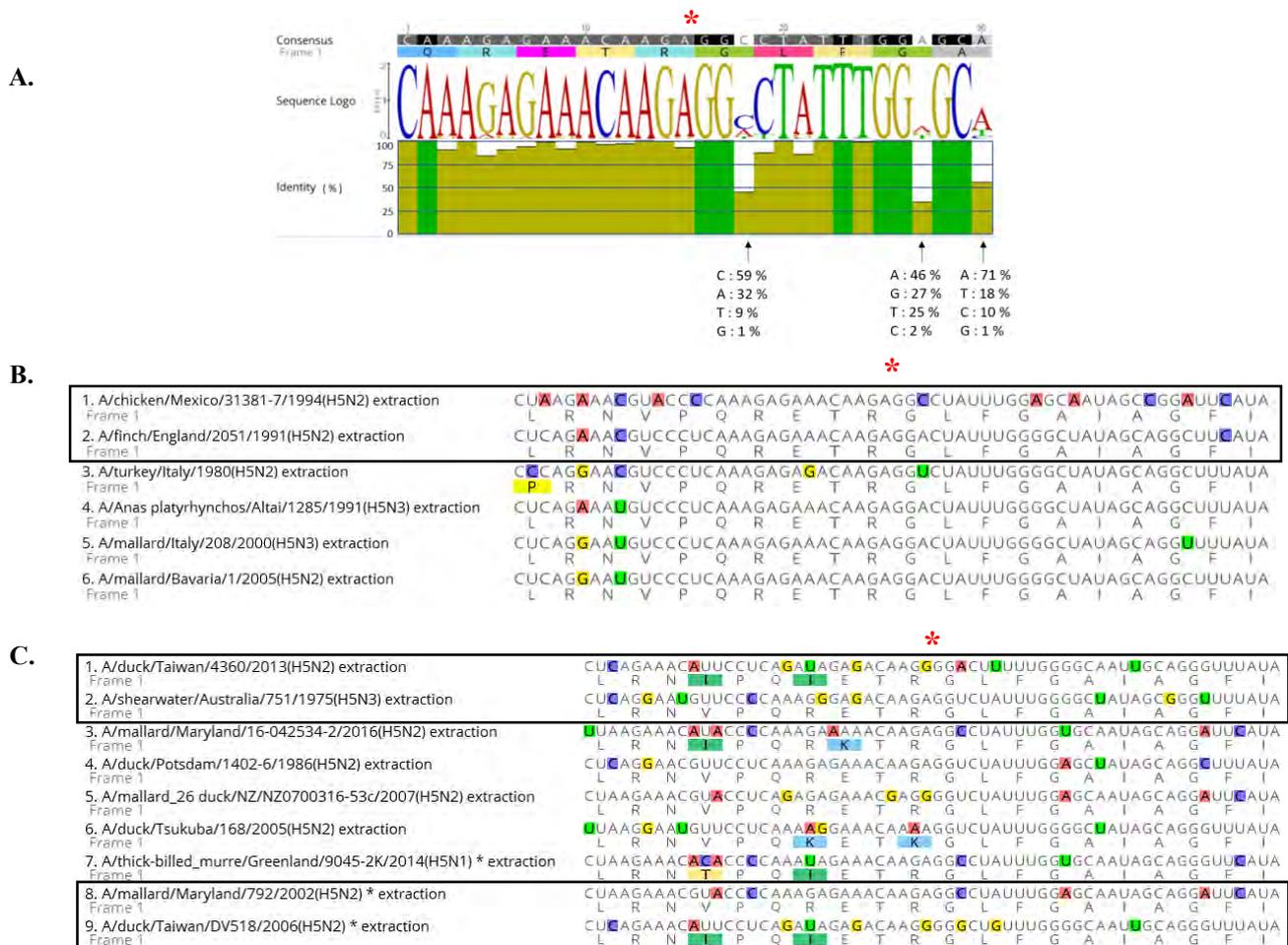


Figure 15 : Alignement des séquences nucléotidiques (ADNc) HA H5 des LPAIV correspondant à la région du site de clivage (indiquée par le symbole *, au-dessus des séquences consensus) de la HA des LPAIV étudiés. La séquence logo de l'alignement de toutes les séquences HA H5 de LPAIV avec le pourcentage d'identité et la fréquence de chaque nt au niveau des positions les plus variantes a été représentée (A). L'alignement des séquences HA (ARNc) génétiquement proches (B.) et éloignées (C.) d'émergences HPAIV par insertions sont aussi représentées. Les ambiguïtés sont surlignées en couleur. La traduction en protéines est indiquée en dessous de chaque séquence nucléotidique. Les séquences HA (et le nom de la souche correspondant) qui ont été sélectionnées comme représentatives des différents groupes de séquences HA de LPAIV H5 sont encadrées.

2.2.3. Construction et optimisation du système minigénome

Afin d'analyser si les différences de structure ARN prédites corrélaient avec une probabilité d'insertion, nous avons élaboré un système expérimental permettant de mesurer la probabilité d'insertions dans plusieurs fragments de séquences nucléotidiques de HA. Pour cela un système minigénome, similaire à celui créé par Nao et ses collègues, a été élaboré (Figure 16). Il consiste en un plasmide dans lequel un fragment de séquence nucléotidique de HA (appelé linker), dérivé d'une souche LPAIV, a été cloné en amont d'un gène rapporteur Nluc. Celui-ci est délégué de son ATG initiateur qui est fourni par le linker après sa ligation. Plusieurs tailles de linker sont testées. Dans le cas où les linkers correspondant à la séquence HA clonés dans le minigénome mesurent 29 et 28 nt, la Nluc ne pourra s'exprimer que si, respectivement, une ou deux insertions de nucléotides ont lieu, rétablissant le cadre de lecture du gène rapporteur. La construction [linker-Nluc(Δ ATG)] est flanquée par les NCR des séquences HA afin de permettre la production d'ARNm polymérase virale-dépendante après la synthèse d'un ARN pseudo-viral (-) AIV par l'action de la polymérase I. L'enzyme de restriction BsaI a été utilisée pour le clonage des linkers. En effet, ce type d'enzyme de restriction reconnaît des séquences ADN asymétriques et clive à l'extérieur de son site de reconnaissance, permettant des clonages non cicatriciels.

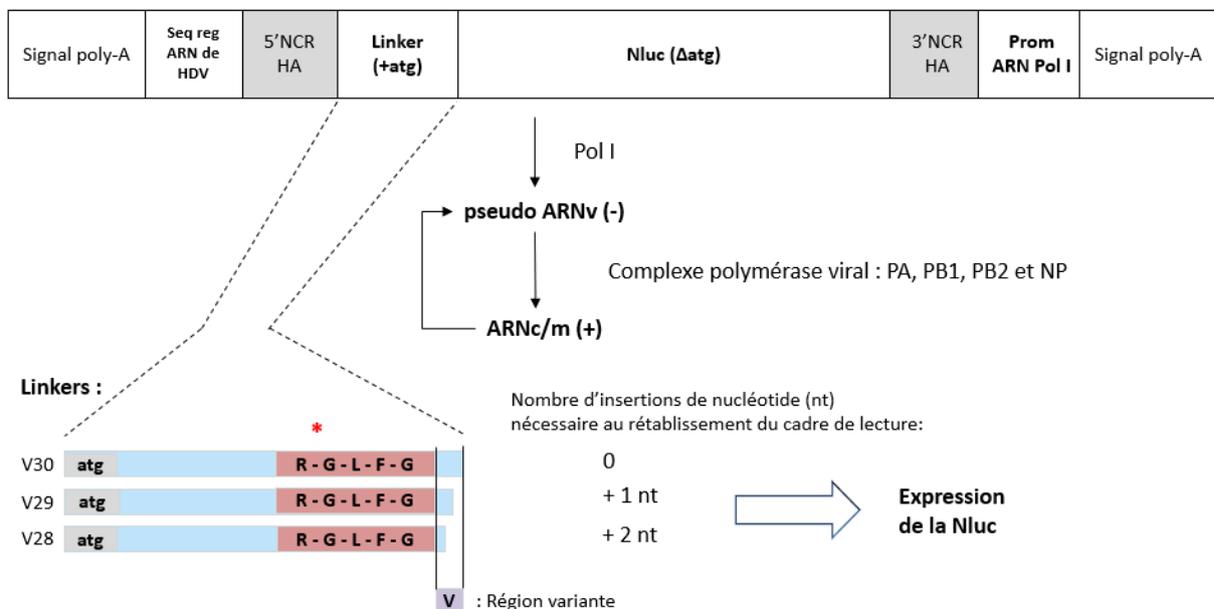


Figure 16 : Schéma du système minigénome rapporteur des phénomènes d'insertions. (Légende : Le symbole * indique le site de clivage de la HA. Les acides aminés correspondant à la séquence protéique environnant le site de clivage sont représentés en gras et en lettres capitales. Les nucléotides sont représentés en lettre minuscules).

Tout d'abord un système minigénome (voir Matériels et Méthodes) en pUC57 a été construit. Des cellules d'une lignée de fibroblastes de poulets ont été transfectées par ces constructions et testées en essai luciférase (**Figure 17**). Une première série d'expériences a été réalisée afin de déterminer les quantités d'ADN à utiliser concernant la construction minigénome en pUC57 et celles des différents plasmides d'expression (pcDNA3.1(+)) codant pour les sous-unités du complexe polymérase virale WSN (A/WSN/H1N1/1933). La construction pUC57- Δ STARTminig (sans l'ATG au niveau de la région de clonage des linkers) en présence du complexe polymérase WSN complet a aussi été testée afin de déterminer le bruit de fond apporté par le système. De plus chaque construction a été évaluée en présence du complexe de polymérase incomplet (-NP) afin de déterminer la spécificité de notre signal. Deux et 0,1 μ g de la construction pUC57-minig (avec l'ATG initiateur dans la zone de clonage du linker) ou pUC57- Δ STARTminig (sans l'ATG) ont été initialement utilisés pour 1×10^6 cellules avec 500 ng de chaque plasmide codant pour PA, PB1, PB2 et NP (WSN) en pcDNA3.1(+).

Concernant la condition pUC57-minig (+NP), on observe un signal Nluc qui est 180x plus important par rapport à la construction pUC57- Δ STARTminig (+NP), pour 2 μ g des constructions relatives. En diminuant la quantité de chaque construction à 0,1 μ g (tout en préservant la même quantité de plasmides codant pour le complexe polymérase virale WSN), le signal Nluc est augmenté de près de 10x pour la condition pUC57-minig (+NP) et devient 2x plus important concernant la construction pUC57- Δ STARTminig (+NP). Par ailleurs, le rapport du signal Nluc entre pUC57-STARTminig(+NP) / pUC57- Δ STARTminig(+NP) augmente de près de 2x (ratio des signaux luciférase passant de 380 à 780 avec l'utilisation de 2 et 0,1 μ g de construction en pUC57, respectivement). Le système devient donc légèrement plus sensible et spécifique lorsqu'une quantité plus faible de construction minigénome est apportée. En absence de NP, le signal est diminué d'environ 10x par rapport à la condition pUC57- Δ STARTminig pour les deux quantités d'ADN testées.

Compte tenu de l'optimisation du signal Nluc, la quantité de 0,1 μ g de construction minigénome pour 1×10^6 cellules par puits a donc été choisie pour la suite des expériences.

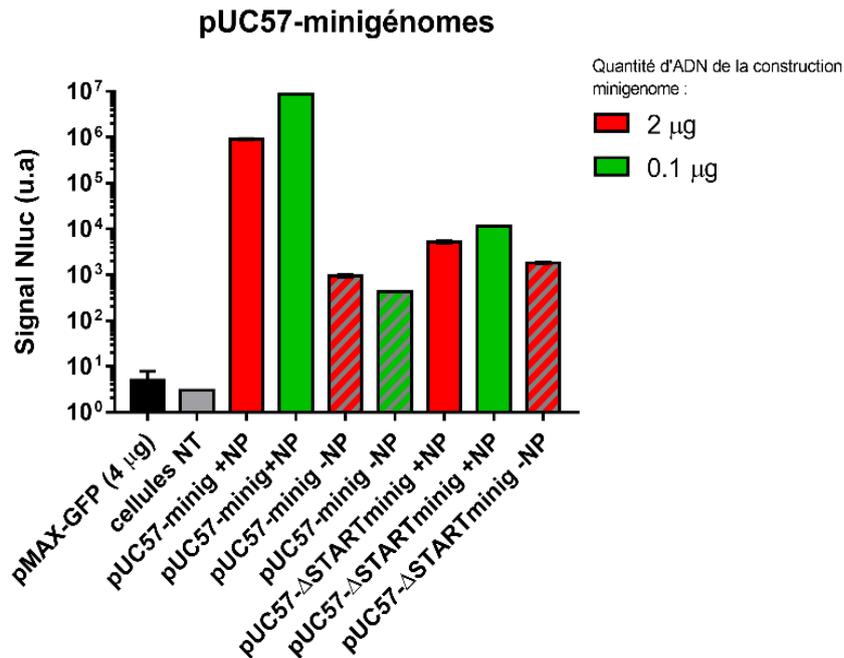


Figure 17 : Résultats des essais luciférase après la transfection de cellules CDF1 avec les constructions minigénomes en pUC57 et les plasmides pcDNA3.1(+)-PA, PB1, PB2 et NP (WSN). Les conditions avec le complexe de polymérase incomplet (sans apport de pcDNA3.1(+)-NP, motifs avec les lignes transversales) ont été réalisées. Deux expériences indépendantes ont été réalisées. Les moyennes des valeurs brutes du signal Nluc ainsi que les écarts-types de chaque groupe sont représentés.

Du fait de la présence d'un bruit de fond important avec pUC57-ΔSTARTminig, nous avons décidé de transférer notre construction minigénome dans un nouveau plasmide pGL4.10. Ce dernier possède plusieurs terminateurs permettant d'éviter d'éventuelles transcriptions alternatives de la Nluc, non spécifiques de la polymérase virale (Figure 13). De plus, un gène codant la Fluc est présent sur ce plasmide, ce qui permet de normaliser le signal du rapporteur Nluc (voir Matériels et méthodes).

Des activités luciférase similaires ont été retrouvées avec les constructions minigénomes en pGL4.10 (Figure 18). Un ratio des activités luciférase pGL4.10-minig (+NP) / pGL4.10-ΔSTARTminig (+NP) approche les 400x, ce qui équivaut à ce que l'on a pu observer précédemment en système pUC57.

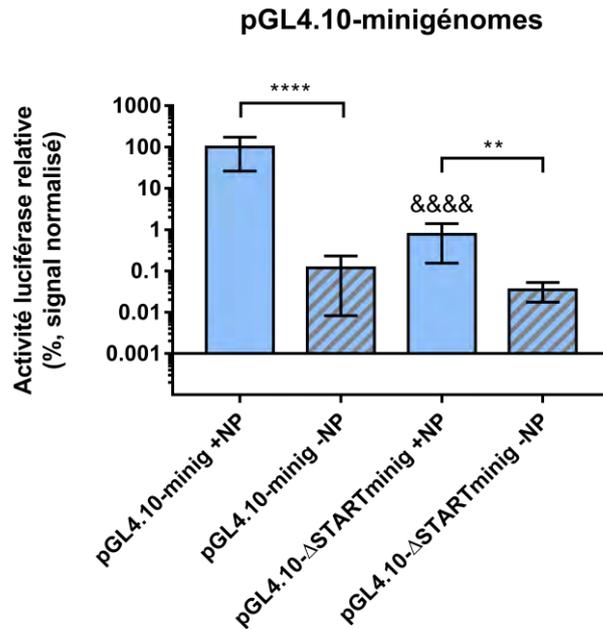


Figure 18 : Résultats des essais luciférase après la transfection de cellules CDF1 avec les constructions minigénomes en pGL4.10 et les plasmides pcDNA3.1(+)-PA, PB1, PB2 et NP (WSN). Les conditions avec le complexe de polymérase incomplet (sans apport de pcDNA3.1(+)-NP, motifs avec les lignes transversales) ont été représentées. Trois expériences indépendantes ont été réalisées. Les moyennes des valeurs brutes des signaux Nluc normalisées avec les signaux Fluc (Nluc/Fluc) ainsi que les écarts-types de chaque groupe sont représentés. ** et **** indiquent une p-value < 0.01 et 0.0001 respectivement. &&&& indique une différence significative entre le groupe pGL4.10-minig +NP et pGL4.10-ΔSTARTminig +NP avec une p-value < 0.0001.

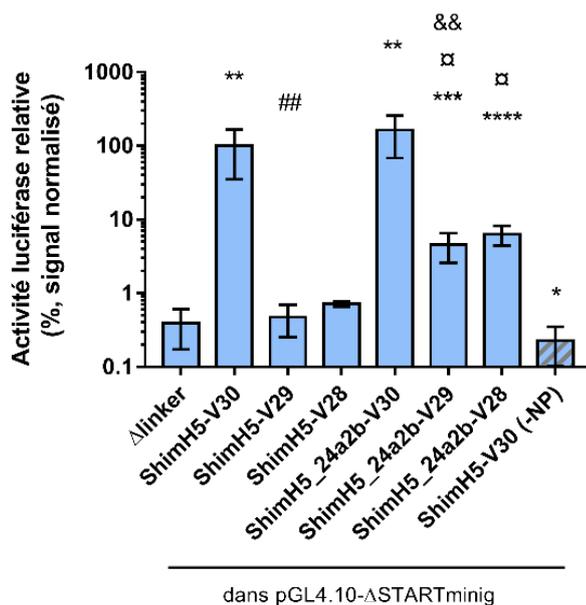
Il est important de noter qu'un signal luciférase non négligeable a été observé pour les conditions pUC57-ΔSTARTminig(+NP) et pGL4.10-ΔSTARTminig (+NP). Cette activité, dont l'origine reste inconnue, correspond au bruit de fond de notre système et a été évalué pour chaque expérience.

Les différents fragments de séquence nucléotidique (ou linkers) dérivant de la région environnante du site de clivage de la HA virale de ShimH5 ont été clonés et testés pour l'évaluation de notre système minigénome *in vitro*, concernant la capacité de certaines séquences HA d'induire des insertions, sous l'action de la polymérase virale (**Figure 19**). Deux versions de linkers dérivant de ShimH5 ont été testées : une première correspondant au fragment sauvage de la séquence HA de ShimH5, puis une seconde version correspondant à une forme mutée (24a2b) de la même séquence (2 substitutions : G vers A et C vers A, séquences des linkers, **Figure 19**). Cette forme mutée possède une structure secondaire d'ARNc prédite de type épingle à cheveux plus grande que la séquence sauvage, flanquant la région du site de clivage (Nao N, Yamagishi J et al. 2017).

En présence du complexe polymérase viral complet, un signal important, similaire à la construction pGL4.10-minig a été observé pour la condition avec le linker ShimH5-V30, ce qui

correspond à notre activité luciférase maximale. Un signal relatif de 0,39 % par rapport à la condition linker ShimH5-V30 a pu être enregistré avec la construction sans linker (Δ linker pGL4.10- Δ minig). Une augmentation très modeste de l'activité luciférase est observée pour les conditions avec linkers ShimH5-V29 et V28 (0,47 % et 0,71 % respectivement). En ce qui concerne la condition avec linker ShimH5_24a2b-V30, le signal équivaut à celui avec la forme non-mutée ShimH5-V30. Une augmentation significative de l'activité luciférase est observée pour les conditions avec les linkers ShimH5_24a2b-V29 et V28 en comparaison au bruit de fond (condition Δ linker), avec pour valeurs 4,54 % et 6,34 %, respectivement. De plus, cette augmentation est aussi significativement différente de ce qui a pu être enregistrée pour les linkers non mutés ShimH5-V29 et V28.

A. pGL4.10-Linkers_ShimH5_minigénomes



B. Linkers dérivant du segment HA H5 de ShimH5 (A/whistling swan/Shimane/499/83 (H5N3)) :

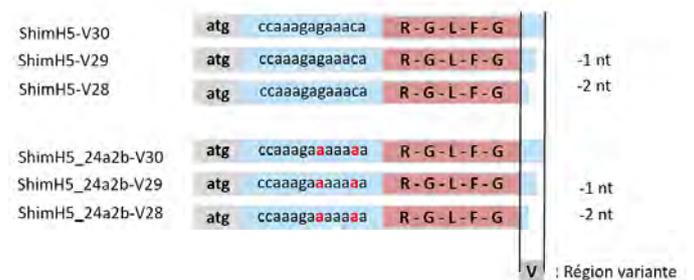


Figure 19 : Résultats des essais luciférase après la transfection de cellules CDF1 avec les constructions pGL4.10- Δ STARTminig (A.) avec les linkers dérivés de la séquence HA de ShimH5 (B.), et les plasmides pcDNA3.1(+)-PA, PB1, PB2 et NP (WSN). Les conditions avec le complexe de polymérase complet ont été représentées. La condition avec ShimH5-V30 sans NP apparaît avec les motifs en lignes transversales). Quatre expériences indépendantes ont été réalisées. Les moyennes des valeurs brutes des signaux Nluc normalisés avec les signaux Fluc (Nluc/Fluc) relatives au signal obtenu en ShimH5-V30 ainsi que les écarts-types de chaque groupe sont représentés. * : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$, **** : $p < 0.0001$ vs Δ Linker ; ## : $p < 0.01$ vs ShimH5-V30 ; □ : $p < 0.05$ vs 24a2b-V30 ; && : $p < 0.01$ vs ShimH5-V29.

Ces résultats suggèrent une augmentation des insertions en présence de la forme mutée 24a2b des linkers dérivant de la séquence HA de ShimH5, comme ce qui a pu être observé précédemment (Nao N, Yamagishi J et al. 2017). La structure secondaire d'ARN étant différente entre la forme sauvage et mutée (24a2b) des linkers, l'hypothèse d'une corrélation entre la présence et l'ampleur de la structure en épingle à cheveux flanquant la séquence correspondant au site de clivage de la HA semble être plausible.

Néanmoins, des différences notables en termes de signal peuvent être notées entre nos résultats et le travail précédemment publié (Nao N, Yamagishi J et al. 2017). Nos activités luciférase enregistrées pour les mêmes conditions sont globalement 2 à 3x inférieures par rapport à celles de Nao et de ses collègues. Par ailleurs, il est important de notifier que dans leur travail, un « plasmide vide » a été utilisé et présenté comme contrôle du bruit de fond de leur système. Du fait de l'imprécision sur l'identité de ce plasmide nous ne pouvons pas être certains de sa qualité. En début d'optimisation de notre système minigénome, nous avons observé un signal persistant notable même en l'absence de linker et de l'ATG initiateur (condition pGL4.10- Δ STARTminig, **Figure 18** et **Figure 19**). Il est donc nécessaire de mettre en relation les activités luciférase observées des séquences linkers testées en système minigénome (dérivés du segment HA de ShimH5) avec la construction complète minigénome utilisée, sans linker (Δ linker = pGL4.10- Δ STARTminig), qui apparait comme un contrôle plus rigoureux, représentant le bruit de fond.

Nous avons également testé notre système minigénome en présence du complexe polymérase aviaire H5N8 issue de la souche A/duck/France/161108h/2016 (H5N8) (numéro d'accèsion GISAID EPI_ISL_240012). Chaque sous-unité du complexe polymérase viral est codée sur un plasmide pHW2000.

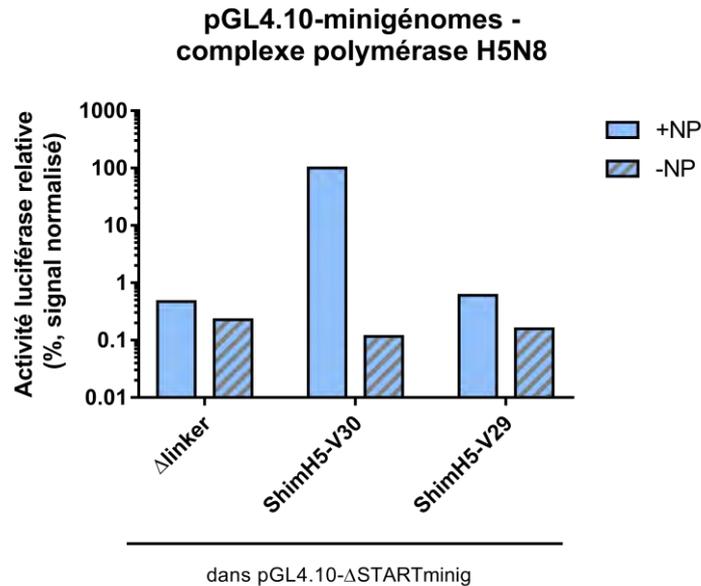


Figure 20 : Résultats des essais luciférase après la transfection de cellules CDF1 avec la construction pGL4.10-ΔSTARTminig avec et sans les linkers dérivés de la séquence HA de ShimH5 (ShimH5-V30 et V29, ainsi que les plasmides pHW2000-PA, PB1, PB2 et NP (H5N8). Les conditions avec le complexe de polymérase complet et sans NP (motifs avec les lignes transversales) ont été représentées. Une expérience a été réalisée (chaque condition en duplicat). Les moyennes des signaux Nluc normalisés avec les signaux Fluc (Nluc/Fluc) de chaque groupe sont représentées.

Nous pouvons observer des signaux équivalents à ceux obtenus en présence du complexe de polymérase WSN (H1N1) (Figure 20). Le ratio entre le signal obtenu en pGL4.10-minig(+NP) et pGL4.10-ΔSTARTminig (+NP) est de 41x, puis augmente jusqu'à 210x si l'on compare l'activité luciférase pGL4.10-ΔSTARTminig_ShimH5-V30 (+NP) / pGL4.10-ΔSTARTminig (+NP). Ces résultats sont légèrement inférieurs par rapport à ce qui a pu être espéré et mesuré avec l'utilisation du complexe polymérase WSN.

En conclusion des expériences précédentes, nous avons élaboré un système rapporteur minigénome dépendant de l'activité de la polymérase virale utilisable en cellules aviaires (de poulets ou de canards). Les complexes polymérase viraux mammifères (WSN) et aviaires (H5N8) peuvent être utilisés et donnent des résultats similaires. Toutefois, l'essai minigénome en présence du complexe polymérase aviaire reste à être répété plusieurs fois afin de vérifier la reproductibilité de nos premiers résultats, de même que le test des autres linkers dérivés de ShimH5 (ShimH5-V28, ShimH5_24a2b-V30, V29 et V28).

2.2.4. Contrôle de la spécificité du système minigénome pour la mesure du phénomène d'insertions

Afin de pouvoir nous assurer que l'augmentation du signal observé avec notre système minigénome est due à l'action duale du complexe polymérase virale et de la séquence linker, nous avons cloné les constructions linker ShimH5 et ShimH5_24a2b suivies de l'ORF de la Nluc (Δ ATG) dans un plasmide pcDNA3.1(+) sous contrôle du promoteur CMV (CMVp). De ce fait, les constructions sont testées dans un contexte où la polymérase virale est absente, sous la dépendance de l'ARN polymérase II cellulaire.

Des signaux luciférase relatifs de 0,18 % et 0,19 % ont été observés pour les groupes linkerShimH5-V29 et V28 respectivement par rapport à l'activité enregistrée avec pcDNA3.1-linker ShimH5-V30 (Figure 21). Le signal luciférase des versions mutées 24a2b des linkersShimH5-V29 et V28 a augmenté respectivement de près de 11x et 2x en comparaison avec leurs homologues sauvages. Cette augmentation du signal est observée de manière plus importante pour le groupe linkerShimH5_24a2b-V29, avec pour valeur 1,91 % du signal de la construction avec linkerShimH5-V30.

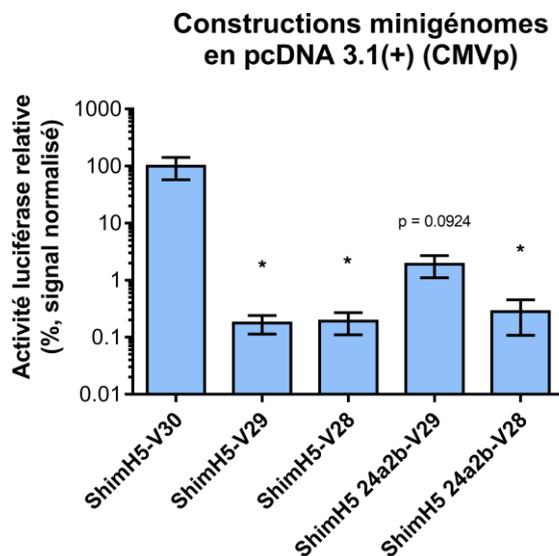


Figure 21 : Résultats des essais luciférase après la transfection de cellules CDF1 avec les constructions linkersShimH5(sauvage et 24a2b)-ORF_Nluc (Δ ATG) clonées en pcDNA3.1(+), avec le promoteur CMV en amont. Deux expériences indépendantes ont été réalisées. Les moyennes des valeurs brutes du signal Nluc ainsi que les écarts-types de chaque groupe sont représentés. * indique une différence significative avec une p-value <0.05 par rapport au groupe ShimH5-V30.

L'augmentation de l'activité luciférase premièrement observée dans un contexte mimant la transcription/réplication virale (**Figure 19** et **Figure 20**) est retrouvée dans un environnement polymérase-virale indépendant (**Figure 21**). L'ARN polymérase II cellulaire étant décrite comme très fidèle, nous ne pouvons pas avoir la certitude que notre système minigénome soit utilisable afin d'observer des événements d'insertions de nucléotides dépendant de la transcription et/ou de la réplication virale. L'expression de la Nluc dans ce contexte pourrait être due à des phénomènes de traduction alternative.

2.3. Discussion

La séquence nucléotidique codant la région du site de clivage de HA de virus de sous-type H5 (et H7) évolue par deux mécanismes : des mutations et/ou des insertions de nucléotides. A partir d'une banque de séquences nucléotidiques de HA H5 dans laquelle tous les virus ont été catégorisés comme HPAIV ou LPAIV, un arbre phylogénétique de la séquence HA H5 a pu être réalisé, permettant de replacer les principales émergences HPAIV. Les séquences HA de LPAIV les plus proches phylogénétiquement des émergences par insertion ont pu être identifiées et elles ont été dénommées séquences de LPAIV progénitrices de HPAIV. L'observation de l'arbre phylogénétique montre que les souches HPAIV ne sont pas regroupées sur des clades particuliers, indiquant que les émergences peuvent survenir dans plusieurs clades, ce qui souligne le caractère parallèle de ce type d'évolution (Escalera-Zamudio et al. 2020). A l'inverse, d'autres clades, totalement dépourvus de séquences HA de souches HPAIV émergentes ont été identifiés.

Les prédictions de structures secondaires d'ARN montrent la présence d'une structure conservée caractéristique de type épingle à cheveux au sein des séquences LPAIV (H5) progénitrices des émergences. Cette structure est située en amont de la région correspondant au site de clivage de la HA et flanque cette dernière. Cette dernière est différente chez la très grande majorité des séquences HA des souches de LPAIV isolées phylogénétiquement des émergences. En revanche, certaines séquences de LPAIV, aussi éloignées de toutes émergences, possèdent des structures similaires décrites comme favorables au phénomène d'insertions. Il serait intéressant d'investiguer leur pouvoir favorisant aux insertions de nucléotides. Une prédisposition génétique à l'émergence présente chez la totalité des souches LPAIV de sous-type H5 et/ou H7 a déjà été suggérée, impliquant la séquence nucléotidique environnant le site de clivage de la HA (Gulyaev et al. 2016; Nao N, Yamagishi J et al. 2017). Cependant, notre approche bio-informatique incluant les données phylogénétiques de séquences HA H5 et les prédictions de structures d'ARN suggère qu'une catégorisation des séquences nucléotidiques de HA des LPAIV (de sous-type H5) est réalisable. La majorité des séquences HA les plus proches génétiquement des émergences par insertion possèdent des structures en épingle à cheveux, décrites comme potentiellement favorables aux insertions par glissement ou bégaiement de la polymérase (Abolnik 2017; Garcia et al. 1996; Nao N, Yamagishi J et al. 2017; Perdue et al. 1997). Ces structures d'ARN sont structurellement différentes dans la majorité des séquences de souches LPAIV éloignées phylogénétiquement

des émergences connues. Ainsi, il semblerait que les séquences HA des LPAIV (de sous-type H5) puissent être catégorisées à la fois selon la structure secondaire prédite de la séquence ARN correspondant au site de clivage de HA et la distance génétique relative à l'événement d'émergence le plus proche. Cette classification ouvre des pistes de recherche pour compléter ces analyses bio-informatiques et tester expérimentalement les hypothèses générées.

Cependant, il est important de garder à l'esprit que l'ARN_v et l'ARN_c sont tous deux présents naturellement sous une forme recouverte par les NP. Sa fixation sur l'acide nucléique viral peut vraisemblablement désorganiser les structures secondaires d'ARN. Néanmoins, des études ont montré que l'association de NP le long de l'ARN_v n'est en réalité pas uniforme. Des zones moins denses en termes de fixation de NP existent, suggérant que les structures secondaires prédites d'ARN à ces niveaux exercent potentiellement des rôles fonctionnels (Williams et al. 2018). L'environnement autour de la séquence codant le site de clivage de HA en fait partie. D'autre part, il est difficile de conclure sur l'impact de structures secondaires d'ARN lors du passage de la matrice au niveau du site catalytique de la polymérase virale. En effet, comme il a été signalé dans cet article (Gulyaev et al. 2019), les insertions ayant lieu au niveau de la boucle de la structure en épingle à cheveu prédite, cette dernière devrait donc être présente dans la poche catalytique de la sous-unité PB1 afin de valider le modèle proposé par Nao et ses collègues (Nao N, Yamagishi J et al. 2017). Or, d'après les données issues des analyses ultra structurales connues sur le complexe polymérase virale des IAV, les contraintes exercées sur la matrice lors de son passage au sein de la polymérase devrait vraisemblablement désorganiser la présence de telles structures. Une longueur de 20 nucléotides sur la matrice d'ARN est présente au sein du complexe d'élongation, sous une forme simple brin (Chang et al. 2015; Wandzik et al. 2020). Il est donc peu probable qu'une structure secondaire de type épingle à cheveu d'une vingtaine de nucléotides puisse se reformer au sein du complexe d'élongation. Il est vraisemblable que d'autres éléments structuraux, indépendamment de la boucle variable de la structure en épingle à cheveux flanquant la séquence codant le site de clivage de HA, soient à l'origine des insertions. Considérant une fenêtre d'analyse plus large au niveau de la séquence codant le motif de clivage de HA, la zone correspondant à la tige de la structure prédite en épingle à cheveux et la stabilité des appariements intra-segment au niveau de cette zone contrainte, pourraient avoir un impact sur l'induction des insertions. Les extrémités libres entrante et sortante de la matrice d'ARN correspondant à la tige de la structure décrite précédemment pourraient se réhybrider, induisant un arrêt de l'élongation, suivie *in fine* d'un redémarrage de la polymérase (Figure 22). Ce modèle théorique pourrait potentiellement expliquer les insertions observées au niveau de la séquence située au sein du site catalytique correspondant à la séquence codant le motif de clivage de HA, et coïnciderait avec les

contraintes structurales de la polymérase virale (Gulyaev et al. 2019). Une analyse de prédiction de structures d'ARN des séquences HA a été réalisée en considérant une fenêtre d'analyse plus large (80 nucléotides de long, avec le site de clivage au centre). Le programme locARNA combinant une approche simultanée basée sur un alignement multiple et une prédiction de la structure la plus stable d'un point de vue thermodynamique a été utilisé. Deux groupes de séquence HA ont été étudiés : **groupe 1**, composé par les séquences HA des progéniteurs HPAIV par insertions, et le **groupe 2**, constitué par des séquences HA de LPAIV phylogénétiquement éloignées de toutes émergences connues. Par rapport aux résultats précédents (Tableau 5 et Tableau 6), une structure prédite de type épingle à cheveux flanquant la séquence codant le motif de clivage est présente dans les deux groupes (Figure 23). Or, le **groupe 1** montre un score de stabilité plus important (-21 kcal/mol) par rapport au **groupe 2** (-13 kcal/mol). Une différence moins importante en termes de stabilité de la structure a été également observée sur l'ARNv. Ces résultats suggèrent que des caractéristiques inhérentes à la séquence HA permettent toujours de différencier les souches LPAIV entre elles.

Des données préliminaires obtenues sur les séquences HA de LPAIV de sous-type H7 en termes de prédictions de structures d'ARN semble être en partie similaires à ce qui a pu être observé chez les séquences HA de sous-type H5. Des analyses bio-informatiques complémentaires sont en cours de traitement.

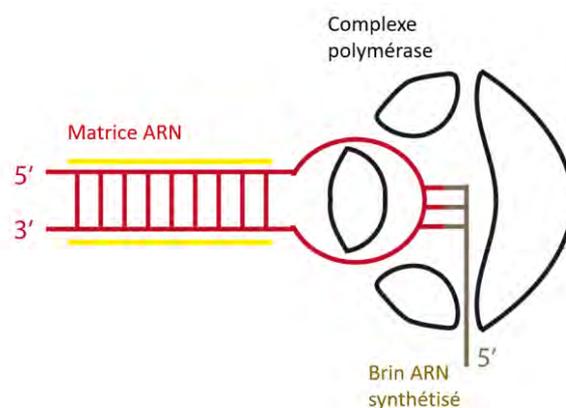


Figure 22 : Modèle théorique simplifié représentant la tige-boucle conservée prédite au niveau de la région codant le site de clivage de HA en présence du complexe polymérase virale. Les interactions entre le brins 5' et 3' de la matrice ARN sont soulignées en jaunes (figure adaptée de Gulyaev et al. 2019).

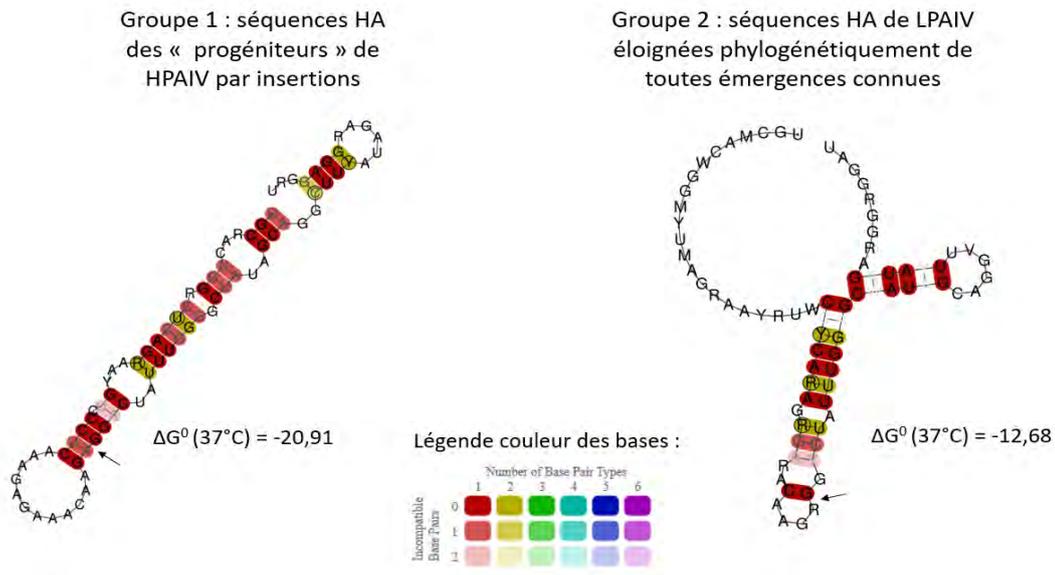


Figure 23 : Structures consensus issues de l'analyse comparative de la structure d'ARN prédite au niveau de la région codant le motif de clivage de HA, chez les groupes 1 et 2. La fenêtre d'analyse correspond au cadre suivant : 40 nucléotides (nt) – site de clivage de HA (représentée par une flèche noire) – 40 nt. Le programme LocARNA disponible en ligne (<http://rna.informatik.uni-freiburg.de/LocARNA/Input.jsp>) a été utilisé. Les groupes correspondant respectivement aux séquences HA (ARNc) des souches progénitrices des émergences HPAIV (sous-type H5) par insertions (groupe 1) et phylogénétiquement éloignées de toutes émergences connues (groupe 2) ont été analysées.

Malgré la grande différence de type de structures secondaires d'ARN prédites au niveau des séquences HA en fonction de la proximité génétique aux émergences HPAIV, d'autres méthodes non prédictives sont nécessaires et doivent compléter notre analyse. Les techniques de cartographie chimique du génome et l'analyse des signatures cinétiques de polymérase permettraient de vérifier l'intégrité de ces structures dans le contexte cellulaire ainsi qu'en présence de NP, afin de mieux appréhender leur impact réel sur la polymérase virale (Dulin et al. 2017; Merino et al. 2005; B. Lee et al. 2017; Feng, Chan, and Spitale 2017).

Dans ce but, une autre stratégie expérimentale consistant en un système minigénome basé sur le travail de Nao *et al.* a été développé et optimisé, dans le but d'analyser les phénomènes d'insertions de nucléotides (Nao N, Yamagishi J et al. 2017). Nous avons construit un système rapporteur « dual », codant à la fois pour la Nluc et un gène normalisateur Fluc. Ce système minigénome est utilisable en cellules aviaires de poulet et de canard, tant avec la polymérase virale d'origine aviaire (complexe polymérase H5N8), que d'origine mammifère (complexe polymérase WSN). Cette dernière souche, ayant subi de nombreux passages, est considérée comme fortement adaptée à de nombreuses conditions.

Comme il a pu être observé précédemment (Nao N, Yamagishi J et al. 2017), une augmentation de l'activité luciférase avec le système rapporteur a été retrouvée contenant les linkers dérivant de la séquence HA mutée de ShimH5 (24a2b), suggérant de potentiels phénomènes d'insertions qui rétablissent le cadre de lecture du gène rapporteur Nluc (**Figure 19**). Cependant une augmentation du signal luciférase avec les mêmes séquences linkers a aussi été retrouvée dans un contexte indépendant de la polymérase virale (**Figure 21**). La construction minigénome a été clonée dans un plasmide pcDNA3.1(+) sous le contrôle du promoteur CMV, en absence de polymérase virale. Ce contrôle important n'a pas été réalisé par Nao et ses collègues dans leur étude. Nous pouvons donc remettre en question l'utilité de leur système minigénome et du nôtre pour mesurer spécifiquement la probabilité d'insertion de nucléotides. D'autres mécanismes non transcriptionnels semblent induire la production d'une protéine Nluc fonctionnelle.

Du fait de l'observation d'un signal indépendant du complexe polymérase viral, nous proposons l'hypothèse suivante : des phénomènes de traduction alternative tels que le décalage du cadre de lecture ribosomal et/ou l'initiation au niveau de sites non-canoniques pourraient avoir lieu lors de nos tests, amenant à la formation d'une protéine Nluc fonctionnelle.

Chez les eucaryotes, il a été montré qu'une grande partie des ORFs fait l'objet d'initiation de la traduction au niveau de codons non-AUG. Bien que ce type d'initiation de la traduction semble être moins efficace que celle débutée par l'AUG canonique, il pourrait avoir des rôles importants en termes de régulation de l'expression protéique. Lors de situations de stress de la cellule (par exemple : le stress du réticulum endoplasmique), un arrêt de la traduction est observé pour la majorité des transcrits cellulaires, caractérisé par la phosphorylation du facteur d'initiation de la traduction eIF2 α . Cette inhibition générale de la traduction est conjointe à l'augmentation de la production spécifique de protéines nécessaires au rétablissement physiologique de la cellule, telle que la protéine chaperonne Bip. La traduction de cette dernière est initiée au niveau d'un codon alternatif non-AUG (Kearse and Wilusz 2017; de Klerk and 't Hoen 2015; Machkovech, Bloom, and Subramaniam 2018). De ce fait, nous avons choisi d'investiguer l'hypothèse d'initiations alternatives de la traduction qui pourraient avoir lieu en aval de l'AUG apportée par la séquence linker, au niveau de la partie interne de l'ORF de la Nluc. Une prédiction des ORFs possibles du gène codant la Nluc dans notre système minigénome a été réalisée en prenant compte les autres codons initiateurs non-AUG de la traduction (aTIS : *alternative translation initiation site*) et le codon initiateur AUG (cTIS : *canonic translation initiation site*) (**Figure 24**).



Figure 24 : Représentation des ORFs possibles du gène de la Nluc (utilisé comme rapporteur dans le système minigénome). Les codons AUG, CUG, UUG et GUG ont été pris en compte. Les codons alternatifs de l'initiation de la traduction (aTIS) présents sur la première moitié de l'ORF classique de la Nluc, ainsi que le codon AUG interne (cTIS) sont indiqués par les flèches en rouge.

Sept aTIS et un codon AUG interne ont pu être identifiés sur la première moitié du gène de la Nluc. Etant donné qu'une Nluc tronquée, correspondant au cadre 66°-171° (extrémité C terminale) résidu codé sur le gène Nluc, possède une activité enzymatique résiduelle (Zhao et al. 2016), il est possible que ces codons initient la traduction de protéines Nluc également fonctionnelles.

Un nouveau système minigénome dont l'ORF codant la Nluc muté pour ces codons initiateurs internes pourrait potentiellement diminuer le bruit de fond précédemment observé (signal du pUC57-ΔSTARTminig et pGL4.10-ΔSTARTminig avec NP, **Figure 17**, **Figure 18** et **Figure 19**). Si cette optimisation est réalisable, d'autres linkers représentatifs de la variabilité observée dans l'environnement nucléotidique codant le site de clivage de HA et dévoilant des structures secondaires d'ARN prédites différentes pourraient être testés aisément, à moindre coût. Les séquences HA de LPAIV sélectionnées en fonction de leur distance génétique avec les émergences par insertions et les données de prédiction de structures secondaires d'ARN semblent tout à fait indiquées (**Tableau 5** et **Tableau 6**). Néanmoins, afin de valider que le signal provient bien d'insertions, nous ferons appel au séquençage haut-débit. Cette dernière technique reste à l'heure actuelle la plus intéressante en termes de sensibilité et de par sa capacité à détecter différents événements ayant un impact potentiel sur l'évolution des souches LPAIV : les insertions/délétions et aussi les substitutions de nucléotides éventuelles, au niveau de la région du site de clivage de la HA. En effet, une transition G vers A a été décrite chez certains isolats LPAIV durant l'épidémie de HPAIV H5N2 au Mexique entre 1993-1995 (Garcia et al.

1996). Cette transition a également été retrouvée sur des souches HPAIV ayant émergé par duplication de codons, suggérant que certaines substitutions pourraient influencer l'apparition secondaires d'insertions de nucléotides. Le modèle selon lequel le basculement de certaines souches LPAIV vers des formes HPAIV est sous la dépendance d'un phénomène évolutif séquentiel est fortement possible et mériterait d'être investigué davantage (Abolnik 2017).

3. Références bibliographiques

- Abdelwhab, El-Sayed M, Jutta Veits, and Thomas C Mettenleiter. 2013. “Genetic Changes That Accompanied Shifts of Low Pathogenic Avian Influenza Viruses toward Higher Pathogenicity in Poultry.” *Virulence* 4 (6): 441–52. <https://doi.org/10.4161/viru.25710>.
- Abolnik, Celia. 2017. “Evolution of H5 Highly Pathogenic Avian Influenza: Sequence Data Indicate Stepwise Changes in the Cleavage Site.” *Archives of Virology* 162 (8): 2219–30. <https://doi.org/10.1007/s00705-017-3337-x>.
- Ayllon, Juan, and Adolfo García-Sastre. 2015. “The NS1 Protein: A Multitasking Virulence Factor.” In *Influenza Pathogenesis and Control - Volume II*, edited by Michael B A Oldstone and Richard W Compans, 73–107. Current {Topics} in {Microbiology} and {Immunology}. Cham: Springer International Publishing. https://doi.org/10.1007/82_2014_400.
- Barbezange, Cyril, Louis Jones, Hervé Blanc, Ofer Isakov, Gershon Celniker, Vincent Enouf, Noam Shomron, Marco Vignuzzi, and Sylvie van der Werf. 2018. “Seasonal Genetic Drift of Human Influenza A Virus Quasispecies Revealed by Deep Sequencing.” *Frontiers in Microbiology* 9 (September). <https://doi.org/10.3389/fmicb.2018.02596>.
- Bjørnstad, Ottar N, and Cecile Viboud. 2016. “Timing and Periodicity of Influenza Epidemics.” *PNAS* 113 (46): 12899–901. <https://doi.org/10.1073/pnas.1616052113>.
- Bragstad, Karoline, Lars P Nielsen, and Anders Fomsgaard. 2008. “The Evolution of Human Influenza A Viruses from 1999 to 2006: A Complete Genome Study.” *Virology Journal* 5 (September): 40. <https://doi.org/10.1186/1743-422X-5-40>.
- Brown, Justin D., Ginger Goekjian, Rebecca Poulson, Steve Valeika, and David E. Stallknecht. 2009. “Avian Influenza Virus in Water: Infectivity Is Dependent on PH, Salinity and Temperature.” *Veterinary Microbiology* 136 (1–2): 20–26. <https://doi.org/10.1016/j.vetmic.2008.10.027>.
- Brunotte, Linda, Joe Flies, Hardin Bolte, Peter Reuther, Frank Vreede, and Martin Schwemmler. 2014. “The Nuclear Export Protein of H5N1 Influenza A Viruses Recruits Matrix 1 (M1) Protein to the Viral Ribonucleoprotein to Mediate Nuclear Export.” *Journal of Biological Chemistry* 289 (29): 20067–77. <https://doi.org/10.1074/jbc.M114.569178>.
- Castro Martin, Isabel Fernández De, Guillaume Fournier, Martin Sachse, Javier Pizarro-Cerda, Cristina Risco, and Nadia Naffakh. 2017. “Influenza Virus Genome Reaches the Plasma Membrane via a Modified Endoplasmic Reticulum and Rab11-Dependent Vesicles.” *Nature Communications* 8 (1). <https://doi.org/10.1038/s41467-017-01557-6>.
- Chang, Shenghai, Dapeng Sun, Huanhuan Liang, Jia Wang, Jun Li, Lu Guo, Xiangli Wang, et al. 2015. “Cryo-EM Structure of Influenza Virus RNA Polymerase Complex at 4.3 Å Resolution.” *Molecular Cell* 57 (5): 925–35. <https://doi.org/10.1016/J.MOLCEL.2014.12.031>.
- Cros, Jerome F., and Peter Palese. 2003. “Trafficking of Viral Genomic RNA into and out of the Nucleus: Influenza, Thogoto and Borna Disease Viruses.” *Virus Research* 95 (1–2): 3–12. [https://doi.org/10.1016/S0168-1702\(03\)00159-X](https://doi.org/10.1016/S0168-1702(03)00159-X).
- Dadonaite, Bernadeta, Brad Gilbertson, Michael L. Knight, Sanja Trifkovic, Steven Rockman, Alain Laederach, Lorena E. Brown, Ervin Fodor, and David L. V. Bauer. 2019. “The Structure of the Influenza A Virus Genome.” *Nature Microbiology*, July, 1–9. <https://doi.org/10.1038/s41564-019-0513-7>.
- Dawson, Wayne K, Michal Lazniewski, and Dariusz Plewczynski. 2018. “RNA Structure Interactions and Ribonucleoprotein Processes of the Influenza A Virus.” *Briefings in Functional Genomics* 17 (6): 402–14. <https://doi.org/10.1093/bfpg/elx028>.
- Dou, Dan, Rebecca Revol, Henrik Östbye, Hao Wang, and Robert Daniels. 2018. “Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement.” *Frontiers in Immunology* 9 (July): 1581. <https://doi.org/10.3389/fimmu.2018.01581>.
- Dulin, David, Jamie J Arnold, Theo van Laar, Hyung-Suk Oh, Cheri Lee, Angela L Perkins, Daniel A Harki, Martin Depken, Craig E Cameron, and Nynke H Dekker. 2017. “Signatures of Nucleotide Analog Incorporation by an RNA-Dependent RNA Polymerase Revealed Using High-Throughput Magnetic Tweezers.” *Cell Reports* 21 (4): 1063–76. <https://doi.org/10.1016/j.celrep.2017.10.005>.
- Edinger, Thomas O., Marie O. Pohl, and Silke Stertz. 2014. “Entry of Influenza A Virus: Host Factors and Antiviral Targets.” *Journal of General Virology*. <https://doi.org/10.1099/vir.0.059477-0>.

- Escalera-Zamudio, Marina, Michael Golden, Bernardo Gutiérrez, Julien Thézé, Jeremy Russell Keown, Loic Carrique, Thomas A Bowden, and Oliver G Pybus. 2020. “Parallel Evolution in the Emergence of Highly Pathogenic Avian Influenza A Viruses.” *Nature Communications* 11 (1): 5511. <https://doi.org/10.1038/s41467-020-19364-x>.
- Evseev, Danyel, and Katharine Magor. 2019. “Innate Immune Responses to Avian Influenza Viruses in Ducks and Chickens.” *Veterinary Sciences* 6 (1): 5. <https://doi.org/10.3390/vetsci6010005>.
- Feng, Chao, Dalen Chan, and Robert C. Spitale. 2017. “Assaying RNA Structure inside Living Cells with SHAPE.” In *Methods in Molecular Biology*, 1648:247–56. Humana Press Inc. https://doi.org/10.1007/978-1-4939-7204-3_18.
- Ferhadian, Damien, Maud Contrant, Anne Printz-Schweigert, Redmond P Smyth, Jean-Christophe Paillart, and Roland Marquet. 2018. “Structural and Functional Motifs in Influenza Virus RNAs.” *Frontiers in Microbiology* 9: 559. <https://doi.org/10.3389/fmicb.2018.00559>.
- Flynn, Ryan A., Qiangfeng Cliff Zhang, Robert C. Spitale, Byron Lee, Maxwell R. Mumbach, and Howard Y. Chang. 2016. “Transcriptome-Wide Interrogation of RNA Secondary Structure in Living Cells with IcSHAPE.” *Nature Protocols* 11 (2): 273–90. <https://doi.org/10.1038/nprot.2016.011>.
- Fodor, E. 2013. “The RNA Polymerase of Influenza A Virus: Mechanisms of Viral Transcription and Replication.” *Acta Virologica* 57 (2): 113–22. https://doi.org/10.4149/av_2013_02_113.
- França, M., D. E. Stallknecht, and E. W. Howerth. 2013. “Expression and Distribution of Sialic Acid Influenza Virus Receptors in Wild Birds.” *Avian Pathology* 42 (1): 60–71. <https://doi.org/10.1080/03079457.2012.759176>.
- Gamblin, Steven J., and John J. Skehel. 2010. “Influenza Hemagglutinin and Neuraminidase Membrane Glycoproteins.” *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc. <https://doi.org/10.1074/jbc.R110.129809>.
- Garcia, M., J. M. Crawford, J. W. Latimer, E. Rivera-Cruz, and M. L. Perdue. 1996. “Heterogeneity in the Haemagglutinin Gene and Emergence of the Highly Pathogenic Phenotype among Recent H5N2 Avian Influenza Viruses from Mexico.” *Journal of General Virology* 77 (7): 1493–1504. <https://doi.org/10.1099/0022-1317-77-7-1493>.
- Gavazzi, Cyrille, Catherine Isel, Emilie Fournier, Vincent Moules, Annie Cavalier, Daniel Thomas, Bruno Lina, and Roland Marquet. 2013. “An in Vitro Network of Intermolecular Interactions between Viral RNA Segments of an Avian H5N2 Influenza A Virus: Comparison with a Human H3N2 Virus.” *Nucleic Acids Research* 41 (2): 1241–54. <https://doi.org/10.1093/nar/gks1181>.
- Gultyaev, Alexander P, Mathilde Richard, Monique I Spronken, René C L Olsthoorn, and Ron A M Fouchier. 2019. “Conserved Structural RNA Domains in Regions Coding for Cleavage Site Motifs in Hemagglutinin Genes of Influenza Viruses.” *Virus Evolution* 5 (2): 1–11. <https://doi.org/10.1093/ve/vez034>.
- Gultyaev, Alexander P, Monique I Spronken, Mathilde Richard, Eefje J A Schrauwen, René C L Olsthoorn, and Ron A M Fouchier. 2016. “Subtype-Specific Structural Constraints in the Evolution of Influenza A Virus Hemagglutinin Genes.” *Scientific Reports* 6: 38892. <https://doi.org/10.1038/srep38892>.
- Gultyaev, Alexander P, Anton Tsyganov-Bodounov, Monique I J Spronken, Sander van der Kooij, Ron A M Fouchier, and René C L Olsthoorn. 2014. “RNA Structural Constraints in the Evolution of the Influenza A Virus Genome NP Segment.” *RNA Biology* 11 (7): 942–52. <https://doi.org/10.4161/rna.29730>.
- Harris, Audray, Giovanni Cardone, Dennis C. Winkler, J. Bernard Heymann, Matthew Brecher, Judith M. White, and Alasdair C. Steven. 2006. “Influenza Virus Pleiomorphy Characterized by Cryoelectron Tomography.” *Proceedings of the National Academy of Sciences* 103 (50): 19123–27. <https://doi.org/10.1073/PNAS.0607614103>.
- Hu, Jiao, Chunxi Ma, and Xiufan Liu. 2018. “PA-X: A Key Regulator of Influenza A Virus Pathogenicity and Host Immune Responses.” *Medical Microbiology and Immunology* 207 (5–6): 255–69. <https://doi.org/10.1007/s00430-018-0548-z>.
- Jiang, Tian, Aitor Nogales, Steven F Baker, Luis Martinez-Sobrido, and Douglas H Turner. 2016. “Mutations Designed by Ensemble Defect to Misfold Conserved RNA Structures of Influenza A Segments 7 and 8 Affect Splicing and Attenuate Viral Replication in Cell Culture.” Edited by Stefan Pöhlmann. *PLOS ONE* 11 (6): e0156906. <https://doi.org/10.1371/journal.pone.0156906>.
- Kamal, Ram P, Irina V Alymova, and Ian A York. 2017. “Evolution and Virulence of Influenza A Virus Protein PB1-F2.” *International Journal of Molecular Sciences* 19 (1).

- <https://doi.org/10.3390/ijms19010096>.
- Kearse, Michael G, and Jeremy E Wilusz. 2017. "Non-AUG Translation: A New Start for Protein Synthesis in Eukaryotes." *Genes & Development* 31 (17): 1717–31. <https://doi.org/10.1101/gad.305250.117>.
- Klerk, Eleonora de, and Peter A.C. 't Hoen. 2015. "Alternative mRNA Transcription, Processing, and Translation: Insights from RNA Sequencing." *Trends in Genetics* 31 (3): 128–39. <https://doi.org/10.1016/j.tig.2015.01.001>.
- Kobayashi, Yuki, Bernadeta Dadonaite, Neeltje van Doremalen, Yoshiyuki Suzuki, Wendy S. Barclay, and Oliver G. Pybus. 2016. "Computational and Molecular Analysis of Conserved Influenza A Virus RNA Secondary Structures Involved in Infectious Virion Production." *RNA Biology* 13 (9): 883–94. <https://doi.org/10.1080/15476286.2016.1208331>.
- Kolakofsky, Daniel. 2016. "Paramyxovirus RNA Synthesis, MRNA Editing, and Genome Hexamer Phase: A Review." *Virology* 498 (November): 94–98. <https://doi.org/10.1016/j.virol.2016.08.018>.
- Koscielniak, Dawid, Ewa Wons, Karolina Wilkowska, and Marian Sektas. 2018. "Non-Programmed Transcriptional Frameshifting Is Common and Highly RNA Polymerase Type-Dependent." *Microbial Cell Factories* 17 (1): 184. <https://doi.org/10.1186/s12934-018-1034-4>.
- Kosik, Ivan, and Jonathan W. Yewdell. 2019. "Influenza Hemagglutinin and Neuraminidase: Yin–Yang Proteins Coevolving to Thwart Immunity." *Viruses*. MDPI AG. <https://doi.org/10.3390/v11040346>.
- Krammer, Florian, Gavin J. D. Smith, Ron A. M. Fouchier, Malik Peiris, Katherine Kedzierska, Peter C. Doherty, Peter Palese, et al. 2018. "Influenza." *Nature Reviews Disease Primers* 4 (1): 3. <https://doi.org/10.1038/s41572-018-0002-y>.
- Kuribayashi, Saya, Yoshihiro Sakoda, Takeshi Kawasaki, Tomohisa Tanaka, Naoki Yamamoto, Masatoshi Okamoto, Norikazu Isoda, et al. 2013. "Excessive Cytokine Response to Rapid Proliferation of Highly Pathogenic Avian Influenza Viruses Leads to Fatal Systemic Capillary Leakage in Chickens." Edited by Elankumaran Subbiah. *PLoS ONE* 8 (7): e68375. <https://doi.org/10.1371/journal.pone.0068375>.
- Leahy, M. B., H. C. Dobbyn, and G. G. Brownlee. 2001. "Hairpin Loop Structure in the 3' Arm of the Influenza A Virus Virion RNA Promoter Is Required for Endonuclease Activity." *Journal of Virology* 75 (15): 7042–49. <https://doi.org/10.1128/jvi.75.15.7042-7049.2001>.
- Leahy, M B, D C Pritlove, L L Poon, and G G Brownlee. 2001. "Mutagenic Analysis of the 5' Arm of the Influenza A Virus Virion RNA Promoter Defines the Sequence Requirements for Endonuclease Activity." *Journal of Virology* 75 (1): 134–42. <https://doi.org/10.1128/JVI.75.1.134-142.2001>.
- Lee, Byron, Ryan A. Flynn, Anastasia Kadina, Jimmy K. Guo, Eric T. Kool, and Howard Y. Chang. 2017. "Comparison of SHAPE Reagents for Mapping RNA Structures inside Living Cells." *RNA* 23 (2): 169–74. <https://doi.org/10.1261/rna.058784.116>.
- Lee, Mi-Kyung, Hee-Eun Kim, Eun-Byeol Park, Janghyun Lee, Ki-Hun Kim, Kyungeun Lim, Seoyun Yum, et al. 2016. "Structural Features of Influenza A Virus Panhandle RNA Enabling the Activation of RIG-I Independently of 5'-Triphosphate." *Nucleic Acids Research* 44 (17): 8407–16. <https://doi.org/10.1093/nar/gkw525>.
- Lim, Chun Shen, and Chris M. Brown. 2018. "Know Your Enemy: Successful Bioinformatic Approaches to Predict Functional RNA Structures in Viral RNAs." *Frontiers in Microbiology* 8 (January): 2582. <https://doi.org/10.3389/fmicb.2017.02582>.
- Lin, Chenyu, and Wayne O Miles. 2019. "Beyond CLIP: Advances and Opportunities to Measure RBP-RNA and RNA-RNA Interactions." *Nucleic Acids Research* 47 (11): 5490–5501. <https://doi.org/10.1093/nar/gkz295>.
- Lowen, Anice C. 2017. "Constraints, Drivers, and Implications of Influenza A Virus Reassortment." *Annual Review of Virology* 4 (1): 105–21. <https://doi.org/10.1146/annurev-virology-101416-041726>.
- Luo, G X, W Luytjes, M Enami, and P Palese. 1991. "The Polyadenylation Signal of Influenza Virus RNA Involves a Stretch of Uridines Followed by the RNA Duplex of the Panhandle Structure." *Journal of Virology* 65 (6): 2861–67. <http://www.ncbi.nlm.nih.gov/pubmed/2033659>.
- Machkovech, Heather M, Jesse D Bloom, and Arvind R Subramaniam. 2018. "Comprehensive Profiling of Translation Initiation in Influenza Virus Infected Cells." *BioRxiv*, January, 326967. <https://doi.org/10.1101/326967>.

- Mehedi, Masfique, Thomas Hoenen, Shelly Robertson, Stacy Ricklefs, Michael A Dolan, Travis Taylor, Darryl Falzarano, Hideki Ebihara, Stephen F Porcella, and Heinz Feldmann. 2013. “Ebola Virus RNA Editing Depends on the Primary Editing Site Sequence and an Upstream Secondary Structure.” *PLOS Pathogens* 9 (10): e1003677. <https://doi.org/10.1371/journal.ppat.1003677>.
- Mercier, Philippe Le, and Daniel Kolakofsky. 2019. “Bipartite Promoters and RNA Editing of Paramyxoviruses and Filoviruses.” *Rna* 25 (3): 279–85. <https://doi.org/10.1261/rna.068825.118>.
- Merino, Edward J., Kevin A. Wilkinson, Jennifer L. Coughlan, and Kevin M. Weeks. 2005. “RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE).” *Journal of the American Chemical Society* 127 (12): 4223–31. <https://doi.org/10.1021/ja043822v>.
- Monne, Isabella, Alice Fusaro, Martha I Nelson, Lebana Bonfanti, Paolo Mulatti, Joseph Hughes, Pablo R Murcia, et al. 2014. “Emergence of a Highly Pathogenic Avian Influenza Virus from a Low-Pathogenic Progenitor.” *Journal of Virology* 88 (8): 4375–88. <https://doi.org/10.1128/JVI.03181-13>.
- Mostafa, Ahmed, Elsayed Abdelwhab, Thomas Mettenleiter, and Stephan Pleschka. 2018. “Zoonotic Potential of Influenza A Viruses: A Comprehensive Overview.” *Viruses* 10 (9): 497. <https://doi.org/10.3390/v10090497>.
- Moumen, Abdeladim, Lucette Polomack, Torsten Unge, Michel Véron, Henri Buc, and Matteo Negroni. 2003. “Evidence for a Mechanism of Recombination during Reverse Transcription Dependent on the Structure of the Acceptor RNA.” *Journal of Biological Chemistry* 278 (18): 15973–82. <https://doi.org/10.1074/jbc.M212306200>.
- Munster, V. J., and R. A.M. Fouchier. 2009. “Avian Influenza Virus: Of Virus and Bird Ecology.” *Vaccine*. <https://doi.org/10.1016/j.vaccine.2009.02.082>.
- Muramoto, Y., T. Noda, E. Kawakami, R. Akkina, and Y. Kawaoka. 2013. “Identification of Novel Influenza A Virus Proteins Translated from PA mRNA.” *Journal of Virology* 87 (5): 2455–62. <https://doi.org/10.1128/jvi.02656-12>.
- Nao N, Yamagishi J, Miyamoto H, Tsuda Y Igarashi M, Manzoor R, Ohnuma A, Kishida N Furuyama W, Shigeno A, Kajihara M, and Takada A. Yoshida R. 2017. “Genetic Predisposition To Acquire a Polybasic Cleavage Site for Highly.” *MBio* 8 (1): 1–15. <https://doi.org/10.1128/mBio.02298-16>.
- Nayak, Debi P., Eric Ka Wai Hui, and Subrata Barman. 2004. “Assembly and Budding of Influenza Virus.” *Virus Research* 106 (2 SPEC.ISS.): 147–65. <https://doi.org/10.1016/j.virusres.2004.08.012>.
- Nelson, Martha I., and Edward C. Holmes. 2007. “The Evolution of Epidemic Influenza.” *Nature Reviews Genetics*. <https://doi.org/10.1038/nrg2053>.
- Neumann, G, M R Castrucci, and Y Kawaoka. 1997. “Nuclear Import and Export of Influenza Virus Nucleoprotein.” *Journal of Virology* 71 (12): 9690–9700. <http://www.ncbi.nlm.nih.gov/pubmed/9371635>.
- Noda, Takeshi. 2011. “Native Morphology of Influenza Virions.” *Frontiers in Microbiology* 2: 269. <https://doi.org/10.3389/fmicb.2011.00269>.
- Nogales, Aitor, Luis Martinez-Sobrido, David J Topham, and Marta L DeDiego. 2018. “Modulation of Innate Immune Responses by the Influenza A NS1 and PA-X Proteins.” *Viruses* 10 (12). <https://doi.org/10.3390/v10120708>.
- Ohuchi, Masanobu, Michaela Orlich, Reiko Ohuchi, Barry E.J. Simpson, Wolfgang Garten, Hans-Dieter Klenk, and Rudolf Rott. 1989. “Mutations at the Cleavage Site of the Hemagglutinin Alter the Pathogenicity of Influenza Virus a/Chick/Penn/83 (H5N2).” *Virology* 168 (2): 274–80. [https://doi.org/10.1016/0042-6822\(89\)90267-5](https://doi.org/10.1016/0042-6822(89)90267-5).
- Olsen, Björn, Vincent J Munster, Anders Wallensten, Jonas Waldenström, Albert D M E Osterhaus, and Ron A M Fouchier. 2006. “Global Patterns of Influenza a Virus in Wild Birds.” *Science (New York, N.Y.)* 312 (5772): 384–88. <https://doi.org/10.1126/science.1122438>.
- Olsper, Allan, Betty Y-W Chung, John F Atkins, John P Carr, and Andrew E Firth. 2015. “Transcriptional Slippage in the Positive-sense RNA Virus Family *Potyviridae*.” *EMBO Reports* 16 (8): 995–1004. <https://doi.org/10.15252/embr.201540509>.
- Parsch, J, J M Braverman, and W Stephan. 2000. “Comparative Sequence Analysis and Patterns of Covariation in {RNA} Secondary Structures.” *Genetics* 154 (2): 909–21.
- Pasick, John, Katherine Handel, John Robinson, John Copps, Deidre Ridd, Kevin Hills, Helen Kehler,

- et al. 2005. “Intersegmental Recombination between the Haemagglutinin and Matrix Genes Was Responsible for the Emergence of a Highly Pathogenic H7N3 Avian Influenza Virus in British Columbia.” *The Journal of General Virology* 86 (Pt 3): 727–31. <https://doi.org/10.1099/vir.0.80478-0>.
- Pauly, Matthew D, Megan C Procaro, and Adam S Lauring. 2017. “A Novel Twelve Class Fluctuation Test Reveals Higher than Expected Mutation Rates for Influenza A Viruses.” Edited by Karla Kirkegaard. *ELife* 6 (May): e26437. <https://doi.org/10.7554/eLife.26437>.
- Perdue, M L, M García, D Senne, and M Fraire. 1997. “Virulence-Associated Sequence Duplication at the Hemagglutinin Cleavage Site of Avian Influenza Viruses.” *Virus Res.* 49 (2): 173–86.
- Pflug, Alexander, Delphine Guilligay, Stefan Reich, and Stephen Cusack. 2014. “Structure of Influenza A Polymerase Bound to the Viral RNA Promoter.” *Nature* 516 (7531): 355–60. <https://doi.org/10.1038/nature14008>.
- Pflug, Alexander, Maria Lukarska, Patricia Resa-Infante, Stefan Reich, and Stephen Cusack. 2017. “Structural Insights into RNA Synthesis by the Influenza Virus Transcription-Replication Machine.” *Virus Research* 234 (April): 103–17. <https://doi.org/10.1016/J.VIRUSRES.2017.01.013>.
- Pielak, Rafal M., and James J. Chou. 2011. “Influenza M2 Proton Channels.” *Biochimica et Biophysica Acta - Biomembranes*. <https://doi.org/10.1016/j.bbamem.2010.04.015>.
- Riel, Debby Van, Vincent J. Munster, Emmie De Wit, Guus F. Rimmelzwaan, Ron A.M. Fouchier, Ab D.M.E. Osterhaus, and Thijs Kuiken. 2006. “H5N1 Virus Attachment to Lower Respiratory Tract.” *Science* 312 (5772): 399. <https://doi.org/10.1126/science.1125548>.
- Rossman, Jeremy S., and Robert A. Lamb. 2011. “Influenza Virus Assembly and Budding.” *Virology*. <https://doi.org/10.1016/j.virol.2010.12.003>.
- Ruigrok, Rob WH, Thibaut Crépin, Darren J. Hart, and Stephen Cusack. 2010. “Towards an Atomic Resolution Understanding of the Influenza Virus Replication Machinery.” *Current Opinion in Structural Biology*. <https://doi.org/10.1016/j.sbi.2009.12.007>.
- SantaLucia, J. 1998. “A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (4): 1460–65. <https://doi.org/10.1073/pnas.95.4.1460>.
- Simon-Loriere, Etienne, Darren P Martin, Kevin M Weeks, and Matteo Negroni. 2010. “RNA Structures Facilitate Recombination-Mediated Gene Swapping in HIV-1.” *Journal of Virology* 84 (24): 12675–82. <https://doi.org/10.1128/JVI.01302-10>.
- Spackman, E. 2009. “The Ecology of Avian Influenza Virus in Wild Birds: What Does This Mean for Poultry?” *Poultry Science* 88 (4): 847–50. <https://doi.org/10.3382/ps.2008-00336>.
- Stallknecht, D. E., M. T. Kearney, S. M. Shane, and P. J. Zwank. 1990. “Effects of PH, Temperature, and Salinity on Persistence of Avian Influenza Viruses in Water.” *Avian Diseases* 34 (2): 412. <https://doi.org/10.2307/1591429>.
- Steinhauer, David A. 1999. “Role of Hemagglutinin Cleavage for the Pathogenicity of Influenza Virus.” *Virology* 258 (1): 1–20. <https://doi.org/10.1006/viro.1999.9716>.
- Swayne, D E, and M Pantin-Jackwood. 2006. “Pathogenicity of Avian Influenza Viruses in Poultry.” *Developments in Biologicals* 124: 61–67. <http://www.ncbi.nlm.nih.gov/pubmed/16447495>.
- Swayne, David E. 2007. “Understanding the Complex Pathobiology of High Pathogenicity Avian Influenza Viruses in Birds.” *Avian Diseases* 51 (s1): 242–49. <https://doi.org/10.1637/7763-110706-regr.1>.
- Vandegrift, Kurt J., Susanne H. Sokolow, Peter Daszak, and A. Marm Kilpatrick. 2010. “Ecology of Avian Influenza Viruses in a Changing World.” *Annals of the New York Academy of Sciences*. Blackwell Publishing Inc. <https://doi.org/10.1111/j.1749-6632.2010.05451.x>.
- Veits, Jutta, Siegfried Weber, Olga Stech, Angele Breithaupt, Marcus Gräber, Sandra Gohrbandt, Jessica Bogs, et al. 2012. “Avian Influenza Virus Hemagglutinins H2, H4, H8, and H14 Support a Highly Pathogenic Phenotype.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (7): 2579–84. <https://doi.org/10.1073/pnas.1109397109>.
- Velthuis, Aartjan J W Te, and Ervin Fodor. 2016. “Influenza Virus RNA Polymerase: Insights into the Mechanisms of Viral RNA Synthesis.” *Nature Reviews. Microbiology* 14 (8): 479–93. <https://doi.org/10.1038/nrmicro.2016.87>.
- Vlugt, Corey De, Dorota Sikora, and Martin Pelchat. 2018. “Insight into Influenza: A Virus Cap-Snatching.” *Viruses* 10 (11): 641. <https://doi.org/10.3390/v10110641>.
- Wahlgen, John. 2011. “Influenza A Viruses: An Ecology Review.” *Infection Ecology &*

- Epidemiology* 1 (1): 6004. <https://doi.org/10.3402/iee.v1i0.6004>.
- Wandzik, Joanna M., Tomas Kouba, Manikandan Karuppasamy, Alexander Pflug, Petra Drncova, Jan Provaznik, Nayara Azevedo, and Stephen Cusack. 2020. "A Structure-Based Model for the Complete Transcription Cycle of Influenza Polymerase." *Cell* 181 (4): 877-893.e21. <https://doi.org/10.1016/j.cell.2020.03.061>.
- Wang, Jingfeng, Jinghua Li, Lili Zhao, Mengmeng Cao, and Tao Deng. 2017. "Dual Roles of the Hemagglutinin Segment-Specific Noncoding Nucleotides in the Extended Duplex Region of the Influenza A Virus RNA Promoter." *Journal of Virology* 91 (1). <https://doi.org/10.1128/jvi.01931-16>.
- Wang, Qiao, Ranran Liu, Qinghe Li, Fei Wang, Bo Zhu, Maiqing Zheng, Huanxian Cui, Jie Wen, and Guiping Zhao. 2019. "Host Cell Interactome of PB1 N40 Protein of H5N1 Influenza A Virus in Chicken Cells." *Journal of Proteomics* 197 (April): 34–41. <https://doi.org/10.1016/j.jprot.2019.02.011>.
- Webster, R. G., W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka. 1992. "Evolution and Ecology of Influenza A Viruses." *Microbiological Reviews*. <https://doi.org/10.1111/j.1751-0813.1992.tb09940.x>.
- Williams, Graham D., Dana Townsend, Kristine M. Wylie, Preston J. Kim, Gaya K. Amarasinghe, Sebla B. Kutluay, and Adrianus C. M. Boon. 2018. "Nucleotide Resolution Mapping of Influenza A Virus Nucleoprotein-RNA Interactions Reveals RNA Features Required for Replication." *Nature Communications* 9 (1): 465. <https://doi.org/10.1038/s41467-018-02886-w>.
- Wise, Helen M., Edward C. Hutchinson, Brett W. Jagger, Amanda D. Stuart, Zi H. Kang, Nicole Robb, Louis M. Schwartzman, et al. 2012. "Identification of a Novel Splice Variant Form of the Influenza A Virus M2 Ion Channel with an Antigenically Distinct Ectodomain." Edited by Andrew Pekosz. *PLoS Pathogens* 8 (11): e1002998. <https://doi.org/10.1371/journal.ppat.1002998>.
- Yamayoshi, Seiya, Mariko Watanabe, Hideo Goto, and Yoshihiro Kawaoka. 2015. "Identification of a Novel Viral Protein Expressed from the PB2 Segment of Influenza A Virus." *Journal of Virology* 90 (1): 444–56. <https://doi.org/10.1128/JVI.02175-15>.
- Zhang, Jinwei, and Robert Landick. 2016. "A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure." *Trends in Biochemical Sciences* 41 (4): 293–310. <https://doi.org/10.1016/j.tibs.2015.12.009>.
- Zhao, Jia, Travis J Nelson, Quyen Vu, Tiffany Truong, and Cliff I Stains. 2016. "Self-Assembling NanoLuc Luciferase Fragments as Probes for Protein Aggregation in Living Cells." *ACS Chemical Biology* 11 (1): 132–38. <https://doi.org/10.1021/acscchembio.5b00758>.

TITLE: Study of the evolution of H5-subtype avian influenza viruses to highly pathogenic phenotypes

ABSTRACT:

Avian influenza viruses (AIV) appear to be a significant threat to both animal and human health as certain strains can be zoonotic. Highly pathogenic avian influenza viruses (HPAIV) are characterized by the acquisition of a polybasic sequence in the hemagglutinin (HA) cleavage site, which constitutes the main virulence factor of these viruses. Every HPAIV naturally derives from low pathogenic influenza virus (LPAIV) progenitor of H5 or H7 subtype. This genetic predisposition is poorly understood, and some questions remain unanswered. Within the H5 and H7 subtypes, have all the LPAIV the same probability of evolving towards highly pathogenic phenotype? If the emergence frequency of HPAIV among LPAIV is significantly different, does the nucleotide environment of the HA cleavage site sequence influence the evolution to HPAIV?

A phylogenetic analysis of HA nucleotide sequences of H5 subtype-AIV from avian hosts combined with ARN-secondary structures predictions have been performed to reveal potential motifs. Thanks to this bioinformatic approach, it has been possible to categorize the H5-subtype HA sequences of LPAIV in several groups. A correlation is likely to be present between the presence of hairpin ARN structure close to the HA cleavage site and the evolution of specific LPAIV to HPAIV by nucleotide insertions.

Moreover, a reporter minigenome-system based on a previous work has been designed and optimized to detect nucleotide insertions. This promising system is still under the optimization process. This new experimental system could become a handy tool to detect very rare nucleotide insertion, a significant pathway to the emergence of HPAIV in poultry.

KEYWORDS: Avian influenza viruses, H5 and H7 subtypes, hemagglutinin, cleavage site, genetic evolution, insertions, minigenome, secondary RNA structures
