



**HAL**  
open science

# Développement de techniques diagnostiques pour les maladies virales émergentes et ré-émergentes

Barthelemy Margot

► **To cite this version:**

Barthelemy Margot. Développement de techniques diagnostiques pour les maladies virales émergentes et ré-émergentes. Sciences du Vivant [q-bio]. 2023. hal-04493199

**HAL Id: hal-04493199**

**<https://ephe.hal.science/hal-04493199v1>**

Submitted on 7 Mar 2024

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

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



École Pratique  
des Hautes Études



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

ÉCOLE PRATIQUE DES HAUTES ÉTUDES – PSL

Sciences de la vie et de la Terre

MÉMOIRE

Présenté par

**BARTHÉLÉMY Margot**

Pour l'obtention du Diplôme de l'École Pratique des Hautes Études – PSL

**DÉVELOPPEMENT DE TECHNIQUES DIAGNOSTIQUES POUR LES  
MALADIES VIRALES ÉMERGENTES ET RÉ-ÉMERGENTES**

Soutenu le **23 novembre 2023** devant le jury :

Pr. MATHIEU Laurence	- <b>Présidente du jury</b>
Pr. COUTINHO Pedro	- <b>Rapporteur</b>
Pr. CHARREL Rémi	- <b>Examineur</b>
Pr. COUTARD Bruno	- <b>Tuteur scientifique</b>
Pr. ARNAUD Frederick	- <b>Tuteur pédagogique</b>

**Mémoire préparé sous la direction du Pr. COUTARD Bruno :**

**Intitulé de la structure d'accueil :** Unité des virus émergents,  
UVE: Aix-Marseille Univ-IRD 190-Inserm 1207. IHU méditerranée infection, 19-21  
boulevard Jean Moulin, 13005 MARSEILLE  
**Directeur :** Pr. De LAMBALLERIE Xavier

**Et de**

**Pr. ARNAUD Frederick**

**Intitulé de la structure d'accueil :** EPHE-PSL : UMR 754 "Infections Virales et Pathologie  
Comparée (IVPC). 50 Avenue Tony Garnier 69007 Lyon

**Directrice :** Dr. Fabienne ARCHER

## Remerciements

Je tiens tout d'abord à remercier le Pr. Xavier de Lamballerie, directeur du laboratoire Unité des Virus Emergents, qui m'a accueilli au sein de son unité et m'a offert l'opportunité de poursuivre mes études et de me reconverter dans le domaine de la biologie et des maladies infectieuses. Je lui suis reconnaissante de la confiance qu'il m'a accordée.

Je souhaite aussi remercier la Pr. Laurence Mathieu, le Pr. Pedro Coutinho et le Pr. Rémi Charrel, d'avoir accepté d'évaluer mon travail et de constituer mon jury de soutenance. Je tiens à adresser un remerciement particulier au Pr. Rémi Charrel, qui est à l'origine de mon recrutement au sein de l'UVE et qui a suivi mon travail au long de la formation.

J'adresse également mes remerciements au Pr. Frédéric Arnaud, qui a accepté d'être mon tuteur pédagogique lors de cette formation, et qui s'est toujours montré très bienveillant, réactif et disponible pour répondre à la moindre de mes questions. A travers lui, je remercie l'EPHE qui offre des opportunités de formation dont j'ai pu bénéficier.

Évidemment, je remercie infiniment mon tuteur scientifique, le Pr. Bruno Coutard, sans qui ce projet et ce mémoire n'auraient jamais vu le jour. Merci de m'avoir ~~foeée~~ poussée à réaliser cette formation. Merci pour ton temps (beaucoup de ton temps), ta patience (ton courage), ta pédagogie, ta disponibilité. On fera bombance (les + de 46 ans comprendront) pour fêter ça !

Forcément, je remercie la seule et l'unique **K**arine avec un grand **K**, ma meilleure pote, ma collègue de bureau et surtout ma mère, de me supporter aussi bien dans la vie qu'au boulot. Merci d'avoir accepté que ta fille préférée vienne empiéter sur ton lieu de travail pour faire cette formation. Je pourrais écrire un livre entier pour te remercier pour tout que ça ne suffirait pas. Mes chères collègues de bureau, Gégé et Audrey, merci de m'avoir épaulée, chacune à votre manière, pendant ces trois ans. Audrey, merci pour tes précieux conseils scientifiques, bien souvent trop pointilleux pour que j'en tienne compte (t'es trop intelligente pour moi dsl); Gégé, merci de m'avoir aidé pour toute la mise en page de ce mémoire, mais surtout d'avoir été ma fidèle compagne de bouffe en tout genre (on finira grosse toutes les deux mais c'est pas grave c'était bon). Mado, merci de m'avoir déconcentrée pendant toute la rédaction de ce mémoire, et ce dès 7h du matin. Je n'oublie pas mon fidèle binôme de l'EPHE, j'ai nommé Chacha, qui m'a supporté pendant ces semaines de formations, moi et mes calculs rénaux. Merci à toutes les 5 pour ces débriefs qui n'en finissent plus, ces restos, ces sorties et ces fous rires.

Je remercie également tous les membres de l'UVE, et ceux qui m'ont aidé de près ou de loin durant ces trois années. Je pense particulièrement à Sarah et Eva, qui m'ont formée sur la plupart des techniques de ce projet.

Merci à mamie Martine, qui a corrigé l'intégralité de ce mémoire sans comprendre un mot de ce qu'elle lisait, et merci à SteF qui a essayé de le lire sans jamais réussir à dépasser la 6<sup>e</sup> page. Merci également à mon papa qui m'a toujours soutenue.

# Table des matières

<b>Remerciements</b>	<b>1</b>
<b>Liste des abréviations</b>	<b>4</b>
<b>Liste des figures</b>	<b>6</b>
<b>1. Introduction générale</b>	<b>8</b>
<b>1.1. Qu'est-ce qu'un virus ?</b>	<b>8</b>
1.1.1. Présentation générale	8
1.1.2. Cycle infectieux	8
1.1.3. Nomenclatures	10
1.1.3.1. La classification ICTV	10
1.1.3.2. La classification de Baltimore	11
<b>1.2. Notions et causes d'une émergence ou ré-émergence virale</b>	<b>13</b>
<b>1.3. Détection de virus et diagnostic moléculaire</b>	<b>17</b>
<b>1.4. Objectifs du projet de recherche</b>	<b>22</b>
<b>2. Développement, validation et production de contrôles pour le diagnostic moléculaire</b>	<b>23</b>
<b>2.1. Qu'est-ce qu'un contrôle?</b>	<b>23</b>
<b>2.2. Nature des contrôles</b>	<b>25</b>
2.2.1. Matériel biologique	25
2.2.2. Matériel synthétique	26
2.2.2.1. Matériels synthétiques nus.	26
2.2.2.2. Matériels synthétiques encapsidés	27
<b>2.3. Enjeux du projet</b>	<b>32</b>
<b>2.4. Matériels et méthodes</b>	<b>33</b>
2.4.1. Conception des gènes synthétiques	33
2.4.2. Production, quantification et validation des transcrits <i>in vitro</i> .	33
2.4.3. Production des ARN encapsidés	34
2.4.3.1. Synthèse des ArRNA	34
2.4.3.2. Lyophilisation et validation de l'ArRNA	35
2.4.3.3. Contrôle de l'homogénéité	35
2.4.3.4. Production et validation de l'ArRNA MS2	36
2.4.4. ADN Encapsidé	38
2.4.4.1. Capsides recombinantes	38
2.4.4.2. Phagémide encapsidé	38
<b>2.5. Résultats et discussion</b>	<b>40</b>

2.5.1.	Transcrits <i>in vitro</i> _____	40
2.5.1.1.	Optimisation du signal de terminaison _____	41
2.5.1.2.	Production des IVT à partir d'un produit PCR _____	43
2.5.2.	ARN encapsidés _____	45
2.5.2.1.	Purification des ARN encapsidés _____	45
2.5.2.2.	Quantification de la fuite de terminaison de transcription. _____	48
2.5.2.3.	Normalisation de la quantité de matériel génétique dans les ArRNA _____	50
2.5.2.4.	Développement d'un contrôle MS2 synthétique _____	52
2.5.3.	ADN encapsidés _____	54
2.5.3.1.	Capsides recombinantes _____	54
2.5.3.2.	Phagémide encapsidé _____	59
<b>3.</b>	<b>Développement de techniques de génotypage par endonucléase. _____</b>	<b>65</b>
<b>3.1.</b>	<b>Les différentes techniques d'identification. _____</b>	<b>65</b>
3.1.1.	Le séquençage. _____	65
3.1.2.	Le génotypage par endonucléase. _____	66
3.1.2.1.	Génotypage SARS-CoV-2 _____	68
3.1.2.2.	Les endonucléases comme outil de génotypage. _____	70
<b>3.2.</b>	<b>Enjeux du projet _____</b>	<b>72</b>
<b>3.3.</b>	<b>Matériel et méthodes _____</b>	<b>73</b>
3.3.1.	Génotypage SARS-CoV-2 _____	73
3.3.2.	Production, caractérisation et optimisation des endonucléases EndoMS _____	74
3.3.2.1.	Production des endonucléases _____	74
3.3.2.2.	Tests d'activité _____	75
<b>3.4.</b>	<b>Résultats et discussion _____</b>	<b>76</b>
3.4.1.	Génotypage SARS-CoV-2 _____	76
3.4.2.	Production d'une endonucléase recombinante. _____	76
3.4.2.1.	Résultats préliminaires. _____	76
3.4.2.2.	Tentatives d'optimisation _____	78
<b>4.</b>	<b>Conclusion générale et perspectives _____</b>	<b>82</b>

## Liste des abréviations

**ADN** : Acide Désoxyribonucléique.

**ArDNA** (Anglais, Royaume-Unis) : Armored DNA. ADN encapsidé.

**ARN** : Acide Ribonucléique.

**ARNm** : ARN messenger.

**ArRNA** (Anglais, Royaume-Unis) : Armored RNA. ARN encapsidé.

**ASFV** (Anglais, Royaume-Unis) : African Swine Fever Virus. Virus de la peste porcine Africaine.

**Bp** (Anglais, Royaume-Unis) : Base pair. Paire de bases.

**CP** (Anglais, Royaume-Unis) : Coat Protein. Protéine de capsid.

**DO** : Densité Optique.

**ds** (Anglais, Royaume-Unis) : double strand. Double brin.

**E.Coli** : *Escherichia Coli*.

**ELISA** (Anglais, Royaume-Unis) : Enzyme Linked ImmunoSorbent Assay. Dosage immuno-enzymatique.

**EQA** (Anglais, Royaume-Unis) : External Quality Assesment. Evaluation externe de qualité.

**EVA-G** (Anglais, Royaume-Unis) : European Virus Archive – Global. Archives européennes des virus – global.

**FU** (Anglais, Royaume-Unis) : Fluorescence Unit. Unité de fluorescence.

**fIMCP** (Anglais, Royaume-Unis) : Full lenght Major Coat Protein. Protéine majeure de capsid, taille complète.

**HBV** (Anglais, Royaume-Unis) : Hepatitis B Virus. Virus de l'Hépatite B.

**PCR** (Anglais, Royaume-Unis) : Polymerase Chain Reaction. Réaction en chaîne par polymérase.

**Prom** : Promoteur de transcription.

**Term** : Termineur de transcription.

**qPCR** (Anglais, Royaume-Unis) : quantitative PCR. PCR quantitative.

**QSP** : Quantité suffisante pour.

**RDB** (Anglais, Royaume-Unis) : Receptor Binding Domain. Domaine de liaison au récepteur.

**RBM** (Anglais, Royaume-Unis) : Receptor Binding Motif. Motif de liaison au récepteur.

**RBS** (Anglais, Royaume-Unis) : Ribosome Binding Site. Site de liaison au ribosome.

**RPM** : Rotation Par Minute.

**RT-PCR** (Anglais, Royaume-Unis) : Reverse Transcriptase PCR. PCR après transcription inverse.

**RT-qPCR** (Anglais, Royaume-Unis) : quantitative RT-PCR. RT-PCR quantitative.

**RUO** (Anglais, Royaume-Unis) : Research Use Only. A usage de la recherche uniquement.

**SARS-CoV-2** (Anglais, Royaume-Unis) : Severe Acute Respiratory Syndrome Coronavirus 2. Coronavirus 2 du Syndrome Respiratoire Aigu Sévère.

**SDS-PAGE** (Anglais, Royaume-Unis) : Sodium dodécylsulfate-PolyAcrylamide Gel Electrophoresis. Electrophorèse sur gel de polyacrylamide en présence de SDS.

**HIS-Tag** : étiquette poly-histidine.

**IC** (Anglais, Royaume-Unis) : Internal Control.  
Contrôle Interne.

**ICTV** (Anglais, Royaume-Unis) : International  
Committee on Taxonomy of Viruses. Comité  
International de Taxonomy des Virus.

**IgG** : Immunoglobulines de type G

**IgM** : Immunoglobulines de type M.

**IMAC** (Anglais, Royaume-Unis) : Immobilized  
Metal Affinity Chromatography.

Chromatographie d'affinité sur ions métalliques  
immobilisés

**IVT** (Anglais, Royaume-Unis) : *In Vitro*  
Transcript. Produit de Transcription *In Vitro*.

**LOD** (Anglais, Royaume-Unis) : Limit Of  
Detection. Limite de détection.

**MM** : Mis-Match.

**MP** (Anglais, Royaume-Unis) : Maturation  
Protein. Protéine de maturation.

**nt** : Nucléotide.

**NTD** (Anglais, Royaume-Unis) : N-Terminal  
Domain. Domaine N-Terminal.

**PAC** : Séquence d'encapsidation

**SEC** (Anglais, Royaume-Unis) : Size  
Exclusion Chromatography.

Chromatographie d'exclusion de taille.

**Sida** : Syndrome de l'immunodéficience  
acquise.

**ss** (Anglais, Royaume-Unis) : simple strand.  
Simple brin.

**SNP** (Anglais, Royaume-Unis) : Single  
Nucleotid Polymorphism. Polymorphisme  
d'un nucléotide.

**TEV** (Anglais, Royaume-Unis) : Tobacco  
etch virus. Virus de la gravure du tabac.

**T7prom** : Promoteur de transcription à la T7  
polymérase.

**T7Term** : Termineur de transcription T7.

**UVE** : Unité des virus Émergents.

**VoC** (Anglais, Royaume-Unis) : Variant of  
Concern. Variant préoccupant.

**VLP** (Anglais, Royaume-Unis) : Virus Like  
Particule. Particules semblables à celles d'un  
virus.

**WT** (Anglais, Royaume-Unis) : wild-type.  
Littéralement : « type-sauvage ».

## Liste des figures

Figure 1 : Schéma généralisé d'un cycle infectieux viral.....	9
Figure 2: Classification ICTV .....	10
Figure 3 : Classification de Baltimore .....	11
Figure 4 : Impact des changement d'écosystèmes sur l'émergence virale .....	14
Figure 5 : Chronologie des principaux événements du début de la pandémie de COVID-19.....	15
Figure 6 : Représentation graphique de la cinétique d'une infection virale et de la réponse humorale associée.....	17
Figure 7 : Processus de diagnostic d'une analyse par PCR en temps réel.....	19
Figure 8 : Représentation schématique de la localisation de chaque contrôle dans une plaque PCR .....	24
Figure 9 : Matrice permettant la synthèse d'un Transcrit <i>In Vitro</i> . .....	27
Figure 10 : Composants du virion MS2 et son cycle de vie .....	29
Figure 11 : Synthèse d'un ArDNA à partir d'un phagémide et d'un phage auxiliaire .....	31
Figure 12 : Présentation du design d'une séquence codant pour un transcrit <i>in vitro</i> .....	33
Figure 13 : Schéma du processus de validation d'un ArRNA .....	37
Figure 14 : Structure secondaire du T7term et mise en évidence des problèmes de terminaison de transcription.....	41
Figure 15 : Présentation des différentes séquences de terminaison de transcription et combinaisons testées.....	42
Figure 16 : Comparaison des différentes séquences de terminaison de transcription .....	43
Figure 17 : Validation de la méthode de production des Transcrits <i>In Vitro</i> .....	44
Figure 18 : Production et purification des Armored RNA.....	46
Figure 19 : Visualisation des ArRNA par microscopie électronique .....	47
Figure 20 : Design du système TaqMan .....	48
Figure 21 : Évaluation des fuites de transcription lors de la production de l'ArRNA.....	49
Figure 22 : Régressions linéaires des quantifications par RT-qPCR de l'ARN d'actine. ....	50
Figure 23 : Évaluation du test de normalisation de la quantité d'ARN d'actine dans un Armored RNA.....	51
Figure 24 : Quantification par RT-qPCR de l'ArRNA MS2 dans différents échantillons biologiques.....	53
Figure 25 : modèle d'assemblage des virophages, et fonctions des MCP.....	55
Figure 26 : Analyse par SDS-PAGE des protéines recombinantes purifiées .....	56
Figure 27 : Produits IMAC du clivage de la fMCP par les TEV protéases .....	57
Figure 28 : Analyses de l'assemblage des protéines MCPΔC et penton .....	58
Figure 29 : Expression d'une VLP G3p-HIS-Taguée à l'apex du phage. ....	59
Figure 30 : Vérification de la présence de l'ADN cible. ....	60
Figure 31 : Mise en évidence de l'encapsidation de l'ADN. ....	63
Figure 32 : Présence de particules phagiques dans les préparation d'ADN encapsidés. ....	64
Figure 33 : Représentation schématique du principe du génotypage par endonucléase de restriction. ....	67
Figure 34 : Le RBM : région clef pour l'entrée du virus et la séroneutralisation .....	69

Figure 35 : Structure de l'EndoMS issue de <i>Thermococcus kodakarensis</i> .....	71
Figure 36 : Chromatographie d'exclusion de taille d'EndoMS.....	76
Figure 37 : Fonctionnalité d'EndoMS.....	77
Figure 38 : Identification des résidus impliqués dans l'interaction d'EndoMS avec les mésappariements .....	78
Figure 39: Influence des mutations sur l'efficacité et la spécificité de clivage d'EndoMS .....	79
Figure 40: Les mutations n'induisent pas d'augmentation d'efficacité et ne réduisent pas la spécificité d'EndoMS.	80

# **1. Introduction générale**

## **1.1. Qu'est-ce qu'un virus ?**

### 1.1.1. Présentation générale

D'origine latine, le mot virus signifie littéralement « venin » ou « poison ». L'existence des virus a été découverte à la fin du XIXe siècle, lorsqu'une maladie a dévasté des cultures de tabac au Pays-Bas : le virus de la mosaïque du tabac. Plusieurs expériences ont démontré que cette maladie était induite par un agent pathogène dont l'existence était encore inconnue (1). L'extrait de feuille malade conserve ses propriétés infectieuses après filtration dans un filtre de Chamberland, filtre ayant la propriété de retenir les bactéries. Cet agent pathogène sera nommé pour la première fois « virus » par Martinus Beijerinck. Le premier virus ayant été découvert chez l'homme est le virus de la Fièvre Jaune, en 1901. Au cours des siècles, les virus ont été à l'origine de nombreuses maladies infectieuses et sont considérés par l'Organisation Mondiale de la Santé (OMS) comme l'une des principales causes de morbidité et de mortalité à travers le monde. Aujourd'hui, 219 espèces de virus pathogènes pour l'Homme sont recensées (2). La grippe, la varicelle ou le Sida (Syndrome de l'immunodéficience acquise) sont des exemples de maladies causées par des infections virales.

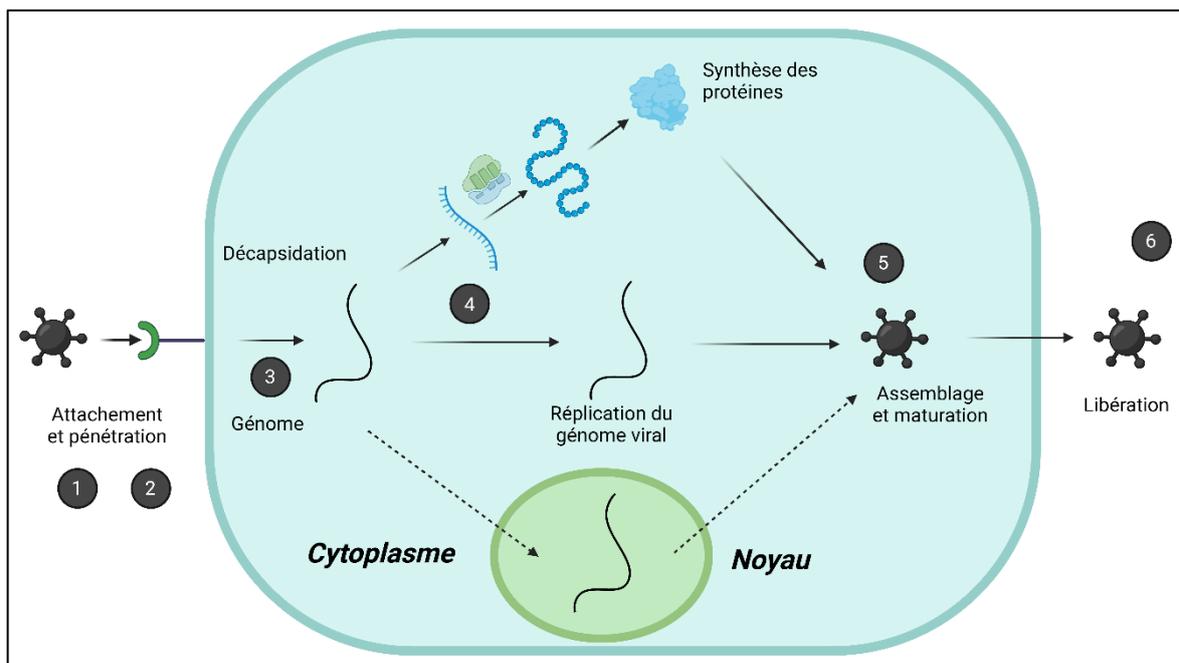
Un virus est un agent infectieux de taille comprise entre dix et plusieurs centaines de nanomètres. Ces micro-organismes sont capables d'infecter tout type d'êtres vivants : humains, animaux, plantes, bactéries... Certains virus peuvent même infecter d'autres virus (3). Ce sont des parasites qui ne peuvent pas se multiplier sans utiliser la machinerie cellulaire d'un hôte. Ils se propagent selon différents modes de transmission : la transmission aérienne, parentérale ou encore de contact. Le mode de transmission d'un virus dépend notamment de son mécanisme d'entrée dans la cellule hôte, et donc de sa structure.

### 1.1.2. Cycle infectieux

La réplication virale est le processus par lequel les virus dupliquent leur matériel génétique. Le cycle infectieux est un mécanisme complexe et propre à chaque virus, mais comporte des étapes universelles, qui peuvent être synthétisées de la manière suivante (**Figure 1**) :

- 1- Attachement : Des protéines présentes à la surface du virus reconnaissent et se lient spécifiquement à des récepteurs de la cellule hôte.
- 2- Fusion : une fois attachés, les virus fusionnent leur membrane externe avec la membrane cellulaire, permettant aux génomes viraux d'entrer dans la cellule.

- 3- Décapsidation : le virus est décapsidé afin de libérer le génome viral à l'intérieur de la cellule. La capside virale est dégradée afin de rendre le matériel génétique du virus accessible.
- 4- Réplication du génome viral et synthèse des protéines : Le virus détourne la machinerie cellulaire pour la réplication et l'expression du génome viral. Virus à génome ADN : le génome viral est répliqué et transcrit en ARN messagers, et les protéines virales sont synthétisées à partir de ces ARN. Virus à génome ARN : Le génome viral est répliqué, et les protéines virales sont synthétisées directement à partir des ARN viraux.
- 5- Assemblage des particules virales : les protéines synthétisées et les génomes viraux sont assemblés afin de former des particules virales complètes.
- 6- Libération : les particules virales néo-formées sont libérées de la cellule et peuvent infecter d'autres cellules.



**Figure 1 : Schéma généralisé d'un cycle infectieux viral**

*Ce schéma illustre les étapes « typiques » de la réplication d'un virus infectant des cellules eucaryotes. Les numéros correspondent aux étapes du cycle infectieux détaillées dans le texte ci-dessus. Schéma généré avec BioRender.*

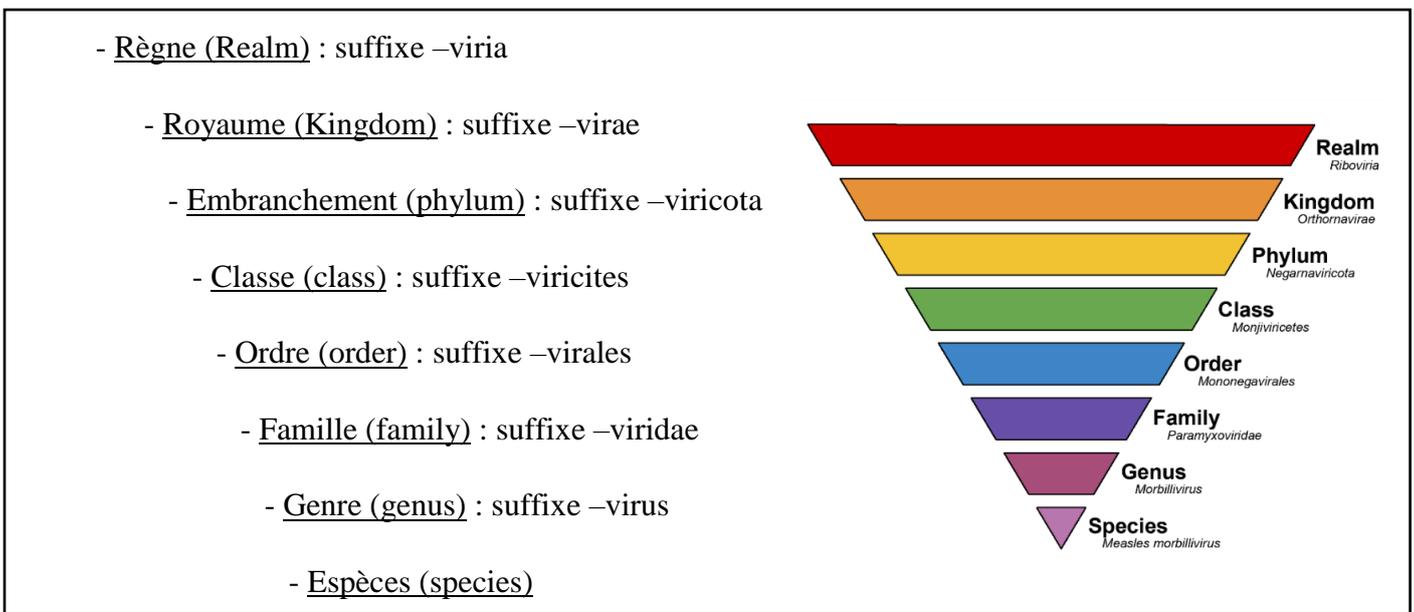
En bref, la réplication virale est un processus complexe qui permet aux virus de se multiplier et de se maintenir dans une population. La compréhension des principes généraux de la réplication virale est essentielle à la prévention et à la lutte contre les infections virales.

### 1.1.3. Nomenclatures

Depuis la découverte des virus, de nombreuses initiatives ont eu pour objectif de classer ces organismes. Ces classifications ont pu se faire sur la base structurale des virus : la présence ou non de bicouche lipidique, une organisation hélicoïdale ou icosaédrique, le type de mutation génétique, la séquence du matériel génétique, le cycle de réplication... À ce jour, deux classifications sont les plus communément utilisées : l'organisation taxonomique ICTV (International Committee on Taxonomy of Viruses) et la classification de Baltimore.

#### 1.1.3.1. La classification ICTV

La classification ICTV est le modèle majoritairement utilisé dans le monde pour classer les virus. Proposé en 1962, cette classification est basée sur la comparaison de diverses caractéristiques qui décrivent le virus et peuvent ensuite être utilisés pour distinguer un virus d'un autre. Ces caractères comprennent la composition moléculaire du génome, la structure de la capsid du virus et le fait qu'il soit enveloppé ou non, le programme d'expression génétique utilisé pour produire les protéines du virus, la gamme d'hôtes, la pathogénicité et la similarité de séquence (4). Sur la base de cette évaluation des caractères, une relation hiérarchique qui regroupe les virus ayant des propriétés similaires a pu être établie. Les propriétés qui définissent les taxons de niveau supérieur sont partagées par tous les taxons de niveau inférieur auxquels ils appartiennent. Les virus sont classés selon huit sous catégories (**Figure 2**):



**Figure 2: Classification ICTV**

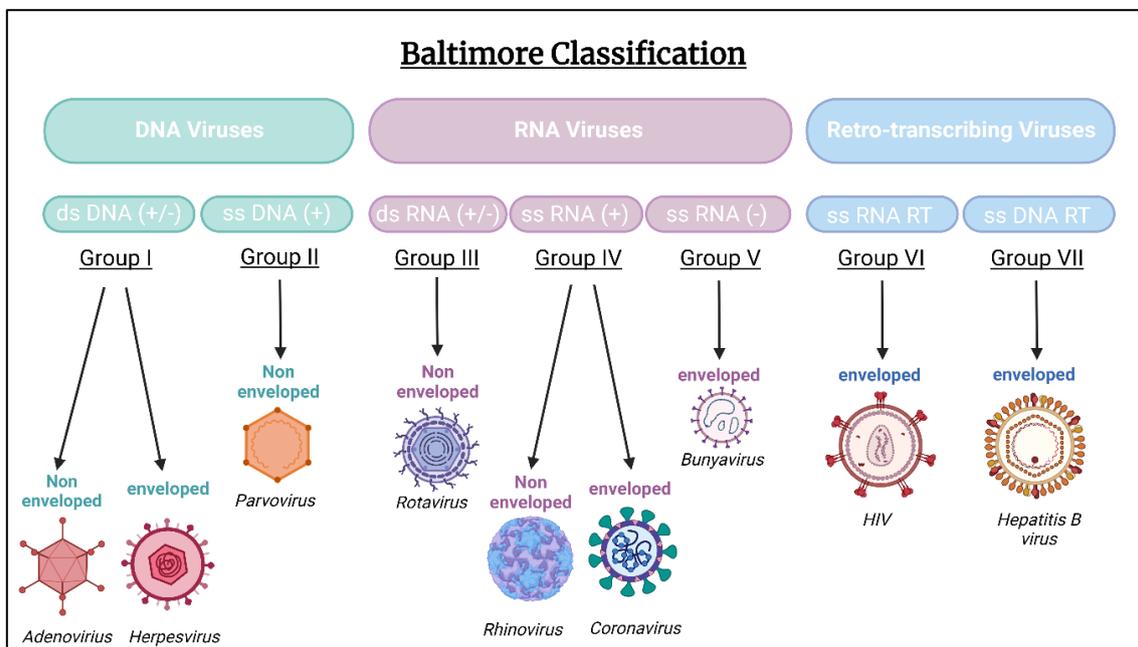
Diagramme de la classification de Baltimore et les suffixes associés. Diagramme récupéré du site [ICTV.global](http://ICTV.global)

En 2021, le Comité International de Taxonomie des virus recense 10434 espèces de virus capables d’infecter les animaux, les plantes, les champignons, les bactéries et les archées, même si à ce jour >99% du virome terrestre n’est pas clairement identifié.

Depuis 2020, l’ICTV a commencé à utiliser une nouvelle classification des virus basée sur la phylogénie, plutôt que sur des caractéristiques physiques et biologiques. Cette nouvelle classification est basée sur l’analyse de séquences génomiques complète des virus. Cette approche permet une meilleure résolution et précision dans la classification des virus, et permet également une classification plus flexible pour les virus émergents et les virus à génomes complexes.

### 1.1.3.2. La classification de Baltimore

Un virus est composé d’un génome, constitué d’un ou plusieurs brins d’Acides Désoxyribonucléiques (ADN) ou d’Acides Ribonucléiques (ARN). On distingue un virus ADN d’un virus ARN, simple brin ou double brin, de polarité positive ou négative, enveloppé ou non. C’est selon ce modèle que la classification de Baltimore est établie (5). Selon cette classification, tous les virus existants sont séparés en sept groupes, répartis dans trois grandes catégories : les virus ADN, les virus ARN et les rétrovirus (6) (**Figure 3**).



**Figure 3 : Classification de Baltimore**

*Schéma de la classification de Baltimore, illustré par des virus appartenant à chaque groupe. Échelle des particules virales non respectée. Figure générée avec BioRender. ds : double strand ; double brin. ss : simple strand, simple brin. RT : Retro-transcribing.*

Groupe I : groupe des virus à ADN double brin. Leur génome est généralement linéaire et non segmenté, d'une taille comprise entre 5 et 2,500 kilobases. Les virus de ce groupe se répliquent à l'intérieur du noyau de la cellule hôte (à l'exception des poxvirus). Ces virus ont besoin de l'ADN polymérase de la cellule hôte pour se répliquer.

Groupe II : comprend les virus à ADN simple brin, qui encapsident de l'ADN simple brin, qui sera par la suite répliqué et exprimé par l'intermédiaire d'un ADN double brin.

Groupe III : il s'agit du groupe des virus à ARN double brin. Comme la majorité des virus à ARN, ils se répliquent dans le cytoplasme. L'ARNm de ces virus représente une transcription asymétrique du génome. Les virus connus de ce type ont plusieurs morceaux d'ARN double brin –on parle de génome segmenté- chacun d'entre eux contenant une information pour la synthèse d'une seule protéine ou présentent des cadres de lectures alternatifs.

Groupe IV : cette classe comprend des virus à ARN simple brin de polarité positive, donc l'ARNm est identique à l'ARN viral (en nucléotide). Ces virus doivent dans un premier temps produire un brin d'ARN de polarité négative pour pouvoir le re-transcrire en brin positif, qui servira soit de génome soit d'ARNm.

Groupe V : ce groupe est constitué de virus dont le génome est constitué d'un ARN simple brin de polarité négative, dont la séquence de base est complémentaire de celle de l'ARN messager. La synthèse de leur ARNm implique donc la transcription d'un ARN simple brin de polarité positive, le génome ne pouvant pas coder directement les protéines virales.

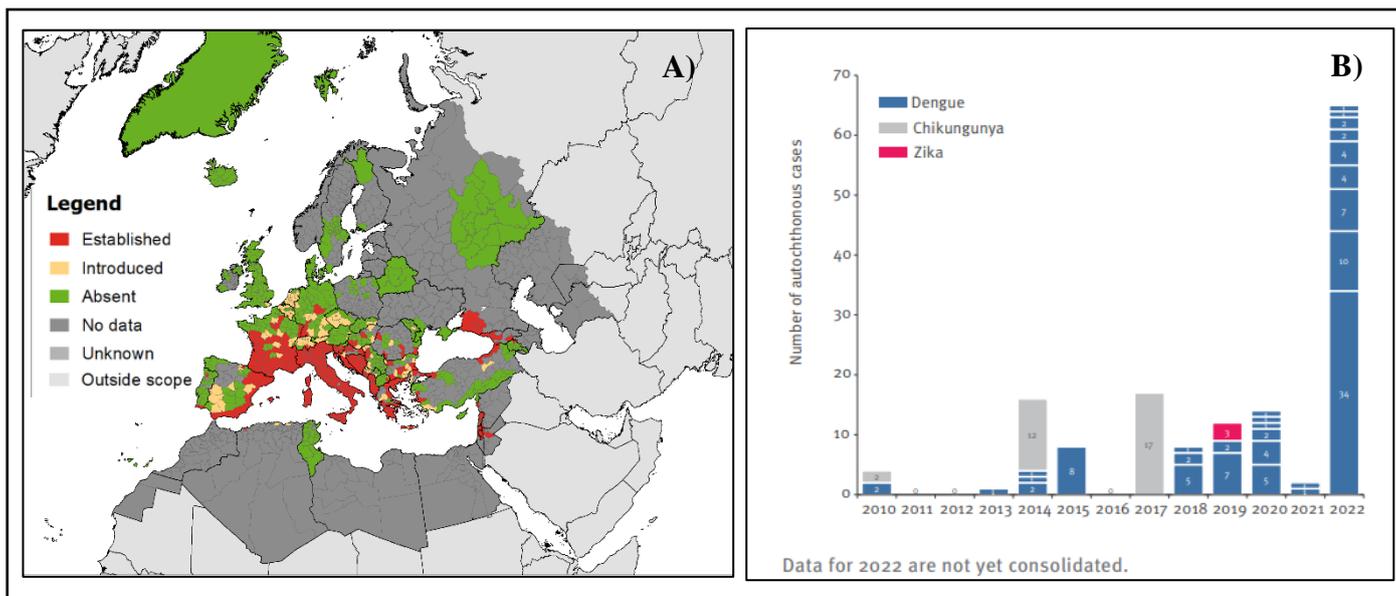
Groupe VI : comprend les virus ARN à transcription inverse (rétrovirus). La présence d'une transcriptase inverse permet la conversion de l'ARN de polarité positive en ADN.

Groupe VII : virus qui contiennent un génome à ADN double brin (bien que l'un des brins soit généralement incomplet) et qui utilisent une transcriptase inverse pour se répliquer par l'intermédiaire d'un ARN intermédiaire (ex : virus de l'hépatite B).

## 1.2. Notions et causes d'une émergence ou ré-émergence virale

La notion d'émergence et de réémergence virale est un concept important en santé publique. Nous pouvons utiliser le terme de « virus émergents » pour désigner des virus qui sont soit récemment apparus dans une population, soit étendent rapidement leur propagation, entraînant une augmentation correspondante de cas d'infection (7). Ces phénomènes d'émergence et de réémergence virale sont une préoccupation majeure pour les autorités de santé publique et nécessitent une surveillance continue pour prévenir leur propagation et leur impact sur la santé humaine et animale. La société civile est de plus en plus fréquemment confrontée aux dangers de l'émergence de nouveaux agents pathogènes. Les mécanismes de ces émergences sont multiples et certaines sont déjà bien identifiées :

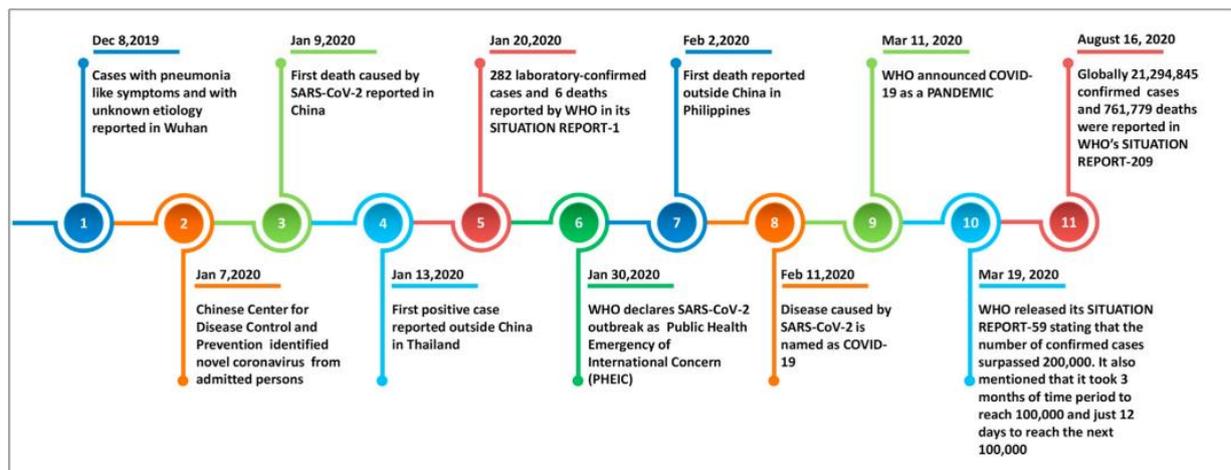
Le changement des écosystèmes : les changements globaux de facteurs environnementaux (précipitations, températures, vent, humidité...) ont un impact conséquent sur l'émergence et la réémergence d'agents pathogènes. Les modifications écologiques, en particulier la déforestation et l'urbanisation, le changement climatique, sont également directement liées aux mécanismes d'émergences. Ces modifications à l'échelle mondiale ont conduit par exemple, en conjonction avec l'augmentation des transports internationaux, à l'apparition de nouvelles espèces de moustiques telles que celles du type *Aedes* en Europe, là où elles ne circulaient auparavant essentiellement que dans des régions au climat tropical (8). Ces moustiques sont vecteurs de nombreux virus, notamment du virus du Chikungunya, du Zika et de la Dengue. Leur présence conduit à des cas autochtones réguliers d'infections par ces virus en Europe (**Figure 4**) (8–10).



#### Figure 4 : Impact des changements d'écosystèmes sur l'émergence virale

*A) Distribution des Aedes albopictus en Europe (Mars 2021 (10)). B) Nombre de cas autochtones de Dengue, Chikungunya et Zika recensés en France métropolitaine (9). La Figure A) a été rognée. La figure complète est disponible en Annexe 1.*

La mobilité mondiale : les arrivées internationales (immigration) ont été multipliées par 56 entre 1950 et 2018 pour atteindre aujourd'hui 1,4 milliards de voyages internationaux par an. Les voyages internationaux et le commerce mondial sont des moteurs importants de l'émergence de nouveaux agents pathogènes exotiques. La récente pandémie de SARS-CoV-2 illustre parfaitement l'importance de la mobilité mondiale dans l'émergence et la propagation virale (11). Le premier cas a été détecté le 8 décembre 2019 à Wuhan en Chine. A peine plus d'un mois après, le premier cas positif est reporté à l'extérieur de la Chine, en Thaïlande. Le 11 mars 2020, l'OMS déclare le COVID-19 comme pandémie (**Figure 5**).



**Figure 5 : Chronologie des principaux événements du début de la pandémie de COVID-19**

*Frise chronologique des événements marquant de la pandémie de COVID-19, de décembre 2019 à août 2020. Figure extraite de l'article (11)*

Le bioterrorisme : le risque de diffusion délibérée d'agents pathogènes existe. Les attaques bioterroristes peuvent concerner des pathogènes humains, vétérinaires ou végétaux, causer des pertes économiques, endiguer les systèmes de transports mondiaux et déséquilibrer le commerce international. Le bioterrorisme n'étant à ce jour pas ou très occasionnellement une des causes de l'émergence, il reste un sujet d'attention pour les agences nationales et internationales ; un exemple : après les attaques terroristes du World Trade Center en 2001, des enveloppes contenant des spores de *Bacillus anthracis* ont été envoyées à des entreprises de presse et à des fonctionnaires, faisant treize malades et trois décès (même si dans ce cas précis il s'agissait d'une bactérie et non pas d'un virus) (12, 13).

Face à ces différents mécanismes d'émergence et de ré-émergence, la mise en place d'un système « Preparedness and Response » (littéralement, Préparation et Réponse) est essentiel pour lutter efficacement contre la propagation des différentes infections virales et limiter leur gravité/mortalité. Les actions mises en œuvre peuvent prendre différentes formes, et couvrent, entre autres, la prévention, la détection et les traitements. Par exemple, le port du masque pour des infections virales respiratoires est un moyen efficace de préparation et réponse à l'émergence, pour éviter de se contaminer ou de contaminer les autres (14). En complément du port du masque, d'autres mesures sont recommandées, comme une distanciation physique, un lavage régulier des mains ainsi qu'une aération suffisante des lieux communs (15). Pour limiter les contaminations virales dans les cas où le vecteur est le moustique, l'utilisation de protections physiques, comme des vêtements longs ou des moustiquaires, ou encore des répulsifs, sont des

moyens efficaces de protection individuelle. La démoustication ainsi que la réduction de gîtes larvaires (lieux de ponte des moustiques) permettent une lutte-anti vectorielle efficace (16).

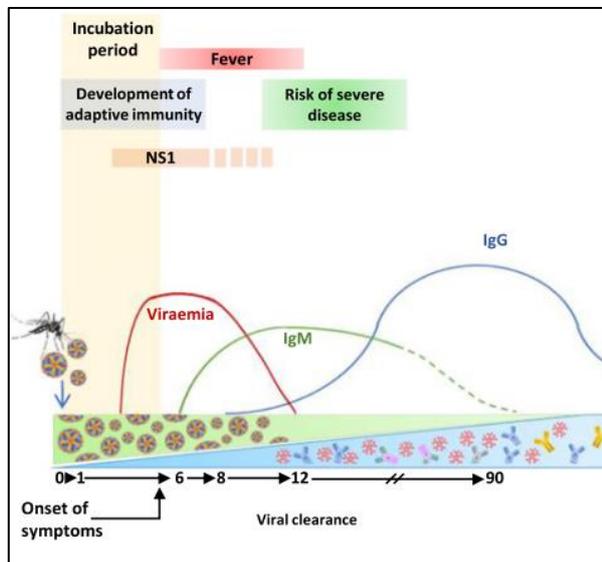
En plus des moyens de protection physique cités plus haut, des moyens scientifiques peuvent être mis en place, comme l'utilisation d'un traitement ou la vaccination. La vaccination peut être utilisée de manière préventive, pour éviter la ré-émergence de certains pathogènes, comme pour la rougeole (<https://www.who.int/fr/news-room/fact-sheets/detail/measles> (17)), ou pour lutter contre une épidémie en cours, comme pour la pandémie de SARS-CoV-2.

La liste des moyens de préparation et de réponse à l'émergence citée ci-dessus n'est évidemment pas exhaustive, car ces moyens sont en grande partie spécifiques à chaque épidémie et à chaque virus. Cependant, dans tous les cas, il est primordial d'identifier précisément quel virus est responsable d'une épidémie, afin d'appréhender au mieux sa dynamique et d'adapter la réponse à l'émergence. Pour cela, le développement et le déploiement de tests de détection directe, moléculaire ou antigénique, est une des actions prioritaires de la réponse à l'émergence.

### 1.3. Détection de virus et diagnostic moléculaire

Le diagnostic est crucial dans le cas des virus émergents pour plusieurs raisons. Tout d'abord, un diagnostic rapide et précis peut aider à identifier les cas de virus émergents et à mettre en place des mesures de prévention et de contrôle pour éviter la propagation de la maladie. Cela est particulièrement important dans le cas de virus qui se propagent rapidement, tels que les virus respiratoires émergents comme le SARS-CoV-2 par exemple. De plus, le diagnostic permet de réaliser des enquêtes de prévalence et d'incidence, et peut aider à comprendre la dynamique de l'émergence d'un virus en fournissant des informations sur les modes de transmission, les symptômes et les facteurs de risques. Cette compréhension est essentielle pour guider la mise en place de mesures de prévention efficaces, comme la mise en quarantaine, la vaccination ou la mise en place de mesures d'hygiène appropriées.

Plusieurs approches permettent de détecter directement ou indirectement la présence d'un agent pathogène viral. L'approche indirecte consiste à détecter une réponse de défense de l'hôte. Il peut s'agir par exemple de la présence d'anticorps spécifiques, témoins d'une infection passée ou en cours. La détection directe consiste quant à elle à détecter des éléments de la particule virale, comme des antigènes, des protéines virales ou du matériel génétique.



**Figure 6 : Représentation graphique de la cinétique d'une infection virale et de la réponse humorale associée.**

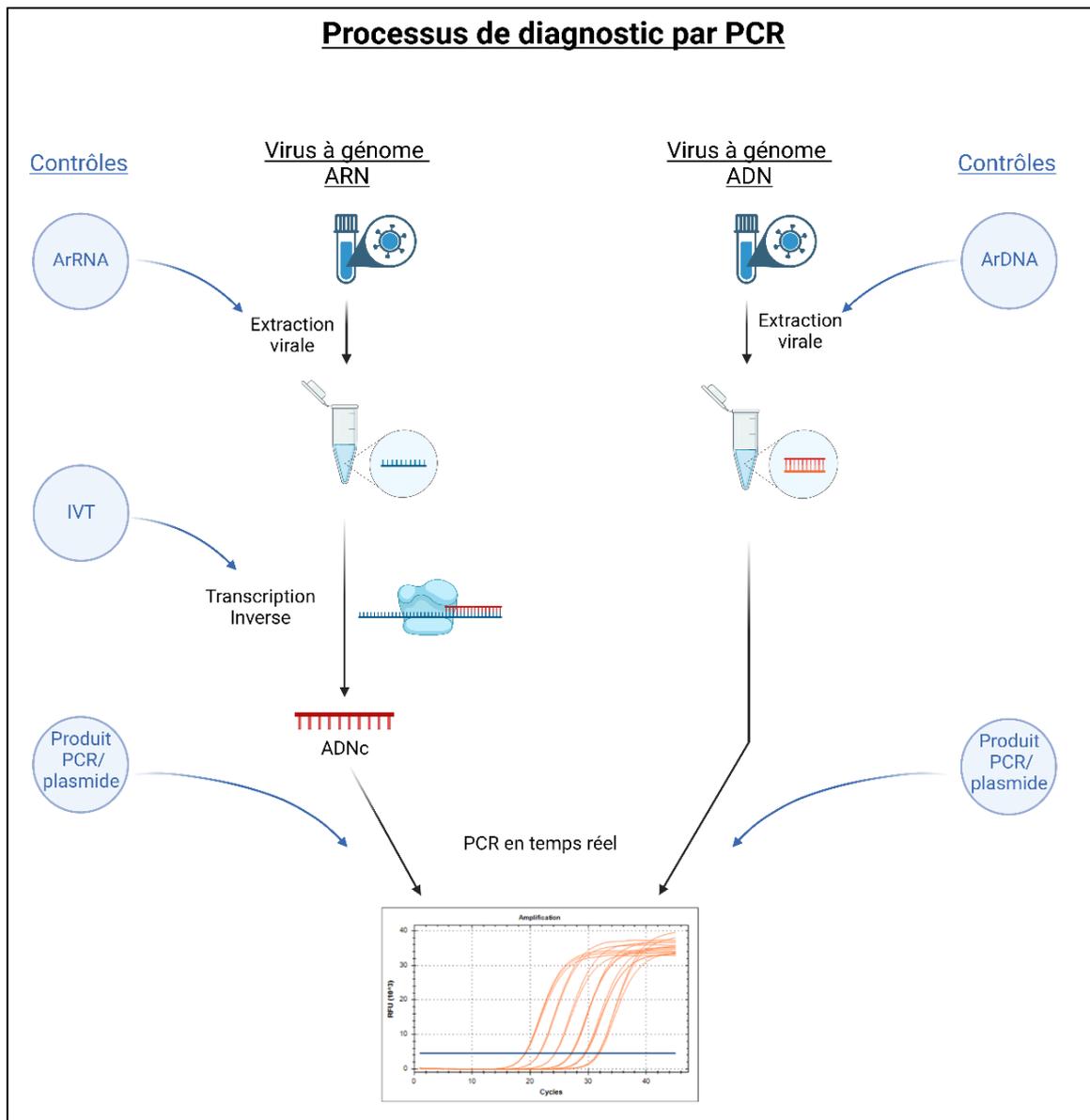
*Figure extraite de l'article (18)*

virus infectieux mesuré dans le sang de l'hôte infecté. Pour les infections virales ne présentant pas de virémie (c'est-à-dire pas de présence de virus infectieux dans le sang), nous parlerons de charge virale. Les Immunoglobulines M (IgM) sont les anticorps qui apparaissent en premier.

La **Figure 6** présente un exemple de la cinétique d'une infection virale – un flavivirus (comme le virus de la dengue) dans ce cas -, ainsi que la production d'anticorps, illustrée par une infection type d'un arbovirus (virus transmis par piqûre d'arthropode, Dictionnaire Larousse) (18). La protéine NS1 (antigène) apparaît en premier au cours de l'infection virale. Il s'agit d'un marqueur viral précoce de l'infection aiguë (la présence de cet antigène est spécifique aux infections aux flavivirus). La période de virémie correspond à la période durant laquelle le virus est détectable dans le sang ; la virémie correspond à la quantité de

Ce sont des marqueurs d'une infection en cours ou récente. Les Immunoglobulines G (IgG) apparaissent dans un second temps et sont des marqueurs d'une infection passée. Il est important de connaître cette cinétique avant de réaliser un test de détection : il faut adapter le test en fonction de ce que l'on veut détecter, et l'adapter en fonction de la date à laquelle l'échantillon a été récolté.

Les techniques de détection peuvent donc être divisées en deux groupes : la détection indirecte et la détection directe. La détection indirecte permet d'identifier des éléments résultants d'une infection virale induisant une réaction immunitaire chez l'hôte. Les techniques permettant de réaliser ce type de détection ne seront pas détaillées ici, car elles n'ont pas été l'objet d'étude dans le cadre de mon diplôme. Néanmoins, on peut citer : i) certains types de tests ELISA, permettant la détection d'une réaction antigène-anticorps ([Annexe 2](#)) ; ii) la séroneutralisation, permettant de quantifier le taux d'anticorps neutralisants d'un virus donné ; ainsi que iii) la technologie Luminex™ (19) : s'appuyant sur le principe de la cytométrie en flux, elle allie l'utilisation de microsphères fluorescentes et une double lecture après excitation par deux lasers, permettant la détection d'anticorps. La détection directe permet quant à elle d'identifier la présence d'un élément de l'agent infectieux. Elle peut reposer sur la détection du matériel génétique de l'agent infectieux, ou de l'un de ses antigènes. Parmi les tests de détection directe, nous pouvons citer les tests immunochromatographiques, permettant la détection d'un antigène viral, ou la technique de Réaction en Chaîne par Polymérase (Polymerase Chain Reaction, PCR) couplée ou non à une étape de transcription inverse selon le type de matériel génétique à détecter. Dans la suite de ce rapport, seuls les tests moléculaires ayant fait l'objet de mon diplôme seront traités. La **Figure 7** représente le déroulé du processus d'une analyse diagnostique par Real Time-PCR.



**Figure 7 : Processus de diagnostic d'une analyse par PCR en temps réel**

*Représentation schématique du processus d'analyse par PCR d'un génome viral ARN ou ADN, avec les contrôles positifs permettant de contrôler chaque étape du processus diagnostique. La nature ainsi que le fonctionnement de ces contrôles sont détaillés dans la section 2 de ce mémoire. ArRNA : Armored RNA. IVT : Transcrit in vitro. ADNc : ADN complémentaire. Figure générée avec BioRender.*

Comme nous l'avons vu précédemment avec la classification de Baltimore, un virus est composé d'un génome, ADN ou ARN, contenu dans une capside. La technique de PCR sera utilisée pour détecter un génome viral qui a été purifié à partir de l'échantillon biologique ou remis en culture, et qui est réalisée après une étape d'extraction, pour laquelle il existe différentes techniques (colonnes, billes de silices, etc.). La PCR repose sur l'utilisation d'une paire d'amorces complémentaires à la séquence virale de part et d'autre d'une séquence cible et de l'utilisation d'une polymérase. Il est ainsi possible d'obtenir un nombre conséquent de

copies d'une séquence cible donnée. La PCR est composée de trois étapes, qui sont la dénaturation des brins d'acides nucléiques, l'appariement des amorces sur ces brins et la polymérisation à partir des amorces par la polymérase. Ce processus peut être répété jusqu'à quarante-cinq fois, jusqu'à l'obtention de millions de copies identiques d'un même gène (20). Cette méthode est adaptable pour l'amplification d'une séquence d'ARN en ajoutant une étape préalable de transcription inverse : la Reverse Transcriptase-PCR (PCR après Transcription Inverse, RT-PCR). Durant cette étape, le simple brin ARN est converti en simple brin ADN complémentaire par une transcriptase inverse de rétrovirus. L'ADN complémentaire produit peut ensuite être amplifié par une PCR classique (21). Cette méthode est largement utilisée en virologie en combinaison avec des techniques utilisant une méthode de détection/quantification par fluorescence, telles que les technologies TaqMan ou SybrGreen, qui permettent de quantifier le nombre de copies d'ARN présent dans l'échantillon biologique (RT-PCR quantitative, RT-qPCR).

Dans des conditions définies, le nombre de copies d'ARN peut être proportionnel à la quantité de fluorescence induite par la PCR. La valeur de seuil correspondant au nombre de cycles de PCR nécessaires à l'apparition du signal de fluorescence est nommée CT (pour Cycle Threshold en anglais). En comparant une gamme étalon et le moment où l'augmentation exponentielle commence, il est alors possible de quantifier précisément la quantité initiale d'acides nucléiques présents dans l'échantillon (20), on parle alors de quantification absolue. Habituellement, des transcrits *in vitro* sont utilisés pour réaliser cette gamme d'étalonnage.

Les technologies utilisant un signal de fluorescence permettant de détecter ou quantifier l'amplification principalement utilisées sont donc :

La technologie SybrGreen: le SybrGreen est une molécule se fixant au petit sillon de l'ADN double brin. Libre, le SybrGreen n'émet pas de fluorescence, mais une fois apparié à l'ADN et excité, il émet une fluorescence pouvant être détectée par l'appareil d'enregistrement de la PCR. A chaque cycle PCR, les molécules de SybrGreen se fixent à chaque double brin d'ADN néoformés et émettent une fluorescence, qui est mesurée à chaque cycle ([Annexe 3](#)). Cette technique permet donc de réaliser une PCR quantitative en s'affranchissant de l'usage d'une sonde spécifique à la région cible.

La technologie TaqMan: cette technologie s'appuie sur une réaction qui utilise la Taq polymérase et une sonde spécifique de la cible marquée par un fluorophore et un quencher, qui émettent à deux longueurs d'ondes différentes. La sonde est construite pour s'hybrider

spécifiquement sur la région ADN cible entre les deux amorces de PCR (22) ([Annexe 3](#)). Dans le cadre de la détection de virus, la technologie TaqMan va permettre d'amplifier une région du génome viral afin de la détecter par fluorescence. Lorsque la réaction de PCR est en cours, la Taq polymérase copie l'ADN cible. Lorsqu'elle atteint la sonde TaqMan, elle la dégrade, séparant ainsi le fluorophore du quencher et libérant un signal de fluorescence. Au cours des cycles de PCR, le signal de fluorescence augmente et est proportionnel à la quantité de produits PCR polymérisés, permettant de réaliser une analyse quantitative. En mesurant la fluorescence à chaque cycle d'amplification, il est ainsi possible de déterminer la quantité initiale d'ADN ou d'ARN dans l'échantillon.

#### **1.4. Objectifs du projet de recherche**

L'Unité des Virus Émergents (UVE) travaille depuis 1995 dans le domaine du diagnostic, de la préparation et des réponses aux virus émergents et ré-émergents d'importance médicale représentant des enjeux de santé publiques majeurs. Au sein de PREPMedVet (Projet ANR, Préparation et intervention d'urgence lors d'émergences de pathogènes d'intérêt médical et/ou vétérinaire (23)), l'UVE doit prendre en charge le volet de diagnostic moléculaire rapide des pathogènes viraux par la technologie d'amplification génique quantitative en temps réel. En particulier, l'Unité a pour mission de développer une filière de produits de référence pour le diagnostic moléculaire, incluant des témoins positifs synthétiques de nouvelle génération. Ces témoins positifs sont un élément crucial dans la stratégie de réponse mise en place et il existe une très forte demande de matériel certifié pour de nombreux pathogènes dits « négligés » et zoonotiques, et également pour le développement et la mise à disposition rapide de matériel de référence en cas d'émergence d'un nouvel agent. Ces témoins permettent de s'assurer du bon fonctionnement des manipulations (extraction et (RT) PCR) et doivent pouvoir être distingués des génomes des agents pathogènes proprement dits afin de permettre la détection d'éventuelles contaminations et de prévenir des résultats faussement positifs. L'UVE a massivement investi dans le développement de technologies et dans une plateforme permettant de préparer à façon, et par des techniques de synthèse *in vitro*, des témoins positifs quantifiés de nouvelle génération adaptés aux exigences diagnostiques.

L'enjeu principal de mon projet de diplôme EPHE est d'améliorer notre préparation à une émergence ou réémergence de certains virus, afin de permettre leur diagnostic de masse dans les meilleures conditions. Pour cela, mon projet de recherche et développement a consisté dans un premier temps à mettre en place et développer une plateforme de production de témoins positifs pour le développement des tests diagnostiques, en améliorant les procédés de design et de production pour permettre une production rapide à façon en cas d'émergence d'un nouvel agent. Dans un second temps, j'ai élaboré de nouveaux tests diagnostiques permettant le contrôle de l'émergence de nouveaux variants d'un pathogène donné.

Ce manuscrit se composera de deux parties : la première abordera le développement des contrôles diagnostiques, et la deuxième les techniques de génotypage développées au cours de ce projet. Il se conclura par un chapitre de discussion générale et perspectives.

## **2. Développement, validation et production de contrôles pour le diagnostic moléculaire**

Il est avéré qu'au moins 80% des directives visant à établir un diagnostic ou à gérer une maladie chez un patient résultent d'un résultat de tests de laboratoire (24). Il est donc primordial que les résultats du test diagnostique et son interprétation soient fiables. Ainsi, il convient de contrôler chaque étape du procédé qui conduit à la détection de la présence d'un agent pathogène afin d'éviter le rendu de résultats faussement positifs ou faussement négatifs.

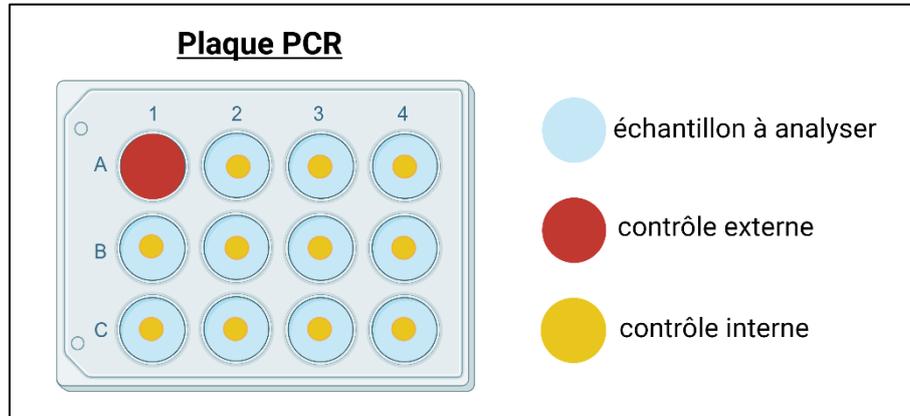
Les tests diagnostiques se basant sur la détection du génome viral sont aujourd'hui largement utilisés. Comme tout autre test diagnostique, sa robustesse doit être évaluée : une erreur de manipulation, le taux d'efficacité de l'enzyme utilisée ou encore la présence de substances inhibitrices dans l'échantillon qui peuvent induire des résultats erronés. Certains réactifs utilisés pour l'extraction du génome viral peuvent par exemple inhiber des réactions enzymatiques, ou certaines substances naturellement présentes dans les échantillons peuvent altérer les protocoles d'amplification (25). L'utilisation de contrôle est donc primordiale pour valider ou invalider le résultat d'un test diagnostique.

### **2.1. Qu'est-ce qu'un contrôle?**

Un contrôle est destiné à valider ou non le bon fonctionnement d'un test. Avant de valider une expérience quelle qu'elle soit, il convient dans un premier temps de vérifier la réponse du contrôle utilisé face à ce test, et s'assurer qu'elle soit cohérente. En effet, il est d'usage d'utiliser deux types de contrôles : les contrôles positifs et les contrôles négatifs. Un contrôle positif devra alors donner une réponse positive au test auquel il est soumis, là où le contrôle négatif ne doit pas en donner. L'utilisation de contrôle dans chacun des tests de laboratoire est obligatoire pour s'affranchir de l'obtention de résultats faussement positifs ou faussement négatifs. Il est donc nécessaire d'introduire à la fois un contrôle positif et un contrôle négatif dans chaque expérience. Dans ce travail, nous allons nous focaliser uniquement sur les contrôles destinés à une analyse par PCR.

Un contrôle négatif est un contrôle ne contenant pas l'analyte d'intérêt. Dans notre cas, n'importe quelle matrice peut servir de contrôle négatif, tant qu'elle ne contient pas la séquence à amplifier par PCR (par exemple de l'eau, un autre virus que celui ciblé, etc.). Un contrôle positif doit quant à lui contenir une séquence en acides nucléiques pertinente pour le système d'amplification PCR utilisé.

Les contrôles peuvent aussi être divisés en deux sous catégories : les contrôles internes et les contrôles externes. Un contrôle externe est un contrôle qui n'est pas mélangé à l'échantillon à analyser, il est ajouté à l'analyse de manière indépendante, alors que le contrôle interne est ajouté à l'échantillon (**Figure 8**).



**Figure 8 : Représentation schématique de la localisation de chaque contrôle dans une** *Figure générée avec BioRender.*

Dans notre cas, les contrôles positifs internes sont ajoutés dans tous les échantillons avant analyse et ne contiennent pas la même séquence en acides nucléiques que celle recherchée dans l'analyse. Ces contrôles seront donc amplifiés avec un système PCR différent de l'analyse principale. Ils sont nécessaires pour contrôler les étapes d'extraction de l'ADN ou de l'ARN viral, de transcription inverse et d'amplification afin de détecter les résultats faussement négatifs résultant d'une inhibition de la PCR ou d'un blocage de l'amplification (26). Un contrôle externe permet quant à lui seulement de s'assurer du bon déroulement de l'étape de l'analyse dans lequel il est utilisé. Ces contrôles peuvent avoir différents niveaux de caractérisation : ils peuvent être seulement qualitatifs, ou peuvent être à la fois qualitatifs et quantitatifs. Lorsqu'un contrôle est suffisamment pur et homogène, il peut être utilisé comme étalon afin de quantifier le titre viral des échantillons analysés. Si un contrôle est correctement produit, qualifié, quantifié, reproductible et stocké, il peut être utilisé comme élément de comparaison de performance intra et inter-laboratoire : on parle alors de matériel de référence (27).

Il existe un autre type de contrôles, appelés « contrôles externes de qualité » (en anglais, External Quality Assessment (EQA)), qui sont des matériels de référence permettant de comparer les tests d'un laboratoire à une source extérieure au laboratoire. C'est un système permettant de vérifier objectivement les performances des tests du laboratoire, fournissent des éléments de qualité et une alerte pour les erreurs de diagnostic en faisant appel à un organisme ou à une installation externe (28).

## 2.2. Nature des contrôles

Les contrôles utilisés en tant que matériels de références peuvent être de deux natures différentes : biologiques ou synthétiques. Ils présentent tous deux certains avantages et inconvénients, et doivent être choisis suivant l'analyse à réaliser.

### 2.2.1. Matériel biologique

Du matériel biologique peut être utilisé en tant que contrôle. Cela peut être un prélèvement de patient (sang, sérum, prélèvement nasopharyngé, etc.), des souches bactériennes ou virales produites en culture.

#### Avantages :

- Ce sont les contrôles les plus proches de l'échantillon à tester. En effet, pour des tests virologiques notamment, ils peuvent représenter la souche virale qui circule réellement dans la population.
- Ils contiennent le génome complet d'un virus.

#### Inconvénients :

Leur quantité peut être limitée ; il faut avoir à disposition suffisamment de patients positifs à un pathogène donné, ou pouvoir cultiver l'agent pathogène, ce qui n'est pas toujours le cas.

- Ils sont limités en variabilité : si un nouveau variant émerge, il faut trouver un patient infecté par ce variant pour en isoler la souche.
- Ces contrôles peuvent potentiellement être infectieux, et doivent être produits dans des environnements de haute sécurité, comme des laboratoires de biosécurité de niveau 3 ou 4 (BLS3, BSL4), pour la production de contrôles respectivement pour le virus de la fièvre jaune ou Ebola par exemple.

Au laboratoire, des bactériophages sont utilisés en tant que contrôles internes de nature biologique. Le bactériophage MS2 est utilisé comme contrôle interne à génome ARN, et le bactériophage T4 comme contrôle interne à génome ADN. Ces virus sont introduits dans l'échantillon à analyser avant extraction virale, constituant ainsi des contrôles positifs d'extraction.

### 2.2.2. Matériel synthétique

Il est également possible d'utiliser des contrôles produits artificiellement en laboratoire pour constituer un matériel de référence.

#### Avantages :

- Les quantités sont quasiment illimitées au regard des besoins.
- Ils peuvent être produits à façon selon les besoins diagnostiques. Différents témoins peuvent être construits en fonction de la région génomique à analyser, et peuvent être modifiés si de nouveaux variant viraux apparaissent.
- Ils permettent d'être rapidement opérationnel. En effet, il faut compter environs un à sept jours de production selon le témoin.

#### Inconvénient :

- Ils ne contiennent qu'une partie du génome et non pas le génome complet, contrairement au matériel biologique.
- Les contrôles synthétiques sont en général des acides nucléiques nus, ce qui ne reflète pas nécessairement le matériel génétique viral encapsidé. Il existe des technologies permettant d'encapsider du matériel génétique, mais la capsidie reste néanmoins différente de la capsidie virale.

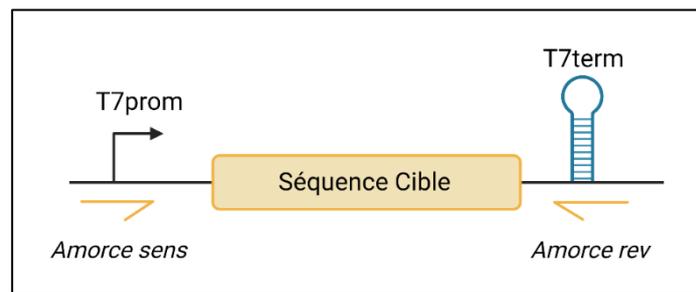
Il est possible de classer les contrôles synthétiques en trois groupes : i) les produits PCR, ii) les transcrits produits *in vitro* (*In Vitro* Transcrits, IVT) ainsi que iii) du matériel génétique encapsidé (Armored RNA (ArRNA) ou Armored DNA (ArDNA)).

#### 2.2.2.1. Matériels synthétiques nus.

Le produit PCR est le contrôle positif le plus simple à développer et produire. Il s'agit de l'amplicon obtenu à la fin d'une réaction par PCR. C'est une séquence en acides nucléiques d'une taille définie par la distance entre les deux amorces utilisées pour réaliser l'amplification. Le produit PCR ne permet cependant de ne contrôler que la dernière étape du processus diagnostique, c'est-à-dire l'étape de PCR.

Afin de contrôler l'étape d'amplification mais également celle de transcription inverse, il est possible d'utiliser des ARN synthétiques, spécifiques du pathogène à détecter. La technique la plus simple pour produire des ARN synthétiques est celle de la transcription *in vitro*. Cette technique consiste en la réalisation d'une PCR sur un plasmide contenant la séquence d'intérêt,

encadrée par un promoteur et d'un terminateur de transcription (**Figure 9**). Le produit PCR ainsi généré sera ensuite transcrit à l'aide d'une ARN polymérase, la matrice ADN est ensuite dégradée par une DNase, générant ainsi une matrice ARN utilisable en tant que contrôle positif. La transcription est réalisée grâce à des séquences promotrices et terminatrices : le promoteur est une séquence nucléotidique placée en amont de la séquence à transcrire, à laquelle la polymérase va se fixer afin d'initier la transcription ; le terminateur est une séquence riche en G-C située en aval de la séquence transcrite, ralentissant le fonctionnement de la polymérase et formant une « tige-boucle » bloquant l'enzyme. En général, l'ARN polymérase du phage T7 est utilisée car elle permet d'obtenir de bons rendements sans limite majeure de type de séquence. Les produits PCR ainsi que les IVT peuvent tous deux servir de contrôles quantitatifs : un dosage par spectrophotométrie après leur synthèse permet de connaître leur nombre de copies d'acide nucléique, et peuvent donc par la suite être utilisés comme gamme d'étalon pour déterminer le nombre de copies d'un échantillon.



**Figure 9 : Matrice permettant la synthèse d'un Transcrit *In Vitro*.**

*Représentation schématique et simplifiée d'une matrice permettant la synthèse d'un IVT. T7prom : Promoteur T7 T7term ; Promoteur T7. Figure générée avec BioRender.*

#### 2.2.2.2. Matériels synthétiques encapsidés

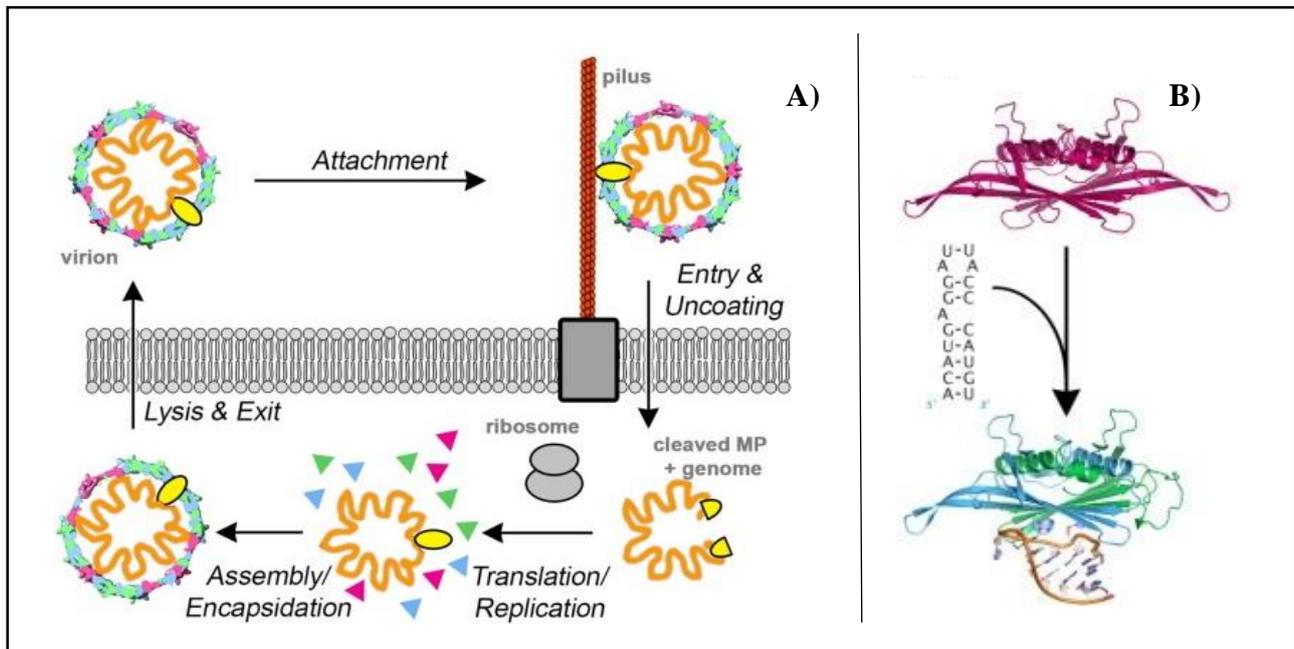
Les contrôles d'extraction doivent correspondre à du matériel génétique encapsidé infectieux ou non, mimant une particule virale. Afin de disposer de contrôles permettant de tester et de contrôler toutes les étapes d'un processus diagnostique, c'est-à-dire de l'extraction virale à la réalisation de la PCR, la technologie « Armored » permet d'encapsider des séquences d'ADN ou d'ARN viraux non infectieux produits en laboratoire. Ces contrôles reproduisent la structure des phages MS2 ou T4 mentionnés plus haut, mais peuvent contenir n'importe quelle séquence cible afin de détecter n'importe quel pathogène.

## Les ARN encapsidés :

Un Armored RNA (ArRNA, ARN encapsidé par plusieurs protéines hétérologues) est un complexe de protéine de capsid de bactériophage MS2 et d'ARN produit dans *Escherichia coli* par l'induction de l'expression d'un gène codant pour des protéines de capsid et une séquence d'ARN (29). La technologie des ArRNA a été développée dans les années 90 par la société Ambion. Cette technique consiste à enfermer l'ARN cible dans une capsid protéique. Cette technologie a pour but de constituer des contrôles positifs d'extraction, les rendant ainsi capables de contrôler l'entièreté du processus diagnostique par PCR (**Figure 7**). De plus, la capsid protéique rend ce contrôle plus stable qu'un ARN nu, permettant ainsi de s'affranchir des problèmes de transport et de stockage dans le temps que l'on rencontre notamment avec les IVT. En effet, les ArRNA ainsi produits, protégés des RNases et de l'environnement extérieur, sont plus stables dans le temps qu'un ARN nu (30), et peuvent être conservés à 4°C, 25°C ou 37°C pendant plus de 30 jours sans perte significative de leur nombre de copies (31).

Le phage MS2 est composé d'un génome ARN simple brin de polarité positive. Il code pour quatre protéines : une protéine de maturation (Maturation Protein, MP), une protéine de lyse, une protéine de capsid (Coat Protein, CP) et une réplicase (32). La CP est l'élément principal de son enveloppe protéique : la capsid icosaédrique est constituée de 90 dimères de CP et d'une unique copie de la MP exposée à la surface (33). Les virus se lient d'abord aux côtés du pilus bactérien par l'intermédiaire de la MP (**Figure 10 A**). Contrairement à d'autres phages, seul le complexe ARN-MP pénètre dans la cellule. Les MP se lient à des structures dites d'emballage, appelées séquences PAC situées à chaque extrémité du génome MS2 (34). Ces séquences PAC jouent un rôle crucial dans la réplication et l'assemblage du phage. Elles forment des éléments de structure secondaire définis, qui sont spécifiquement reconnus par les CP virales (**Figure 10 B**). Ces séquences sont des opérateurs de traduction pour la réplicase, et des sites d'initiation de l'assemblage des CP. En d'autres termes, elles agissent comme des signaux pour guider le processus d'emballage de l'ARN génomique dans les nouvelles particules virales. Lorsque le génome ARN est dans la cellule, il est utilisé comme matrice pour la traduction des protéines virales. Le gène de la CP, qui est la protéine la plus abondante dans cette structure, peut être traduit directement. Une fois internalisée, la MP est clivée en deux fragments distincts par une protéase, ce qui permet à la traduction et à la réplication de commencer. Le contrôle temporel de l'expression des gènes du phage régule ensuite la

production de génomes descendants et de protéines structurales qui s'assemblent avant l'action de la protéine de lyse du phage.



**Figure 10 : Composants du virion MS2 et son cycle de vie**

A) Cycle de vie du bactériophage MS2. Les éléments en jaune correspondent aux séquences PAC ; les triangles correspondent aux protéines de capsid. B) Représentation de la fixation des séquences PAC aux protéines de capsid. Figure extraite de l'article (35)

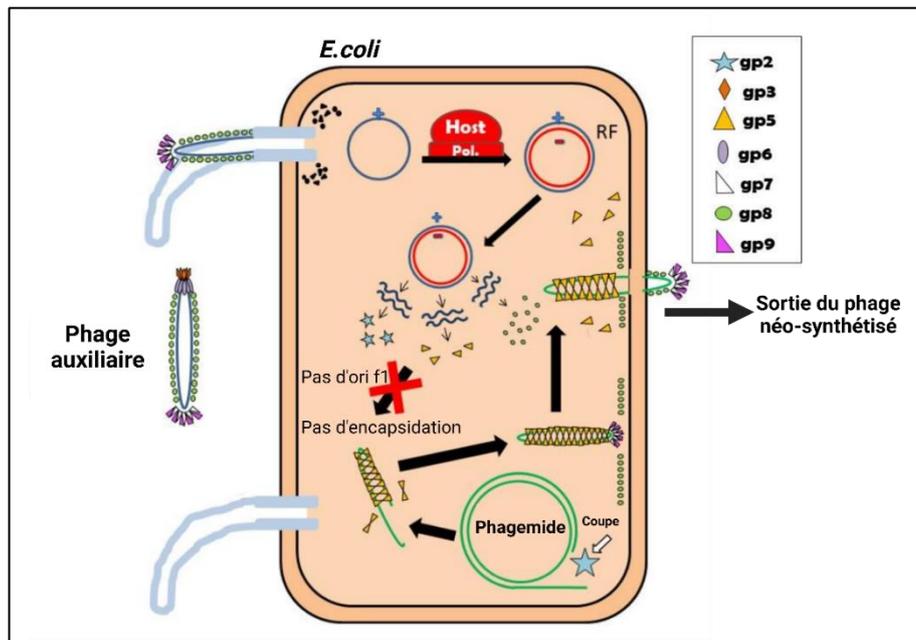
Fusionnée à une séquence PAC, il a été démontré qu'une séquence d'ARN recombinant peut être spécifiquement empaquetée dans une capsid recombinante de MS2 (36). Pour la production de VLPs (Virus like Particles, particules semblables à celles d'un virus), trois éléments sont indispensables : la capsid et la maturase pour permettre la formation de la pseudo-particule, et la séquence PAC pour la fusion de l'ARN avec la capsid virale. Une unique copie de la maturase assemblée avec 90 dimères de protéines de capsid permet la formation de la capsid d'une particule virale. La plus longue séquence à avoir été encapsidée en utilisant un unique site PAC est de 1200 bases (37), mais l'ajout d'un second site PAC permet une liaison coopérative du dimère de protéines de capsid à l'ARN, résultant en une plus haute affinité vis-à-vis d'un unique site PAC (38). De fait, en intégrant plusieurs sites PAC, l'ARN le plus long encapsidé à ce jour est de près de 5000 nucléotides (39).

## Les ADN encapsidés :

Les types de contrôles ou d'étalons couramment utilisés pour les essais sur les virus à génome ADN peuvent être de l'ADN plasmidique nu, des échantillons de patients positifs au virus étudié, des lignées cellulaires spécifiques et de l'ADN encapsidé (Armored DNA, ADN encapsidés par une ou des protéines hétérologues). Les contrôles dérivés de plasmides sont faciles à produire, non infectieux pour l'homme, mais sont susceptibles d'être dégradés par des DNases du fait que l'ADN ne soit pas protégé. De plus, l'ADN plasmidique ne peut servir de contrôle pour l'ensemble du processus de PCR en temps réel, car l'extraction des acides nucléiques n'est pas effectuée sur ce type de contrôle. Les échantillons de patients peuvent quant à eux constituer un contrôle pour l'ensemble du processus, mais peuvent être infectieux et rendent la manipulation et les transports complexes. Afin de générer des contrôles synthétiques capables de contrôler l'entièreté du processus de diagnostic par PCR, plusieurs techniques de production d'Armored DNA ont été publiées. Les phages à génome ADN sont principalement utilisés pour encapsider des ADN, mais des capsides de baculovirus ont également été utilisées pour construire un ArDNA (40). Quatre types de phages ont déjà été utilisés pour construire un ArDNA : le bactériophage T7, les phages filamenteux, le bactériophage MS2 et le phage lambda (41–44). Le phage T7 n'est en général pas choisi pour synthétiser des ArDNA, du fait de sa faible efficacité d'emballage (41). Le phage lambda permet quant à lui de synthétiser simplement des ArDNA doubles brins, avec une meilleure efficacité d'empaquetage que le bactériophage T7 (43). Pour cela, le vecteur d'expression Lambda GT11 est utilisé : il peut tolérer des insertions d'une longueur maximale de 7200 bases dans un site de restriction *EcoRI* unique situé près de l'extrémité C-terminale de la région codante du gène *lacZ*.

Étonnamment, la production d'ArDNA dérivée de la technologie ArRNA est possible avec la capside du phage MS2. Le processus nécessite cependant de dissocier la capside recombinante du phage MS2 en dimère de CP *in vitro*, puis de les rassembler en VLPs en présence d'un site ADN-ARN chimérique avec une séquence PAC modifiée (mélange covalent de PAC ADN et PAC ARN, optimisé par les auteurs de l'article) (44). Cette technique permet d'encapsider de longues séquences, allant jusqu'à 6500 paires de bases (44), mais nécessite de passer par une étape de dénaturation/renaturation des capsides recombinantes laborieuse. Les ArDNA obtenus *via* un phage filamenteux sont simples brins, et ne sont donc pas idéaux pour mimer les propriétés des virus à ADN double brin. Néanmoins le phage filamenteux M13 possède un certain nombre d'avantages. Couramment utilisé en biotechnologie, notamment pour la technique du phage display (45), il permet l'encapsulation de génome phagique recombinant

ou l'encapsidation d'un phagémide, qui est une molécule d'ADN répliquative possédant une origine d'encapsidation. Afin d'utiliser la machinerie du phage M13 pour générer un ADN simple brin encapsidé, il est nécessaire d'utiliser un phage auxiliaire du phage M13. Ce phage auxiliaire possède les protéines d'enveloppe complètes et un génome complet du phage M13, cependant il possède une origine de répllication différente de celle du phage M13 et un gène de résistance à un antibiotique, tous deux insérés à la place de l'origine de répllication de M13 (ori f1). En présence d'un phagémide portant une origine M13 ou f1, le phagémide simple brin est emballé de manière préférentielle (par rapport au génome M13), ce qui permet de produire de l'ADN de phagémide simple brin pouvant être utilisé pour de la mutagenèse, le séquençage, ou pour produire des ArDNA. Lorsque le phage auxiliaire M13 infecte *E.coli*, son génome simple brin utilisera la polymérase de l'hôte pour générer une forme répliquative (RF) double brin. Le génome de M13 contient 9 gènes principaux, le produit de traduction de chaque gène étant appelé gp1, gp2, et ainsi de suite. Comme l'ori f1 du phage auxiliaire est mutée, les gp2 et gp5 ne reconnaissent pas l'origine d'encapsidation de manière efficace et encapsident donc préférentiellement le phagémide. Ensuite, la gp5 forme une structure en dimère et commence à emballer ce brin simple d'ADN à partir du signal d'encapsidation sur l'ori f1. Le phage néo-synthétisé sera alors libéré de la cellule hôte et sera recouvert par les gp3, gp6, gp7, gp8 et gp9. Les phages ainsi libérés contiendront majoritairement le phagémide contenant la séquence d'ADN exogène d'intérêt.



**Figure 11 : Synthèse d'un ArDNA à partir d'un phagémide et d'un phage auxiliaire**  
 Représentation schématique d'une infection d'*E. coli* par un phage auxiliaire en présence d'un phagémide. RF : Replicative form : Forme répliquative. Pol : polymérase. Schéma modifié à partir de celui présent sur le site : [https://2014.igem.org/Talk:Team:TCU\\_Taiwan/M13Phage](https://2014.igem.org/Talk:Team:TCU_Taiwan/M13Phage) (46).

### 2.3. Enjeux du projet

La détection de virus en diagnostic moléculaire nécessite l'utilisation de contrôles adaptés. Différents témoins doivent être produits à façon en fonction du type de virus à détecter. Pour la détection de virus à génome ARN, des témoins synthétiques sous forme de transcrits *in Vitro* ou d'ARN encapsidé sont utilisés. Pour les virus à génome ADN, les plasmides, les produits PCR, ou les ADN encapsidés peuvent être utilisés. Ce projet a pour objectif de développer et valider pour l'activité diagnostic du laboratoire un panel de contrôles positifs synthétiques qui répondent à l'ensemble des besoins.

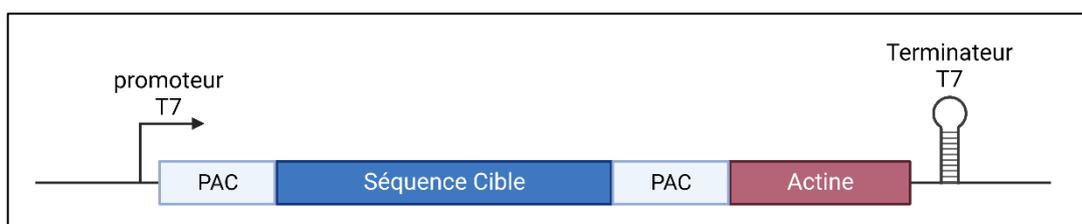
Pour la détection de virus à génome ARN, le premier besoin est de produire des témoins positifs quantitatifs sous forme d'IVT, devant par conséquent être homogènes. Ensuite, ce processus de production doit être robuste et reproductible, afin de générer un protocole générique pour la production de ce type de contrôle. Le deuxième implique le développement d'une plateforme de production d'ARN encapsidés synthétiques, pouvant être produits en grande quantité, homogènes, lyophilisables, distribuables et pouvant servir de contrôle sous forme d'EQA ou mimant du matériel infectieux. Dans ce contexte, la technologie ArRNA devra être implémentée au laboratoire pour répondre aux besoins spécifiques de la plateforme de diagnostic.

Pour la détection de virus à génome ADN, les processus de production des témoins sous forme plasmidique ou de produit PCR sont déjà correctement établis au laboratoire. Cependant, nous ne disposons à l'heure actuelle d'aucun protocole permettant de produire des contrôles positifs sous forme d'ADN encapsidé. Le dernier enjeu de cette partie de mon projet consiste donc au développement d'un protocole de synthèse d'ArDNA, avec les mêmes objectifs que pour les ArRNA.

## 2.4. Matériels et méthodes

### 2.4.1. Conception des gènes synthétiques

Les matrices ADN sont synthétisées et clonées par la société GENEWIZ ou Thermo Fisher Scientific dans des plasmides pUC-GW-Amp. Elles sont conçues selon le modèle détaillé dans la **Figure 12** ci-dessous. Ces plasmides sont reçus sous forme lyophilisée et sont reconstitués avec de l'eau ultra-pure sans nucléases pour obtenir une solution de plasmide à 0,1 µg/µl. La séquence codant l'actine est une séquence hexogène permettant de normaliser la quantité de matériel ARN. T7 prom : TAATACGACTCACTATAGGGAG. PAC : ACATGAGGATCACCCATGT. Actine : GTCTGGATTGGTGGCTCCATCTTGGCTTCCCTGTCCACCTTCCAGCAGATGTGGATCAGCAAGCAGGAGTATGATGAGTC. T7 Term : GCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTG



**Figure 12 : Présentation du design d'une séquence codant pour un transcrit *in vitro***

*Représentation schématique du design d'une matrice ADN compatible à la production d'un IVT ou d'un ArRNA. PAC : séquence encapsidatrice Actine : séquence codant le gène d'actine. Figure générée avec BioRender.*

### 2.4.2. Production, quantification et validation des transcrits *in vitro*.

La PCR avant la transcription *in vitro* est réalisée avec le kit de PCR « Patinum™ PCR SuperMix High Fidelity (Invitrogen, ThermoFisher) ». Le mix PCR contient 15µM de chaque amorce, 0,2µg de plasmide et 145µl de Mix PCR. Le cycle thermique a été défini comme suit : 2 min à 94°C, puis 30 cycles de : 30 sec à 94°C, 30 sec à 55°C, 1 min à 68°C. Le produit PCR obtenu est déposé sur un gel à 1% d'agarose, et migre pendant 30min à 135V, afin de vérifier la taille de l'amplicon obtenu. Le produit PCR est ensuite purifié à l'aide du kit de purification « Monarch PCR & DNA cleanup Kit, Biolabs New England », selon le protocole du fournisseur. Le produit final est dosé par spectrophotométrie à 260nm.

La transcription *in vitro* est réalisée à partir du produit PCR obtenu à l'étape précédente, avec le « Kit de transcription T7 Megashort script Kit, Ambion », selon le protocole du kit. Les ARN transcrits obtenus sont purifiés avec le « kit de purification ARN MEGAclear™ Transcription

clean-up ». L'éluion finale se fait avec 50µl d'eau ultra pure. L'ARN obtenu est dosé par spectrophotométrie à 260nm.

Les mesures de poids moléculaire des ARN transcrits se font par électrophorèse capillaire, avec l'appareil « Bioanalyser Agilent 2100 » et le kit « Agilent RNA 6000 Nano Kit », selon le protocole du kit. Les ARN à analyser sont déposés purs.

Les RT-qPCR conçues pour valider les IVT sont réalisées avec le « Kits SuperScript™ III Platinum™ One-Step qRT-PCR » et avec un profil thermique défini comme suit : 15 min 50°C, 2 min à 95°C, puis 45 cycles de 15 sec à 95°C puis 45 sec à 60°C. Le mix PCR contient 10 µM de chaque amorce, 1µM de sonde, 5µl d'échantillon et 12,5µl de Mix 2X. Les RT-qPCR sont réalisées dans un thermocycleur « CFX96 Touch Real-Time PCR Detection System, Bio-Rad ».

### 2.4.3. Production des ARN encapsidés

#### 2.4.3.1. Synthèse des ArRNA

Des bactéries *E.coli* T7 Expression Iq ont préalablement été transformées avec un plasmide pET28b-CP-HIS-CP-Mat –codant pour les protéines de CP et MP- et cultivées dans un milieu contenant kanamycine et chloramphénicol, puis rendues compétentes. Les bactéries sont à nouveau transformées avec 1µl de plasmide pUC-GW-Amp contenant le gène d'intérêt ainsi que la séquence PAC, et sont étalées sur un mélange d'agarose et de milieu de culture Luria Bertani (LB), et de trois antibiotiques : Ampicilline (0,1 mg/ml), Kanamycine (0,05 mg/ml) et Chloramphénicol (0,034 mg/ml), puis incubées toute une nuit à 37°C.

Les précultures sont faites dans 5 ml de milieu LB + Ampicilline (0,1 mg/ml), Kanamycine (0,05 mg/ml) et Chloramphénicol (0,034 mg/ml), et sont incubées toute une nuit à 37°C à 200 RPM.

Les cultures sont faites dans 200 ml de milieu Terrific Broth (TB) (47,6 mg/ml), 1,6 ml de glycérol + Ampicilline (0,1 mg/ml), Kanamycine (0,05 mg/ml) et Chloramphénicol (0,034 mg/ml). Les cultures sont incubées à 37°C à 200 Rotation Par Minute (RPM) et la densité optique (DO<sub>600nm</sub>) est mesurée. L'induction de la protéine CP se fait avec 0,5mmol/l d'IPTG, puis sont incubées toute une nuit à 25°C à 200 RPM. Les cultures sont arrêtées par centrifugation pendant 10 minutes à 4000g puis congelées à -80°C. Après décongélation, les bactéries sont lysées avec 10ml de tampon de lyse (50 mmol/l TRIS base, 300mmol/l NaCl, 10 mmol/l imidazole, 5% glycérol, 0,1% Triton, 10µg/ml Dnase, 0,25 mg/ml lysosyme, 1 mmol/l PMSF, 1x BugBuster pH8) et incubées pendant 30 minutes sous agitation. Le milieu est ensuite centrifugé pendant 30 minutes à 12000 RPM. Le surnageant est collecté, et incubé pendant 30 minutes avec ajout d'1ml de billes de Nickel. La purification se fait par chromatographie

d'affinité sur ions métalliques immobilisés (IMAC). Deux lavages sont faits, le premier avec 4ml de tampon « A » (50mmol/l TRIS base, 300 mmol/l NaCl, 10 mmol/l imidazole), le deuxième avec 3,6 ml de tampon « A » et 0,4 ml de tampon « B » (50 mmol/l TRIS base, 300 mmol/l NaCl, 500 mmol/l imidazole). L'élution se fait avec 5ml de tampon « B ». Le produit obtenu est concentré dans un concentrateur de protéine « Merck Millipore Ensembles de filtration par centrifugation Amicon™ Ultra 100K Da ». Les protéines obtenues sont dosées par spectrophotométrie à 280nm.

#### 2.4.3.2. Lyophilisation et validation de l'ArRNA

400µl d'ArRNA ainsi produit sont extrait avec l'appareil « EZ1 Advanced XL, Quiagen », avec le kit « Virus mini kit v2.0 » dans un volume de 60µl d'ARN carrier. Ils sont repris dans un volume d'eau de QSP 500µl. Ils sont par la suite testés par RT-qPCR avec le kit le « Kits SuperScript™ III Platinum™ One-Step qRT-PCR », et un profil thermique défini comme suit : 15 min 50°C, 2 min à 95°C, puis 45 cycles de : 15 sec à 95°C puis 45 sec à 60°C avec des amorces spécifiques à l'ArRNA produit. Le mix PCR contient 10 µM de chaque amorce, 1µM de sonde, 5µl d'échantillon et 12,5µl de Mix 2X. Ils sont testés en dilutions au  $1/5e$ , jusqu'à définir la solution ayant une valeur seuil de 30 CT.

Une fois la dilution de travail déterminée, l'ArRNA est lyophilisé, après ajout de 0,2M d'une solution de sucrose (13.7gr de sucrose dilué dans 40ml de HBSS (Gibco™ HBSS 1x)), selon le protocole de lyophilisation décrit en [Annexe 4](#)).

Une dernière RT-qPCR post lyophilisation est réalisée pour valider l'ArRNA produit. Il est repris dans 500µl d'eau, puis extrait et analysé en RT-qPCR de la même manière qu'énoncé précédemment (**Figure 13**).

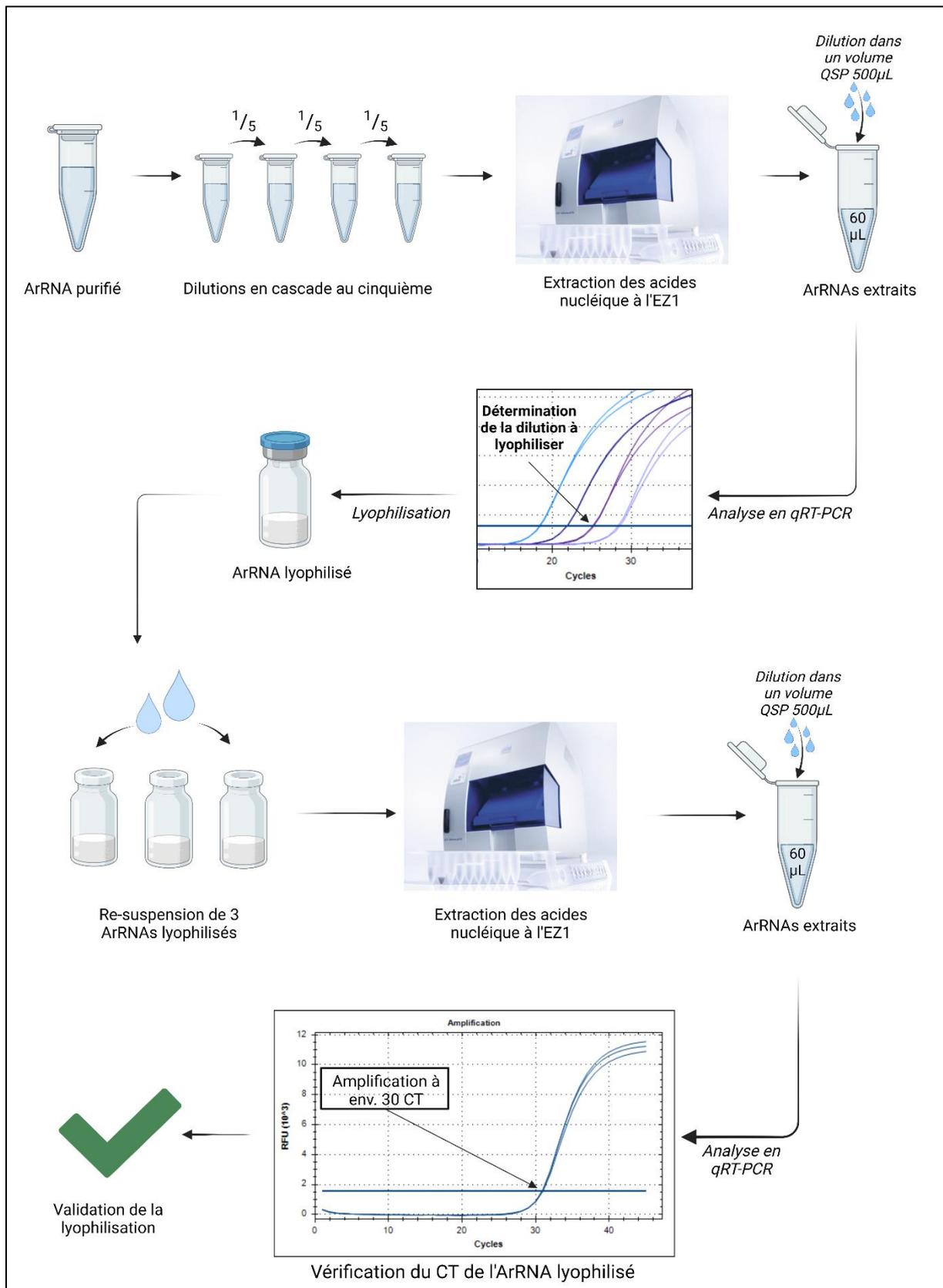
#### 2.4.3.3. Contrôle de l'homogénéité

Le contrôle de l'homogénéité des ArRNA a été réalisé par une analyse PCR. Un système PCR sur le plasmide a été désigné comme suit : Amorce Sens : CATGGTCATAGCTGTTTCCTG ; Amorce reverse : GCTTTACACTTTATGCTTCCGGC ; sonde : TTGTTATCCGCTCACAATTCCACAC

Les RT-qPCR ont été réalisées avec le « Kits SuperScript™ III Platinum™ One-Step qRT-PCR », avec un profil thermique défini comme suit : 15 min 50°C, 2 min à 95°C, puis 45 cycles de : 15 sec à 95°C puis 45 sec à 60°C. Le mix PCR contient 10 µM de chaque amorce, 1µM de sonde, 5µl d'échantillon et 12,5µl de Mix 2X. Les RT-qPCR sont réalisées dans un thermocycleur « CFX96 Touch Real-Time PCR Detection System, Bio-Rad ».

#### 2.4.3.4. Production et validation de l'ArRNA MS2

L'ArRNA MS2 a été produit dans les conditions décrites ci-dessus. Une fois produit, 200 µL d'ArRNA dilué au dixième ont été extraits. Une gamme de sept dilutions en série au 1/5e a été analysée par RT-qPCR afin de déterminer la dilution de travail. Afin de valider cet ArRNA, il a été ajouté (spiké) dans différents échantillons : prélèvement nasopharyngé, salive, urine, clarifiât de selles, sérum et sang total. Pour valider les résultats, un phage MS2 (produit par d'autres membres du laboratoire) a été utilisé en parallèle et ce dans les mêmes conditions ; c'est à dire 10 µL de MS2 (phage ou ArRNA) sont ajoutés à 200µL d'échantillon, extrait à l'EZ1 puis analysé par RT-qPCR.



**Figure 13 : Schéma du processus de validation d'un ArRNA**

*Schéma généré avec BioRender.*

#### 2.4.4. ADN Encapsidé

##### 2.4.4.1. Capsides recombinantes

Les séquences des inserts intégrés dans les plasmides permettant la synthèse des trois protéines d'intérêts sont disponibles en [Annexe 5](#). Des séquences HIS-Tag ont été intégrées dans chaque plasmide, afin de permettre une purification par IMAC. Un site de clivage à la TEV protéase est inséré à la suite du HIS-Tag afin de le supprimer après la purification. Pour la fIMCP (Full length Major Capsid protein, protéine majeure de capside, taille complète), une séquence codant pour la Thioredoxine est ajoutée afin d'augmenter sa solubilisation. Les transformations bactériennes, les pré cultures ainsi que les cultures sont réalisées de la même manière que pour la production des ArRNA, sans utilisation d'ampicilline. L'IMAC se fait de la même manière, à l'exception de la lyse bactérienne. Elle se fait ici sans ajout de BugBuster et par deux sonications de 45 sec.

Les SDS PAGE post-production sont réalisés avec le kit « mPAGE™ 4-20% bis-tris gels, Merck » et la migration se fait sous une tension de 180 volts pendant 40 minutes. La révélation se fait avec le colorant « InstantBlue® Coomassie Protein Stain, Abcam ». Les chromatographies d'exclusions de taille (SEC) se font sur colonne « Cytiva HiLoad™ 16/600 Superdex™ 200 pg, Fisher Scientific ».

Les tests de dégradation à la DNase se font avec 1µl de « DNase-I RNase Free, New England Biolabs » avec 1µl de tampon « CutSmart® Buffer, NEB » et sont par la suite analysés sur gel d'agarose 1%, sous tension 100 volts pendant 30 min.

Les tests de TSA (Thermal shift assay) permettant d'évaluer l'assemblage/désassemblage des protéines se font suivant le protocole thermique suivant : 20°C pendant 1 min, augmentation de la température jusqu'à 45°C, incrément 0.2°C par minute, 45°C pendant 1 minute, diminution de la température jusqu'à 20°C, incrément 0.2C par minute. La mesure de la fluorescence se fait avec le colorant SYPRO orange (Thermo Fischer Scientific).

##### 2.4.4.2. Phagémide encapsidé

Les ArDNA sont synthétisés dans des bactéries TG1 (LU-60502-2 TG1 Electrocompetent Cells, EUROMEDEX). Ces bactéries sont rendues chimio-compétentes, voir protocole détaillé en [Annexe 6](#). L'insert permettant la synthèse de l'ArDNA, basé sur une séquence spécifique de l'Hépatite B, est clonée dans un phagémide pHEN4 (séquences détaillées en [Annexe 7](#)) par la

société Thermo Fischer Scientific. Le produit cloné est reçu sous forme lyophilisé, et resuspendu dans 50µl d'eau ultra pure pour obtenir une solution à une concentration de 1µg/µl.

Les transformations et les précultures sont faites de la même manière que toutes les autres décrites précédemment, avec ajout de 0.1mg/ml d'ampicilline. Les cultures sont faites dans un milieu 2X TY (16 g Bacto tryptone, 5 g NaCl et 10 g d'extrait de levure dans 1L d'H<sub>2</sub>O). La suite des expérimentations sont réalisées selon le protocole décrit dans un article de Nature Protocols, « A general protocol for the generation of Nanobodies for structural biology » (47), étapes 37 à 45, avec le phage auxiliaire « M13K07 helper phage New England BioLab ».

Le produit obtenu est analysé par qPCR après extraction avec l'appareil « EZ1 Advanced XL, QIAGEN » et le kit « Kits SuperScript™ III Platinum™ One-Step qRT-PCR », sans l'étape de rétro transcription. Les tests de dégradation à la DNase sont réalisés sur 100µl d'ArDNA avec le kit « Kit TURBO DNA-free™ AM1907, Thermo Fisher Scientific™. », selon le protocole « rigorous treatment » du kit.

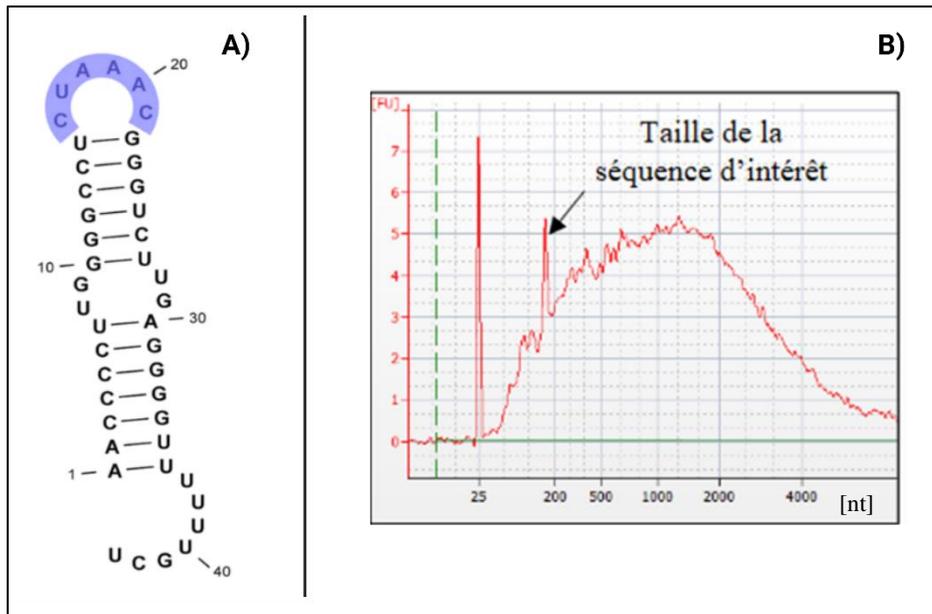
## 2.5. Résultats et discussion

### 2.5.1. Transcrits *in vitro*

Avant mon arrivée au laboratoire, les Transcrits *in vitro* (ITV) étaient initialement produits à partir d'un produit PCR. Ces PCR étaient réalisées avec des amorces ciblant les séquences universelles M13 sur le plasmide, de part et d'autre du site d'insertion. Cette méthode présente différents inconvénients, notamment des problèmes de reproductibilité et d'efficacité. Par ailleurs, la réalisation d'une PCR peut engendrer des mutations au cours des cycles d'amplification. Afin de supprimer l'étape de PCR, nous avons cherché à réaliser la transcription directement à partir du plasmide en travaillant notamment sur l'optimisation d'une séquence terminatrice de transcription pour éviter les fuites de terminaison, responsables de l'hétérogénéité de taille des IVT. Plusieurs séquences terminatrices de transcription sont donc intégrées en aval de la séquence d'intérêt. Un terminateur de transcription est une séquence nucléotidique marquant la fin de la transcription. Classiquement, la séquence « terminateur T7 » (T7term), issu du bactériophage T7 est choisie car l'ARN polymérase dérivée de phages T7 est le système de transcription le plus utilisé dans le domaine de la biologie synthétique (48). Le T7term code pour une séquence d'ARN qui peut former une structure stable de tige-boucle suivie d'une répétition de six résidus uridylés. La transcription se termine au niveau du résidu G en 3' en aval du poly-U (**Figure 14 A**). Les témoins positifs ont donc été synthétisés en réalisant une transcription directement sur le plasmide, grâce au kit « MEGAshortscript™ T7 Transcription (Ambion) »

Afin de déterminer l'homogénéité de taille des témoins positifs ainsi produits, nous les avons analysés par électrophorèse capillaire (**Figure 14 B**). L'électrophérogramme met en évidence un pic étroit à la taille attendu du transcrit (160 nt, repéré par une flèche sur la **Figure 14 B**) mais aussi la présence d'un large pic (de quelques dizaines à plus de 4000 nucléotides), qui suggère une fuite de transcription du terminateur T7. En effet, si le terminateur T7 est efficace à 100%, nous ne devrions observer qu'un seul pic à environ 160 nt, taille théorique de l'ARN de référence.

Pour optimiser la terminaison de transcription, deux options sont possibles : utiliser une matrice courte qui s'arrêterait après le T7term en réalisant une PCR sur le plasmide, ou augmenter la force du signal de terminaison.

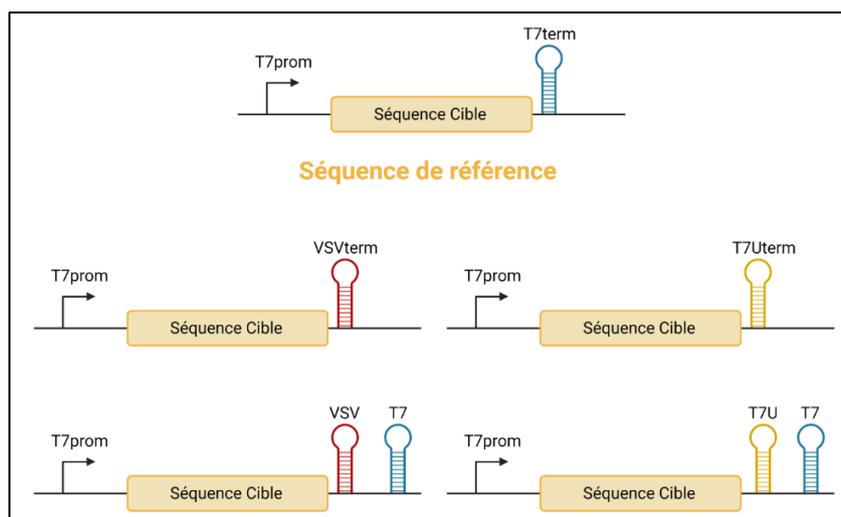


**Figure 14 : Structure secondaire du T7term et mise en évidence des problèmes de terminaison de transcription**

*A) Structure secondaire du terminateur T7. Figure extraite de l'article (48). B) Électrophorégramme obtenu après l'analyse d'un IVT produit grâce à une transcription directement le sur plasmide. Le pic à 25nt correspond au standard de poids moléculaire. Abscisse : taille en nucléotides (nt). Ordonnée : Fluorescence Units (FU).*

#### 2.5.1.1. Optimisation du signal de terminaison

Pour augmenter le signal de terminaison de transcription, plusieurs séquences de terminaison issues de l'article de Mairhofer et al. ont été testées (48). Nous avons testé trois séquences de terminaison différentes et deux combinaisons de séquences de terminaisons : la séquence T7term, qui est notre séquence de référence ; une séquence VSVterm, séquence de terminaison du Virus de la Stomatite Vésiculaire (VSV), qui a démontré sa fonctionnalité pour la polymérase T7 (49). Selon les auteurs, le VSVterm est capable d'interrompre l'activité de la polymérase avec une efficacité de 75 %. Nous avons également testé une séquence T7U, séquence T7term optimisée par les auteurs de l'article. Les structures secondaires de ces terminateurs sont détaillées en [Annexe 8](#). Nous avons également testé deux différentes combinaisons de ces séquences de terminaison (**Figure 15**). En se basant sur les données de l'article, l'association en tandem des séquences T7U et T7term permet une efficacité de terminaison de 93%, et l'association VSVterm T7term de 92%.

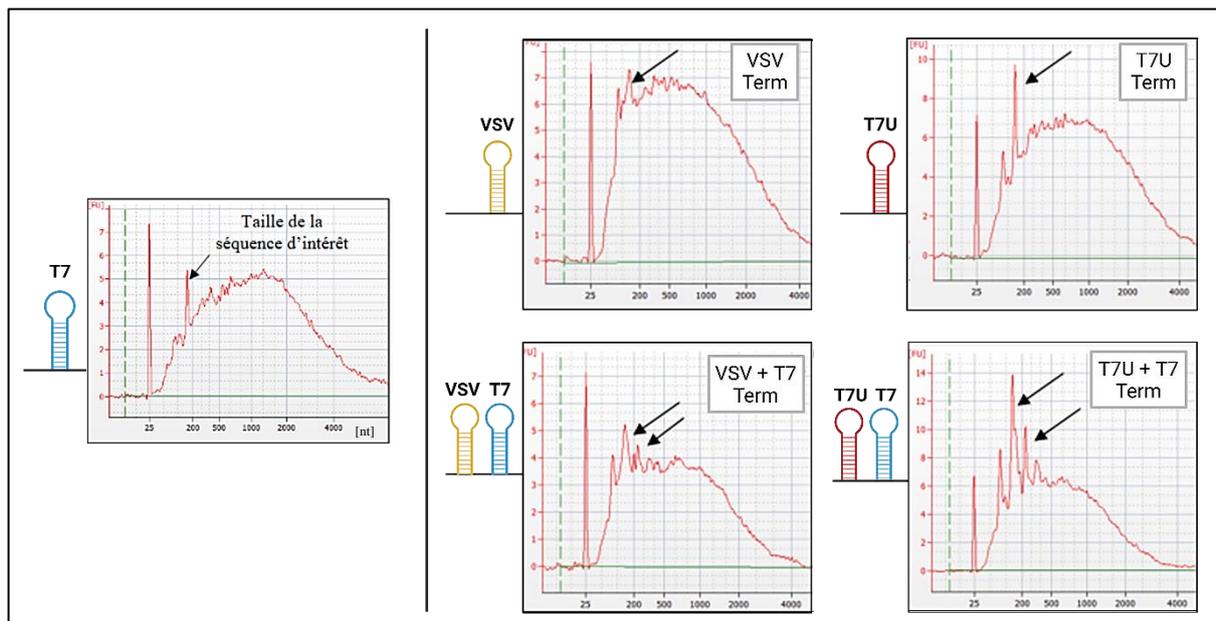


**Figure 15 : Présentation des différentes séquences de terminaison de transcription et combinaisons testées.**

*Représentation schématique des différents sites de terminaison testé. T7prom : Promoteur T7. T7term : terminateur T7. VSVterm : terminateur VSV. T7Uterm : Termineur T7U. Schéma généré avec BioRender.*

Les cinq IVT contenant ces cinq types de séquences de terminaison ont été synthétisés en réalisant une transcription à partir du plasmide correspondant, et ont par la suite été analysés par électrophorèse capillaire (

Figure 16). Pour les IVT contenant les séquences VSVterm et TU7term seules, un large pic avec une fluorescence importante est toujours observé. Pour le VSVterm, le pic correspondant à la taille théorique de l'ARN est à peine visible. De fait, ces séquences terminatrices seules ne montrent pas une efficacité de terminaison de transcription nettement supérieure à celle du T7term. Lorsque deux séquences de terminaison se succèdent, nous observons deux pics relativement étroits correspondant aux tailles des séquences d'intérêt, mais toujours avec présence d'un large pic qui reflète une fuite de terminaison. Néanmoins, le bruit de fond de ce large pic est moins important qu'avec un seul terminateur. Même si l'ajout de plusieurs séquences de terminaison permet une amélioration de l'homogénéité de la taille des transcrits, celle-ci n'est toujours pas optimale ce qui ne permet pas de quantifier de manière précise les IVT.



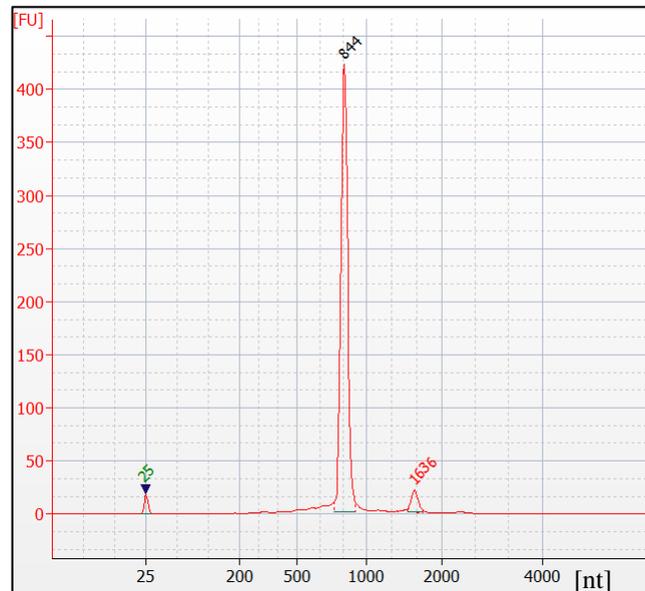
**Figure 16 : Comparaison des différentes séquences de terminaison de transcription**

Résultats d'analyse par électrophorèse obtenus après transcription d'IVT contenant différentes séquences de terminaison de transcription. T7 : terminateur T7. VSV : terminateur VSV. T7U : terminateur T7U. Abscisse : taille en nucléotides. Ordonnée : Fluorescence Units (FU).

### 2.5.1.2. Production des IVT à partir d'un produit PCR

La production d'IVT en réalisant une transcription directement sur le plasmide ne donnant pas de résultats concluants, il a été décidé de revenir au protocole précédent, c'est-à-dire de réaliser une PCR en amont de la transcription. La PCR avec les amorces placées sur les séquences universelles M13 n'étant pas optimale au laboratoire, j'ai construit des amorces sur le promoteur et le terminateur T7 (**Figure 9**), permettant par ailleurs de n'amplifier que la séquence à transcrire. Une fois le produit PCR obtenu, la transcription est réalisée sur celui-ci, permettant ainsi de produire un transcrit homogène en taille. Après analyse de l'IVT ainsi produit (**Figure 17**), nous obtenons un pic majoritaire correspondant à notre séquence d'intérêt, soit environ 900nt, et un pic minoritaire à 1636nt pouvant correspondre à un dimère de produit de PCR et moins probablement à une fuite de terminaison de transcription issue des traces de

plasmides résiduelles après la PCR. Ce résultat permet de conclure que cette technique de production de transcrit *in vitro* est la plus efficace pour obtenir des transcrits homogènes, et nous avons donc choisi de conserver cette méthode pour la production de nos IVT et ce malgré les différents inconvénients que peuvent engendrer la réalisation d'une PCR (augmentation du temps de manipulation, apparition d'éventuelles mutations).



**Figure 17 : Validation de la méthode de production des Transcrits *In Vitro***

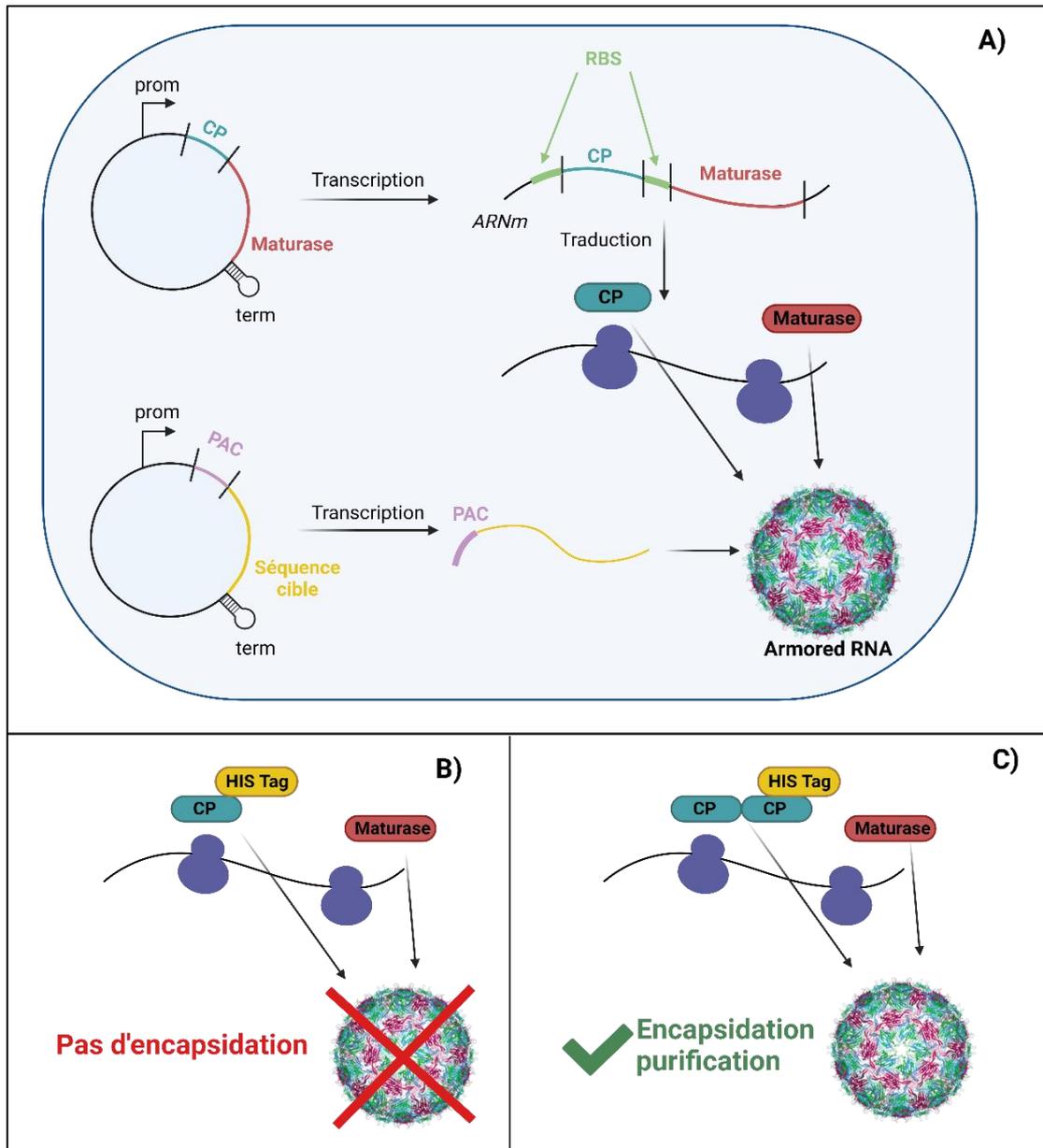
Électrophorégramme d'un IVT synthétisé à partir d'un produit PCR. Premier pic : standard de poids moléculaire. Abscisse : taille en nucléotides. Ordonnée : Fluorescence Units (FU).

## 2.5.2. ARN encapsidés

### 2.5.2.1. Purification des ARN encapsidés

Le développement de la technique de production des ArRNA a été développée avant mon arrivée au laboratoire. Le principe de la production est présenté en (**Figure 18A**). Cette technique permet en effet d'obtenir un ArRNA à partir d'une culture d'*E. coli*, mais elle présente des limites techniques. En effet, lors du processus de purification, l'ultracentrifugation doit être privilégiée (50). La séparation sur colonne d'affinité permettrait de purifier des ArRNA de manière plus rapide et plus efficace. Afin de pouvoir purifier un ArRNA par IMAC, il est nécessaire de fusionner une étiquette poly-histidine (HIS-Tag) à la séquence codante de la CP ou de la maturase. L'affinité du His-Tag pour le  $\text{Cu}^{2+}$ , le  $\text{Ni}^{2+}$ , le  $\text{Co}^{2+}$  ou le  $\text{Zn}^{2+}$  permet de séparer rapidement le produit de fusion protéine-HIS-Tag de l'ensemble des protéines bactériennes avec une pureté pouvant atteindre 95 % grâce à la chromatographie d'affinité aux chélates métalliques (51).

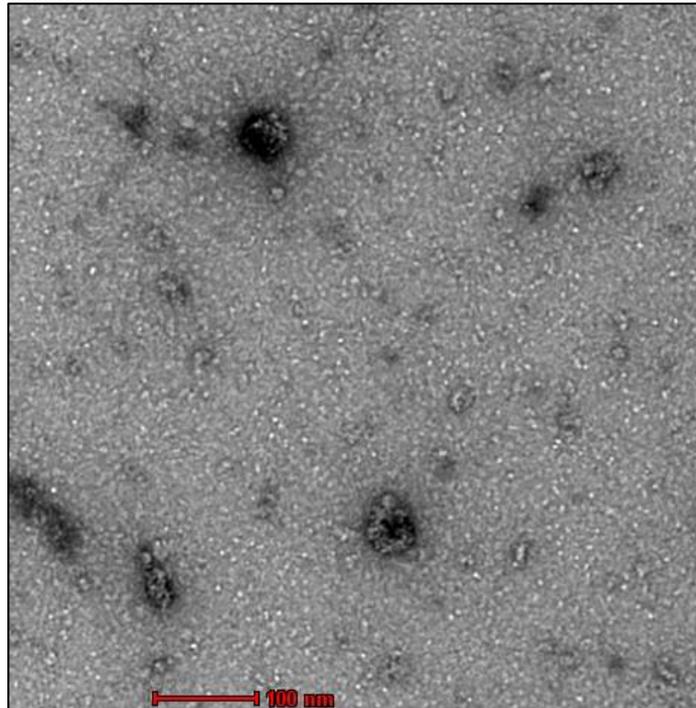
Selon l'article de Mikel *et al.* (50), l'insertion d'un HIS-tag dans la séquence de protéine de capsid (CP) lors de la production VLPs MS2 ne permet pas une purification efficace (**Figure 18B**). En effet, la présence d'un HIS-Tag sur une boucle exposée de la CP peut perturber l'auto assemblage en particules fonctionnelles. En conséquence, l'introduction d'un HIS-tag et l'expression du système MS2 VLP-HIS-tag dans un système de double expression à partir d'un plasmide n'entraînent pas la production de VLP. L'article propose donc une solution consistant à utiliser deux protéines de capsid (2CP) fusionnées, dont seule la séquence de la deuxième CP porte un HIS-Tag (**Figure 18C**). En utilisant un système maturase-CP-CP\_HISTAG, les auteurs de l'article ont mis en évidence qu'il est possible de produire des VLPs MS2 HIS-taguées contenant 90 HIS-tags dans la structure de la capsid exposée à l'extérieur de la particule, permettant ainsi une purification IMAC. Il est essentiel de noter que les VLPs obtenues ne se distinguent pas morphologiquement des VLPs purifiées grâce aux méthodes utilisées précédemment (50).



### Figure 18 : Production et purification des Armored RNA

Schéma du processus menant à la production d'une particule pseudovirale via un ARN recombinant. **A)** représentation schématique globale du processus de production *in vitro* d'un ARN encapsidé. **B)** la présence d'un HIS Tag sur une CP ne permet pas de générer de VLPs. **C)** le format CP-CP<sub>HIS</sub> permet la production et la purification de VLPs. Prom : promoteur de transcription ; Term : terminateur de transcription. CP : Coat Protein (protéine d'enveloppe) ; ARNm : ARN messenger ; RBS : Ribosome Binding Site (site de liaison au ribosome) ; PAC : structures d'emballage.

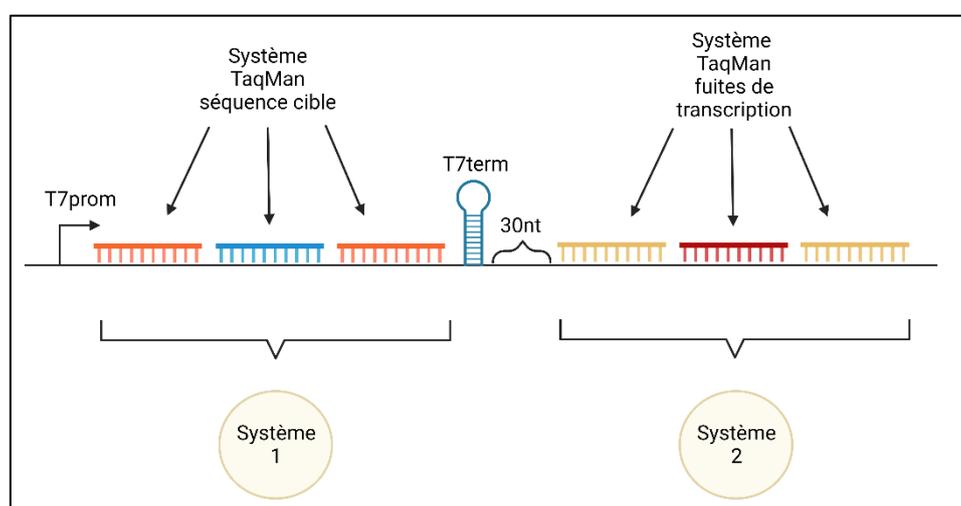
Au laboratoire, les ArRNA sont produits et purifiés grâce à cette approche. La purification par IMAC permet de produire des ARN encapsidés avec des rendements allant jusqu'à 2 milligrammes de pseudo-particules par culture de 200 ml et permettant la réalisation de dizaines de milliers de contrôles de RT-qPCR par préparation. Les ArRNA ainsi produits ont été analysés par microscopie électronique (**Figure 19**). On peut difficilement identifier des particules de formes icosaédriques et d'une taille d'environ 20-30 nanomètre, en adéquation avec la littérature (30, 31).



**Figure 19 : Visualisation des ArRNA par microscopie électronique**  
*Images obtenues après analyse en microscopie électronique en coloration négative de la préparation d'ArRNA. L'échelle est inscrite en rouge en bas de l'image.*

### 2.5.2.2. Quantification de la fuite de terminaison de transcription.

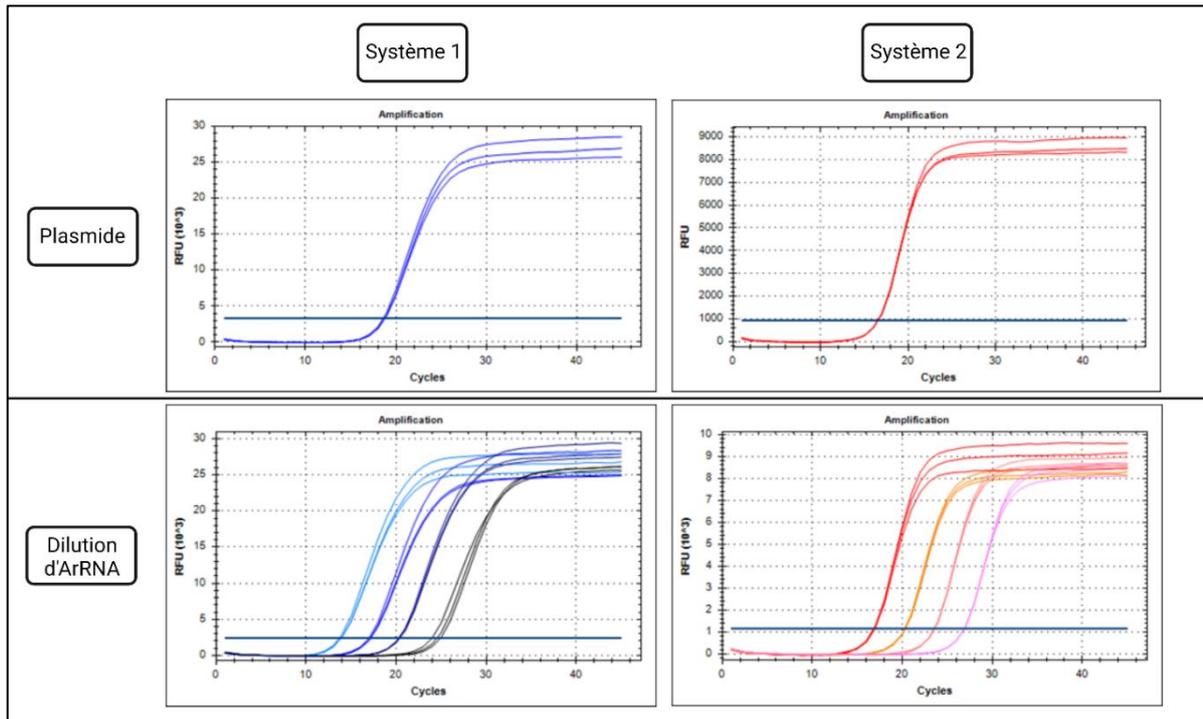
Comme pour la production d'IVT, il est possible que les ArRNA présentent des fuites de terminaison de transcription *in vivo*. La quantité d'ArRNA n'étant pas compatible avec un dosage par spectrophotométrie ou par une analyse en électrophorèse capillaire, une autre méthode a été développée pour faire une quantification relative par RT-qPCR de la séquence cible ARN et de la séquence en aval du T7term, qui serait révélateur de la présence fuite de terminaison ayant engendré des transcrits plus longs qu'attendus. Nous avons développé deux systèmes, le premier (Système 1) ciblant la séquence cible, et le deuxième (Système 2) ciblant une séquence plasmidique, hors de la région transcrite attendue (**Figure 20**).



**Figure 20 : Design du système TaqMan**

Schéma représentant la position des amorces et sonde utilisées pour déterminer les fuites de transcription. Système 1 : séquence cible, Système 2 : séquence plasmidique. T7prom : Promoteur T7. T7term : terminateur T7. Brins jaunes et oranges : amorces. Brins rouges et bleus : sonde. Figure générée avec BioRender

Un ArRNA a été extrait, dilué en cascade au dixième, puis analysé en RT-qPCR avec les deux systèmes décrits. Le plasmide ayant servi de matrice pour la synthèse de cet ArRNA a été utilisé comme témoin positif pour l'amplification. La **Figure 21** présente les courbes d'amplification obtenues après PCR sur le plasmide et l'ArRNA avec le système 1 et 2. Il est notable que les valeurs CT pour les deux systèmes sur l'ADN sont proches. Les CT obtenus entre le système 1 et 2 pour l'amplification de l'ArRNA peuvent donc être comparés et nous pouvons donc procéder à l'analyse des fuites de terminaison de transcription lors de la synthèse de l'ArRNA.



**Figure 21 : Évaluation des fuites de transcription lors de la production de l'ArRNA**

*Courbes d'amplification obtenue par RT-qPCR sur un ArRNA, avec le système 1 et le système 2. Dilutions en cascade au 1/10<sup>ème</sup>. PCR réalisée en triplicat. Abscisse : nombre de cycles PCR. Ordonnée : Relative Fluorescence Units (RFU).*

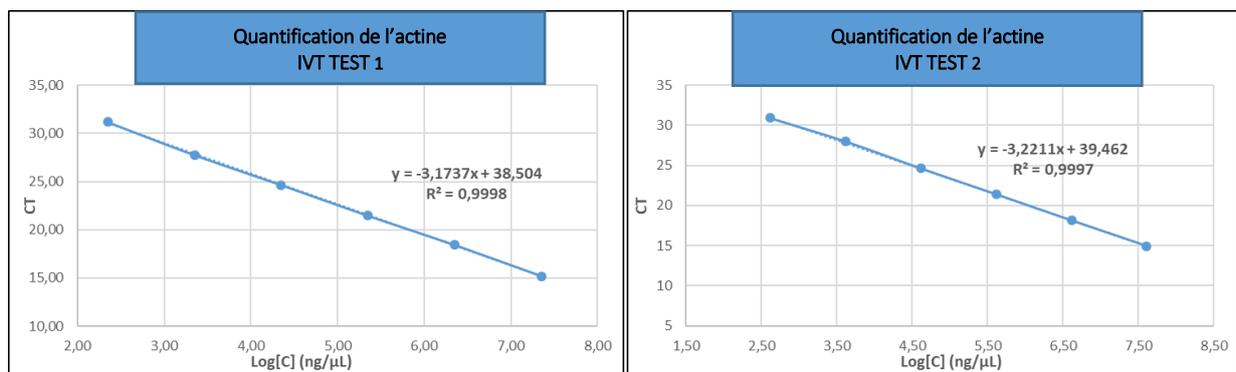
Nous pouvons dans un premier temps noter que les résultats sont positifs pour les deux systèmes d'amplification, signifiant qu'il y a de l'ARN. Les courbes d'amplification pour le Système 2 sur l'ArRNA sont synonymes d'une présence d'ARNs « longs », c'est-à-dire contenant la séquence située après le T7term. Il y a donc une fuite de terminaison de la transcription *in vivo* également. Pour toutes les dilutions de l'ArRNA, nous obtenons une différence moyenne d'environ 3.42 CT entre le Système 1 et le Système 2. Sachant qu'une différence de 3,2 CT correspond à une différence d'un Log de 10, nous pouvons conclure qu'il y a environ 10 fois moins d'ARNs « longs » transcrits que d'ARNs « courts » contenant uniquement la séquence d'intérêt dans l'ArRNA. Ces résultats suggèrent que 90% des transcrits sont correctement

stoppés sur le T7term et 10% correspondent à des transcrits plus longs. Pour affiner ces résultats, il aurait fallu produire les IVT correspondants, les quantifier, et produire une courbe de titration, pour pouvoir faire une comparaison quantitative fine. Faute de temps, et compte tenu des proportions favorables aux transcrits attendus, nous avons validé la construction avec le T7 terminateur en l'état.

### 2.5.2.3. Normalisation de la quantité de matériel génétique dans les ArRNA

Les ArRNA n'étant pas compatibles avec un dosage spectrophotométrique, nous avons mis en place une méthode alternative afin de pouvoir évaluer la quantité d'ARN dans ce type de contrôle. La séquence d'actine, présente dans tous les ArRNA, permet en théorie de normaliser la quantité d'ARN dans tous les contrôles. En effet, les séquences en nucléotides placées en amont de la séquence d'actine peuvent induire de potentiels effets de matrice sur le dosage. Nous avons donc réalisé des tests de référence sur le gène d'actine afin de générer une méthode de quantification universelle à tous les contrôles positifs (Figure 12).

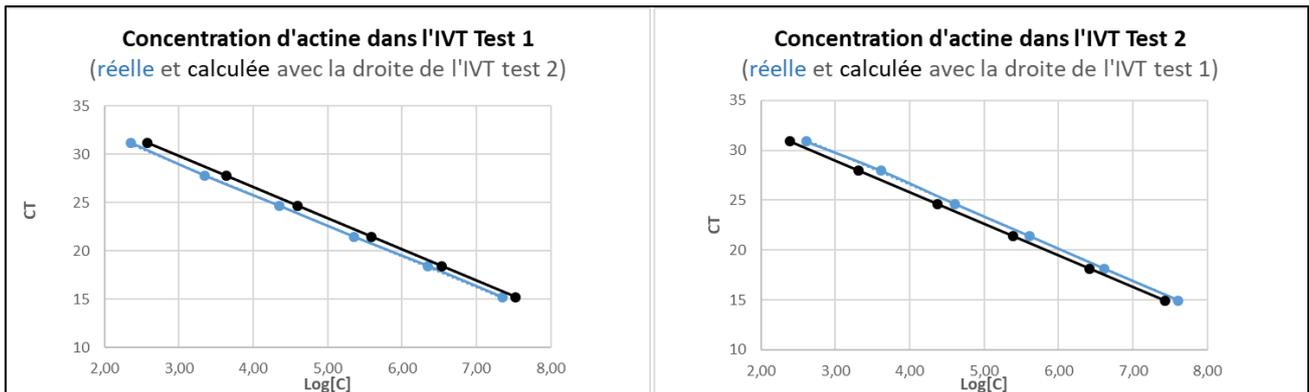
Afin de mettre en place une méthode de quantification des ArRNA, deux IVT ont été synthétisés (appelé ici « IVT Tests »), contenant une séquence en nucléotides de composition et de taille différente, suivi d'une séquence d'actine. Le nombre de copies d'ARN/ $\mu$ L de ces IVT ont été calculés en fonction de leurs dosages par spectrophotométrie. Une série de dilution en cascade de 10 en 10 de ces IVT a été réalisée, puis analysée par RT-qPCR avec une système TaqMan spécifique à la séquence d'actine. Deux droites de régression ont ainsi été générées (**Figure 22**).



**Figure 22 : Régressions linéaires des quantifications par RT-qPCR de l'ARN d'actine.**

*Régressions linéaires obtenues après analyse en RT-qPCR en triplicat des IVT tests 1 et 2, avec un système d'amplification spécifique à l'actine. Abscisse : Logarithme du nombre de copies d'ARN/ $\mu$ L. Ordonnée : CT obtenus après amplification.*

Afin de valider cette méthode de quantification, le nombre de copies d'ARN cible actine/ $\mu\text{l}$  contenue dans l'IVT Test 1 a été calculée avec l'équation de la droite de régression de l'IVT Test 2 et inversement. Ces quantités ont ensuite été comparées avec leur nombre de copies réelles. Nous pouvons voir sur la **Figure 23** que les deux droites de régressions sont quasiment superposées : la concentration réelle et la concentration calculée *via* la droite de régression de l'autre IVT Test sont proches. Ce test permet donc de conclure que la séquence en nucléotide présente en amont de la séquence d'actine n'induit pas de biais de quantification de ce transcrit.



**Figure 23 : Évaluation du test de normalisation de la quantité d'ARN d'actine dans un Armored RNA**

*Régressions linéaires obtenues après calcul du nombre de copie d'ARN d'actine en fonction de la régression linéaire de l'autre IVT test. Courbes bleues : concentrations exactes. Courbes noires : concentrations calculées avec la régression linéaire de l'autre IVT test. Abscisse : logarithme de la concentration. Ordonnée : CT obtenus après amplification de l'IVT testé.*

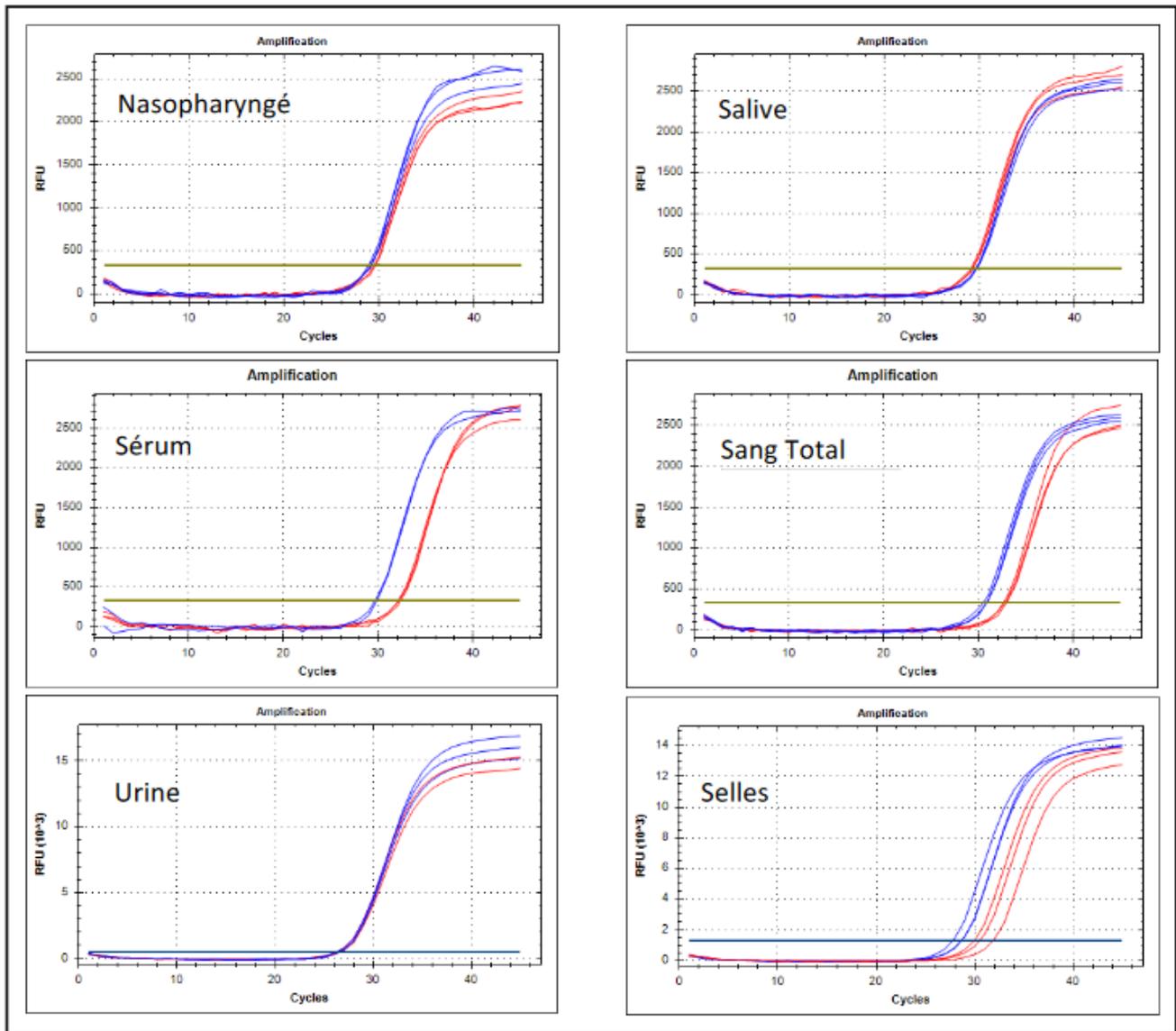
Pour évaluer la quantité d'ARN dans un contrôle de type ArRNA, il suffira donc de l'analyser par RT-qPCR ciblant la séquence d'actine et de calculer le nombre de copies d'ARN / $\mu\text{l}$  grâce aux équations des droites de régressions ainsi générées. Cette méthode pourra être utilisée pour quantifier tous les ArRNA produits au laboratoire, quelle que soit la composition ou la taille en nucléotides et ce à partir de la séquence ARN et hexogène d'actine et du système de quantification RT-PCR développé.

#### 2.5.2.4. Développement d'un contrôle MS2 synthétique

Le phage MS2 est couramment ajouté aux échantillons biologiques pour valider l'extraction de l'ARN viral. Nous avons cherché à remplacer le phage MS2, d'origine commerciale, par un ArRNA contenant la séquence cible utilisée pour la détection de MS2. Pour développer et valider ce contrôle, il est nécessaire de : i) calibrer la quantité d'ArRNA pour qu'après amplification par RT-qPCR on obtienne le même CT que celui obtenu avec un phage MS2 et ii) valider les résultats d'extraction à partir d'échantillons biologiques couramment utilisés.

L'ArRNA MS2 est dans un premier temps dilué en cascade et titré dans de l'eau (10µl d'ArRNA dans 200µl d'eau) afin de déterminer une dilution permettant d'obtenir une courbe d'amplification d'une valeur seuil de 30 CT (résultat non présenté). Il est par la suite titré dans différents échantillons : de la salive, un prélèvement nasopharyngé, du sérum, du sang total, de l'urine et du clarifié de selles, afin d'identifier d'éventuels effets de matrice sur l'ARN encapsidé, en comparaison du phage MS2. Ces échantillons ont par la suite été amplifiés par RT-qPCR (**Figure 24**). Ces résultats montrent que les données obtenues sont répétables, tant pour l'ArRNA MS2 que pour le phage MS2 (les trois répliques ont la même valeur de seuil), avec des CT aux alentours de 30. Comme appréhendé, des différences d'effet de matrice dans certaines conditions sont observées : un retard d'environ 3 CT est observé pour l'ArRNA dans le sérum, le sang total et les selles par rapport au phage MS2. Un ARN encapsidé pourrait être moins stable ou moins bien extrait dans ces matrices contenant des inhibiteurs (52). En ce qui concerne les prélèvements nasopharyngé, salivaires et urinaires, l'amplification de l'ArRNA MS2 a la même valeur seuil que celle du phage MS2. L'ArRNA MS2 peut donc se substituer au phage MS2 dans toutes les matrices testées, dans la dilution de travail prédéfinie.

Une phase de développement d'un contrôle d'extraction remplaçant le phage MS2 a donc pu être conduite, permettant de passer à une phase de production et à la génération de plusieurs milliers de tests sous forme lyophilisée, mis à la disposition de la communauté scientifique dans le catalogue EVA-G (European Virus Archive – GLOBAL : <https://www.european-virus-archive.com/>).



**Figure 24 : Quantification par RT-qPCR de l'ArRNA MS2 dans différents échantillons biologiques.**

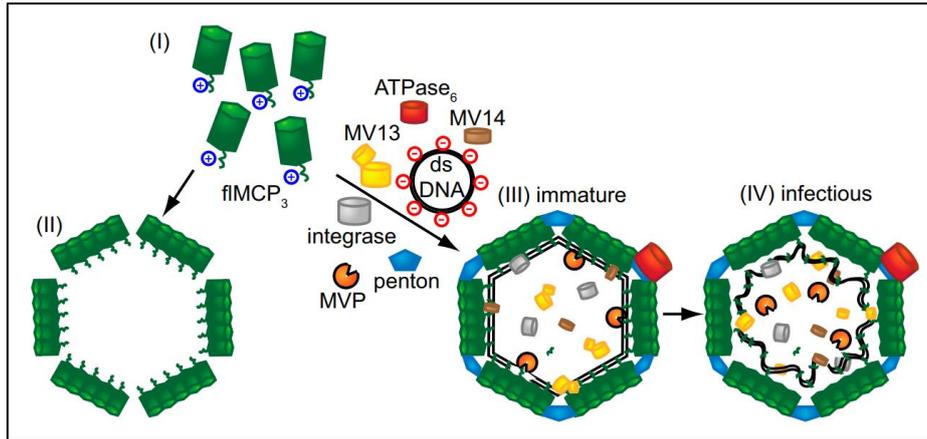
*Courbes bleues : phage MS2. PCR réalisée en triplicats. Courbes rouges : ArRNA. Abscisse : nombre de cycles PCR ; Ordonnée : fluorescence relative (RFU)*

### 2.5.3. ADN encapsidés

Afin de synthétiser des ADN encapsidés, nous avons envisagé deux techniques. La première reproduisant le système MS2 déjà décrit (44), mais avec des capsides recombinantes simples. En l'absence de publications scientifiques décrivant des phages à génome ADN avec des capsides simples à reconstruire, nous avons exploré les virophages, qui sont des virus capables d'infecter les virus géants. En parallèle, nous nous sommes appuyés sur l'utilisation de phagémides et leur encapsidation dans le phage M13, technologie mieux décrite mais qui ne permet que de produire des ArDNA simple brin.

#### 2.5.3.1. Capsides recombinantes

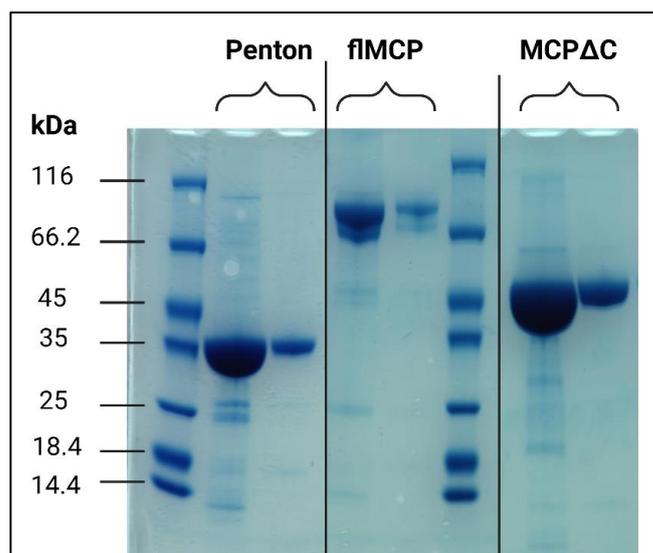
Le mavirus, un virophage marin capable d'infecter le phagotrophe flagellé marin *Cafeteria roenbergensis*, est composé de protéines de capsides capables de s'auto assembler (51) (**Figure 25**). La capside du mavirus est composée d'une protéine appelée fIMCP (Full length Major Capsid protein, protéine majeure de capside, taille complète), capable de s'auto-assembler en VLP. Son domaine C-terminal chargé positivement interagit avec l'ADN viral chargé négativement. Par la suite, la fIMCP s'associe avec des protéines appelées Penton, et l'interaction avec une maturase permet le détachement du domaine C-terminal pour produire une VLP mature et infectieuse. La MCP du mavirus est clivée par une protéase à cystéine virale au niveau de sites de l'extrémité C-terminale qui sont conservés dans diverses MCP de virophages, ce processus est impliqué dans la maturation et l'assemblage de la particule virale. Les auteurs de cet article ont exploré la production d'une protéine appelée MCP $\Delta$ C qui est une fIMCP déléetée de son domaine C-terminal. La production de VLP à partir de MCP $\Delta$ C et de penton pourrait permettre de s'affranchir de l'utilisation d'une maturase pour former des VLP matures. Il est à noter cependant que l'élimination du C-terminal peut empêcher l'interaction VLP-ADN et par conséquent conduire à la non-encapsidation de l'ADN.



**Figure 25 : modèle d'assemblage des virophages, et fonctions des MCP**

(I) La fIMCP est essentiel pour l'assemblage et peut s'auto-assembler en « boules » en l'absence d'autres protéines du virophage (II). (III) Tous les gènes de mavirus sont précédés par le même motif promoteur CroV tardif (51), ce qui suggère une expression et un assemblage parallèles. Le génome ADN double brin chargé négativement interagit avec le C-term fIMCP chargé positivement. (IV) Le clivage du fIMCP par une protéase détache le complexe ADN-C-term de la capsid. MV13 et MV14 : Produits des gènes 13 et 14 du mavirus. Figure extraite de l'article (54)

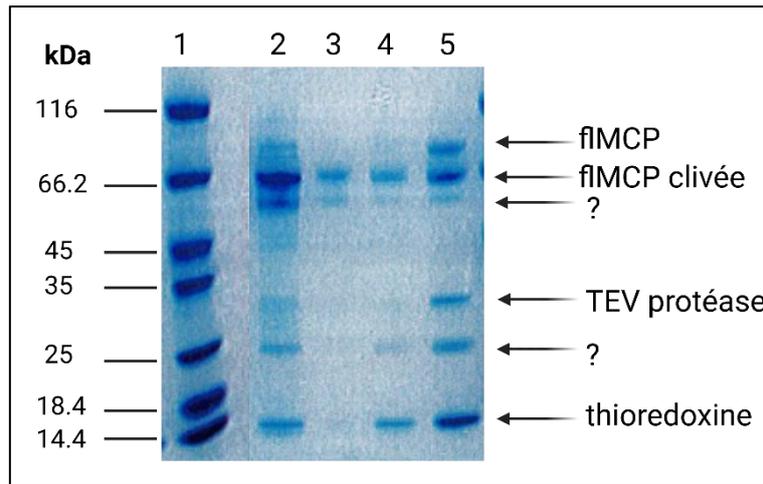
Nous avons donc produit trois protéines de manière recombinantes : fIMCP, MCPΔC et le penton afin de réaliser des assemblages fIMCP/Penton ou MCPΔC/penton. Un HIS-Tag a été ajouté aux séquences de ces trois protéines afin de permettre une purification IMAC. La séquence de la thioredoxine est ajoutée en amont de la séquence codante de la fIMCP (séquences présentées en Annexe 5). La thioredoxine a été montrée comme ayant la propriété d'accroître la solubilisation des protéines étant à l'origine insolubles (55). Un site de clivage à la protéase du virus de la gravure du tabac (Tobacco Etch Virus, TEV) est ajouté en aval du HIS-Tag afin de l'éliminer et d'éliminer la thioredoxine après IMAC. Les produits IMAC du Penton, la fIMCP et la MCPΔC ont ensuite été analysés par SDS-PAGE (Figure 26). Les protéines obtenues sont relativement pures, et correspondent aux poids moléculaires théoriques attendus : 80kDa pour la fIMCP, 58.5 kDa pour la MCPΔC et 37 kDa pour le penton. 1.5 mg de fIMCP, 8.4 mg de MCPΔC et 5 mg de penton ont été obtenus pour un volume de 200 ml de culture : la synthèse de la fIMCP produit des rendement nettement moins importants que les deux autres protéines.



**Figure 26 : Analyse par SDS-PAGE des protéines recombinantes purifiées**

*SDS-PAGE des fractions récoltées après purification des protéines. Les protéines sont éluées dans deux fractions successives de 5ml, correspondant aux pistes de gauche et de droite pour chaque protéine. Gel coloré au bleu de Coomassie.*

Nous avons dans un premier temps choisi de former des VLP en utilisant la fIMCP. La première étape du protocole est de réaliser un clivage à la TEV protéase, afin d'enlever la thioredoxine, pouvant gêner la formation de VLP. Le produit de clivage de la fIMCP par la TEV protéase a été purifié par IMAC afin de ne récupérer que la fIMCP clivée du C-terminal. Celle-ci sera récupérée lors de la première étape de purification, c'est-à-dire lors du premier passage sur la colonne d'affinité (flow-through) et non à l'étape d'élution, car la fIMCP clivée n'est plus HIS-taguée. Les différentes fractions récupérées par IMAC ont été analysées par SDS-PAGE (**Figure 27**). La fIMCP clivée est récupérée dans le puit n°3 sur le SDS-PAGE : la bande correspondant à la fIMCP clivée produit intensité faible, et nous observons deux bandes correspondant à un poids moléculaire d'environ 60kDa et 25 kDa. La première est potentiellement un clivage non attendu de la fIMCP qui dans le cycle viral est maturée par une protéase virale (**Figure 25**). La seconde pourrait être un contaminant habituel des purifications IMAC (SlyD) dont l'identité pourrait être déterminée par spectrométrie de masse. De plus, nous observons une grande quantité de fIMCP clivée dans la fraction d'élution (puits 5), supposant donc que la protéine n'a pas été entièrement digérée par la TEV protéase. Du fait de la faible quantité de fIMCP clivée recueillie, la suite des manipulations a été effectuée uniquement avec la MCPΔC et le penton. D'autres tests de clivage à la TEV protéase sur la fIMCP sont actuellement en cours.



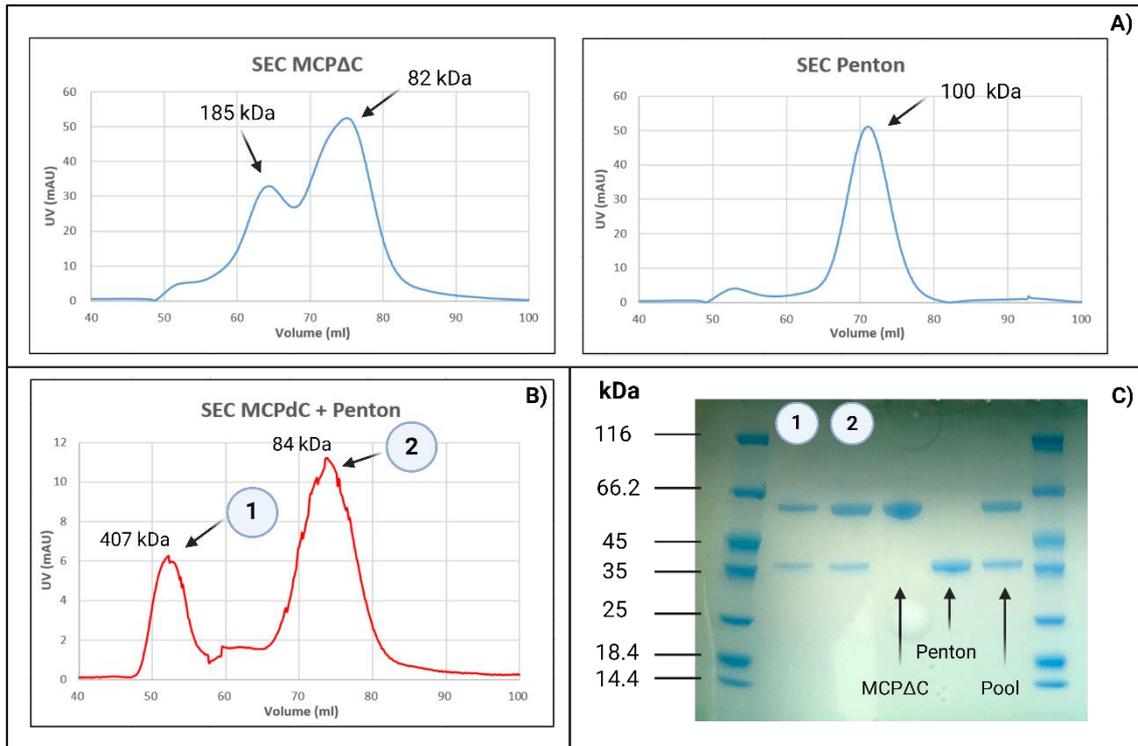
**Figure 27 : Produits IMAC du clivage de la fIMCP par les TEV protéases**

Analyse par SDS-PAGE des différentes fractions IMAC de la fIMCP clivée à la TEV. Puit 1 : standard de poids moléculaire. 2 : premier passage dans la colonne (flow through). 3 : premier lavage. 4 : deuxième lavage avec ajout de 10% de tampon B. 5 : Elution. Gel coloré au bleu de Coomassie.

Par la suite, nous avons essayé de former des VLPs en ajoutant le penton et la MCPΔC. Pour ce faire, ces deux protéines ont été purifiées par Chromatographie d'Exclusion de Taille (SEC), afin de récupérer une protéine la plus pure possible. Les produits de purifications récupérés correspondant aux pics indiqués par une flèche sur la **Figure 28 A**. La MCPΔC, ayant un poids moléculaire théorique de 58.5 kDa, élue à 62.6 ml et 73.8 ml, correspondant à des poids moléculaires apparents de 185 kDa (possible trimérisation) et 82 kDa (possible monomère ou dimère). Dans la littérature, il a déjà été montré que la MCPΔC s'organise en trimère (54). Le penton, ayant un poids moléculaire théorique de 37 kDa, élue à 71 ml, correspondant à un poids moléculaire apparent de 100 kDa (possible dimérisation ou trimérisation). Ces résultats ne sont pas en accord avec ceux décrits dans la littérature car le penton devrait s'organiser en pentamère (54).

Ensuite, la MCPΔC et le penton séparés par chromatographie d'exclusion de taille (Size Exclusion Chromatography, SEC). ont été mélangés en quantité équimolaires puis incubés pendant trois heures à température ambiante, puis à nouveau analysés par SEC (**Figure 28 B**). La séparation du mélange de la MCPΔC et du penton élue en deux fois : la première élution à un volume de 51.7 ml, correspondant à un poids moléculaire apparent de ~400 kDa, et la deuxième à un volume de 73.5ml, correspondant à un poids moléculaire apparent de 85 kDa. Ces résultats d'analyse par SEC couplé avec une analyse par SDS-PAGE (**Figure 28 C**) permet de conclure que la première élution correspond à un complexe MCPΔC-penton. En effet, nous observons sur le SDS-PAGE la présence des deux protéines. La deuxième élution quant à elle pourrait correspondre à une co-élution de la MCPΔC et du penton, compte tenu du poids

moléculaire correspondant. Ces premiers résultats sont donc en faveur d'un assemblage de la MCPΔC avec le penton.



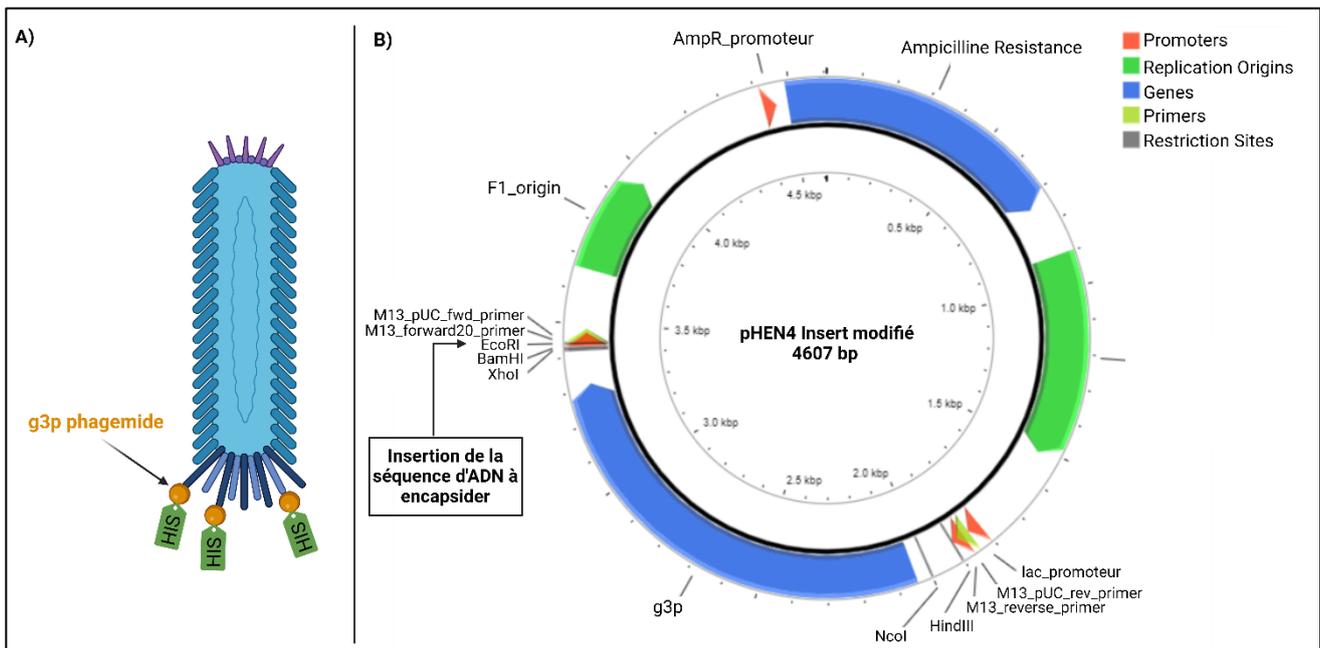
**Figure 28 : Analyses de l'assemblage des protéines MCPΔC et penton**

**A)** Analyse par SEC de MCPΔC et penton. **B)** Analyse en SEC de MCPΔC et penton assemblés. **C)** Analyse par SDS-PAGE des fractions 1 et 2 récoltées après analyse par SEC. MCPΔC et penton : témoins positifs, protéines seules purifiées. Pool : mélange de MCPΔC et penton. Gel coloré au bleu de Coomassie.

Afin de former un ArDNA, une séquence d'ADN recombinant a été intégrée à ce complexe protéique. Cette intégration a été effectuée en réalisant une sonication sur ce complexe afin de désassembler les protéines, en ajoutant l'ADN puis en incubant le mélange à température ambiante afin de permettre au complexe MCPΔC de se reformer. Cette expérimentation a également été effectuée sans sonication, dans l'éventualité où l'ADN s'intégrerait naturellement au complexe protéique. Un test de dégradation à la DNase est actuellement en cours afin d'évaluer la formation d'un complexe ADN-MCPΔC. Si l'ADN est intégré au complexe protéique, il sera alors insensible aux tests de dégradation et sera observable sur gel. Une analyse par microscopie électronique du mélange de la MCPΔC et du penton sera également envisagée afin de visualiser ou non la formation d'une pseudo-particule virale. Les résultats des tests de dégradation à la DNase couplés à ceux de la microscopie nous permettront de conclure sur la faisabilité d'encapsider un ADN recombinant *via* cette méthode.

### 2.5.3.2. Phagémide encapsidé

En parallèle, nous avons choisi le phage M13 comme second candidat pour la production d'un ADN encapsidé. Nous avons utilisé le phagémide pHEN4 (séquence présentée en [Annexe 7](#)). Une séquence, appelée « insert pHEN4 », a été clonée entre les sites de restriction NcoI et EcoRI du pHEN4, afin d'y intégrer une séquence codante pour la protéine g3p, qui se situe de manière aléatoire à l'apex des VLPs, couplée à un HIS-Tag pour une purification IMAC (**Figure 29 A**) (séquence détaillée en [Annexe 9](#)). La séquence cible de l'ADN simple brin à encapsider (ici, une séquence cible du Virus de l'hépatite B, HBV) a été clonée dans le phagémide *via* le site de restriction XhoI (**Figure 29 B**). Les étapes de restrictions/ligation ont été effectuées par un prestataire extérieur au laboratoire.

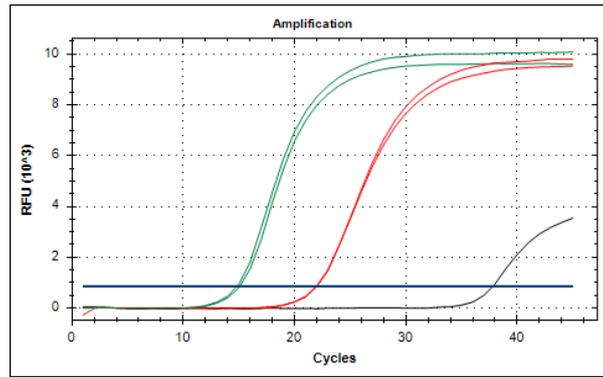


**Figure 29 : Expression d'une VLP G3p-HIS-Taguée à l'apex du phage.**

A) Représentation schématique d'une VLP taguée avec g3p-HIS-Tag. Figure générée avec BioRender.

B) Carte plasmidique du phagémide pHEN4 de séquence modifiée. Carte générée via Plasmapper 3.0.

Un premier ADN encapsidé a été synthétisé selon le protocole décrit dans le matériels et méthodes. Une fois le protocole de synthèse terminé, le produit final a été dosé par spectrophotométrie à 280nm. Nous obtenons un résultat de  $DO_{280nm}$  de 0.16 mg/ml, pouvant indiquer la présence d'une protéine. Ce même produit a par la suite été analysé en qPCR, avec des amorces spécifiques à l'ADN encapsidé (**Figure 30**). La qPCR est positive, avec des courbes d'amplification aux alentours de 22 CT (courbes rouges). Ces premiers résultats nous indiquent donc la présence de protéine et celle de l'ADN cible.



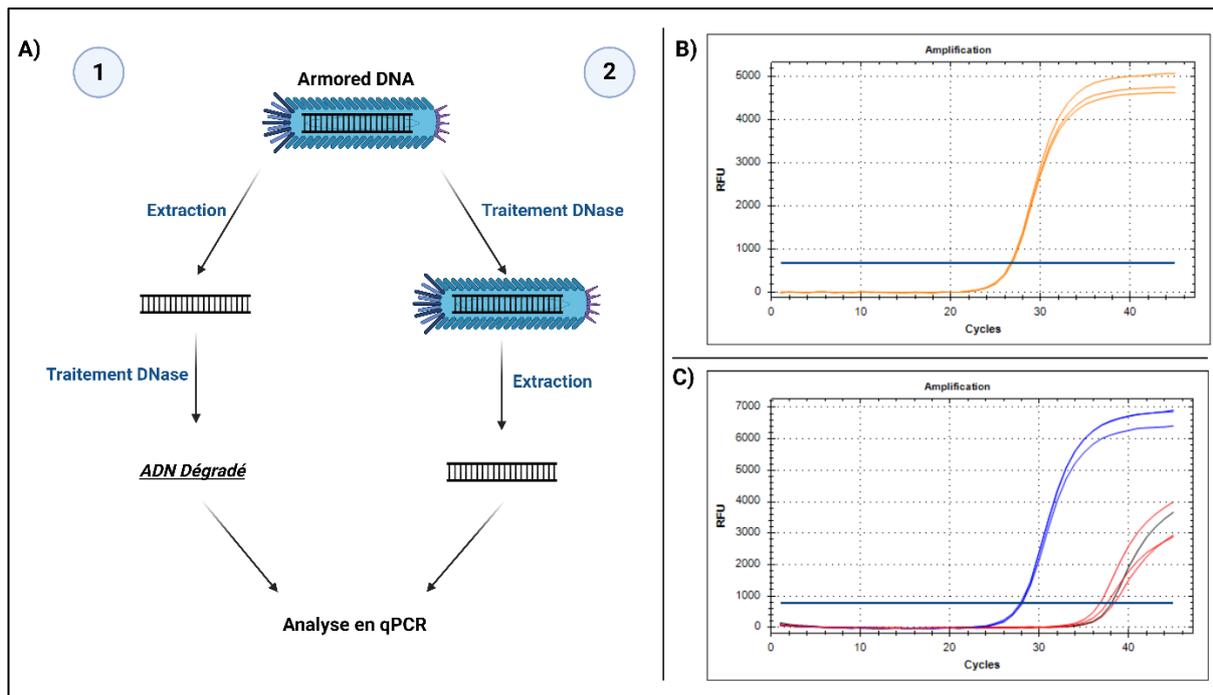
**Figure 30 : Vérification de la présence de l'ADN cible.**

*Courbes d'amplification réalisées en duplicat obtenues après analyse en qPCR de l'ArDNA synthétisé. Courbes vertes : plasmide contenant la séquence cible (témoin positif). Courbes rouges : ArDNA. Courbe noire : H<sub>2</sub>O (témoin négatif).*

Pour s'assurer de l'intégration de l'ADN cible à ces protéines, nous avons réalisé des tests de dégradation à la DNase ; en effet, la capsid est supposée protéger l'ADN et le rendre insensible à une dégradation par une DNase. Le test a été réalisé de la manière suivante (**On peut** en conclure que le matériel génétique est protégé par une capsid).

Figure 31 **A)** : la première étape consiste à extraire l'ADN contenu dans l'ArDNA, puis de traiter à la DNase afin de constituer un contrôle positif de dégradation ; la deuxième constitue le test de dégradation à proprement parlé, c'est-à-dire que l'ArDNA est directement traité à la DNase et l'extraction de l'ADN se fait a posteriori pour pouvoir analyser l'état de dégradation de l'ADN par qPCR avec un système TaqMan spécifique à l'ADN encapsidé(**On peut** en conclure que le matériel génétique est protégé par une capsid.

Figure 31 **B et C**). Pour le premier test, lorsque l'ADN nu est dégradé par la DNase, nous obtenons des courbes d'amplifications (courbes rouges) superposées à celles du contrôle négatif (H<sub>2</sub>O, courbe noire). La PCR est considérée négative pour ce test, ce qui valide l'efficacité de la DNase. Pour le second test, correspondant à la dégradation de l'ArDNA non extrait, nous obtenons des CT (courbes bleues) de mêmes valeurs que pour l'ArDNA non dégradé testé dans des conditions identiques (courbe orange). La DNase est donc inefficace sur un ArDNA, car le CT obtenu est le même qu'il soit traité à la DNase ou non. On peut en conclure que le matériel génétique est protégé par une capsid.

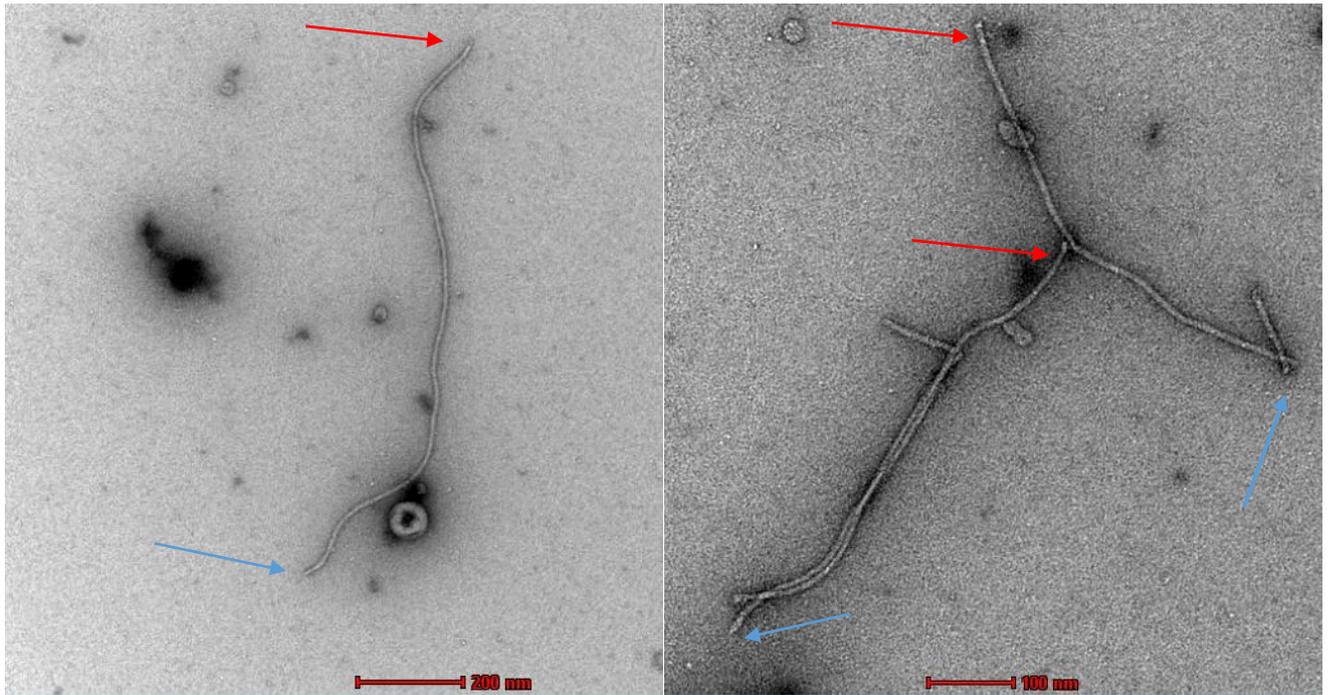


**Figure 31 : Mise en évidence de l'encapsulation de l'ADN.**

*A) Représentation schématique des deux tests de dégradation réalisés. B) Courbes d'amplification obtenues après qPCR de l'ArDNA après extraction sans traitement DNase (contrôle positif). PCR réalisée en triplicat. C) Résultats de l'analyse en qPCR réalisée en duplicat après les tests de dégradation à la DNase. Courbes rouges : résultats du test 1. Courbes bleues : résultat du test 2. Courbe noire : contrôle négatif (H<sub>2</sub>O).*

Ces premiers résultats sont donc en faveur d'une encapsulation de l'ADN. Afin de confirmer ces résultats, le produit obtenu a été analysé par microscopie électronique en coloration négative (**Figure 32**). La présence de particules allongées, composées d'une tête (flèche bleue) et d'une queue (flèche rouge) semblables à celles du phage M13 sont observées.

L'approche développée à partir du phage M13 et d'un phagémide contenant une séquence virale cible permet donc de générer des ADN synthétiques non infectieux – a priori simple brin - encapsidés dans une particule de phage M13.



**Figure 32 : Présence de particules phagiques dans les préparations d'ADN encapsidés.**

*Images obtenues après analyse en microscopie électronique en coloration négative de la préparation d'ArDNA.*

*L'échelle est inscrite en rouge en bas des images.*

Cependant, des améliorations devront être apportées au protocole de production. Dans un premier temps, il faudra augmenter le rendement de production d'ArDNA. En effet, comme nous pouvons le voir sur la **Figure 30**, un ADN encapsidé pur a une valeur d'amplification d'environ 22 CT, pour un volume de production légèrement inférieur à 1 ml. Sachant qu'il est d'usage en diagnostic d'utiliser un contrôle positif ayant une valeur d'environ 28 CT, et connaissant les pertes de matériel lors de la lyophilisation, une préparation d'ArDNA ne permettrait de générer seulement 10 flacons lyophilisés de 100 tests/flacon (contre plusieurs centaines de flacons pour les ArRNA). Étant donné qu'une production d'ADN encapsidé représente plus d'une semaine de travail, la méthode de synthèse d'un ArDNA n'est pas totalement compatible avec une production à grande échelle. Dans un second temps, il nous faudra aussi mener des tests d'évaluation de stabilité de cet ADN encapsidé. Même si par définition un ADN est censé être stable dans le temps et à certaines températures, nous n'avons pour l'instant pas mené d'étude sur la différence de stabilité entre un ADN nu et un ADN protégé. Ensuite, j'ai présenté ici une preuve de concept pour un ADN synthétique d'environ 400 paires de base, mais nous ne connaissons pas la taille maximale qu'il est possible d'encapsider avec cette technique. Pour finir, une étude de la contamination de génome phagique résiduelle dans une préparation d'ArDNA devra être menée.

### **3. Développement de techniques de génotypage par endonucléase.**

#### **3.1. Les différentes techniques d'identification.**

Le génotypage est une technique permettant de déterminer l'identité d'une variation génétique, à une position spécifique sur la totalité ou une partie du génome, pour un individu ou un groupe d'individus donné. Plusieurs techniques peuvent être utilisées à des fins de génotypage, comme la technique de Polymorphisme de longueur des fragments de restriction (Restriction Fragment Length Polymorphism (RFLP)), consistant à découper l'ADN en fragments avec une enzyme de restriction, à les séparer par électrophorèse et ensuite les hybridés avec une sonde pour détecter les variations génétiques ; ou la technique de génotypage par microsatellites, consistant à amplifier les régions contenant des microsatellites par PCR, puis à séparer les fragments par électrophorèse : la taille des microsatellites est déterminée en comparant les tailles des fragments entre différents individus, permettant ainsi l'étude de la diversité génétique et l'identification d'individus. Il existe d'autres techniques de génotypage, mais seules les techniques de séquençage et de génotypage par endonucléase seront développées dans ce travail.

##### 3.1.1. Le séquençage.

Le séquençage est une méthode consistant à déterminer l'ordre de succession des différents nucléotides d'un génome. Le séquençage permet de déterminer la séquence d'un génome inconnu (séquençage *de novo*), d'identifier une ou différentes espèces déjà connue(s) présente(s) dans un échantillon, ou encore de mettre en évidence l'apparition d'éventuelle(s) mutation(s).

La première génération de séquenceur apparaît dans les années 1970, utilisant la méthode Sanger. Cette méthode repose sur l'utilisation de nucléotides radiomarqués ou fluorescents bloquant l'élongation de la chaîne d'ADN par les polymérases ([Annexe 10](#)). Cette technique permet de traiter des petites séries d'échantillons de manière rapide et à faible coût. Cependant, elle ne permet de séquencer que des fragments assez courts (environ 1000nt), elle ne donne pas d'informations précises sur les sous-populations virales et ne permet pas de séquencer un échantillon inconnu.

La deuxième génération de séquençage apparaît environ 30 ans plus tard, au début des années 2010. Cette génération de séquenceur impose la fragmentation de l'ADN avant construction de banques. Les réactions de séquençage sont réalisées soit par fluorimétrie (Séquenceurs MySeq

et HiSeq, Illumina) soit par réaction chimique (séquenceurs Ion Torrent PGM et S5), et génèrent des millions de petites séquences d'environ 50 à 1000 paires de bases, permettant ainsi de reconstruire un génome complet. Ces appareils sont rapides, fiables et permettent de séquencer des échantillons inconnus (séquençage de novo).

La dernière génération de séquençage ne requière pas de fragmentation préalable de l'ADN. Les nucléotides sont directement détectés sans amplification. Cette détection se fait soit par passage dans un nanopore (un courant électrique est appliqué, et le flux sera bloqué d'une certaine manière en fonction du nucléotide présent dans ce pore), soit par fluorescence. Cette troisième génération de séquençage permet de s'affranchir de nombreux biais de manipulations, puisqu'il n'est pas nécessaire de réaliser de fragmentation et d'amplification préalable. Elle permet également de réduire le coût et le temps de manipulation. Ces séquenceurs sont de petite taille (à peine plus gros qu'une clef USB pour l'appareil MinION) et permettent ainsi d'être facilement transportables.

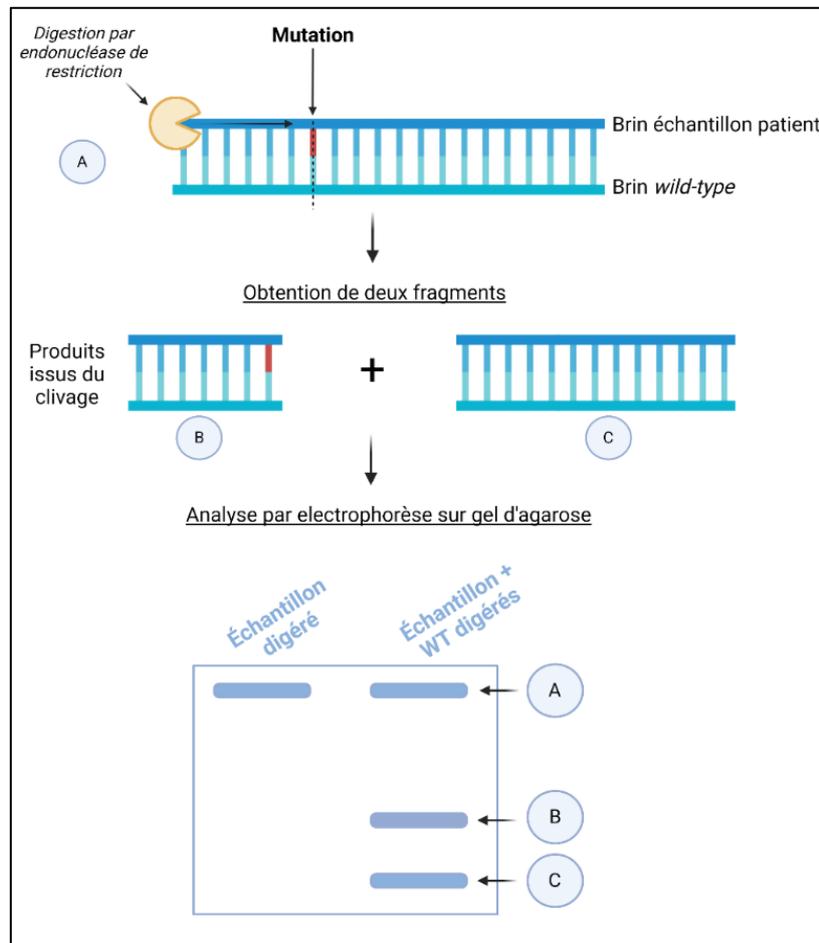
Toutes les générations de séquençages sont complémentaires entre elles et doivent être choisies en fonction de l'échantillon à analyser et du résultat souhaité. Dans le contexte du génotypage, l'utilisation du séquençage Sanger reste importante, après amplification d'une ou de plusieurs régions susceptibles de contenir des SNP (Single Nucleotide Polymorphism, polymorphisme d'un seul nucléotide). Cependant, la technologie Sanger n'est plus maintenue dans la majorité des laboratoires et l'utilisation de plateformes de prestation est devenue commune, impliquant un rendu de résultats plusieurs jours après le traitement des échantillons.

### 3.1.2. Le génotypage par endonucléase.

Le génotypage par endonucléase est une méthode d'identification utilisant des endonucléases mismatch-spécifiques pour analyser des séquences spécifiques d'ADN. Cette technique repose sur l'hybridation d'un ADN de référence avec un ADN échantillon dont le produit génère des « mis-matches » ou mésappariements (MM) en cas de SNP.

Le génotypage par endonucléase est réalisé de la manière suivante : l'ADN ou l'ARN est extrait de l'échantillon et est amplifié par (RT) PCR, afin de produire suffisamment de copies de matériel génétique. L'ADN viral amplifié est par la suite hybridé avec un ADN de référence, appelé « *Wild Type* ». On obtient un ADN ou ADN complémentaire double brin hétéroduplex, contenant un brin de l'échantillon à analyser et un brin *Wild Type*. L'endonucléase MM spécifique, reconnaît d'éventuels mésappariements entre les deux brins et coupe l'ADN de

manière franche. Ces fragments issus du clivage sont ensuite séparés par électrophorèse sur un gel d'agarose ou électrophorèse capillaire (**Figure 33**).



**Figure 33 : Représentation schématique du principe du génotypage par endonucléase de restriction.**

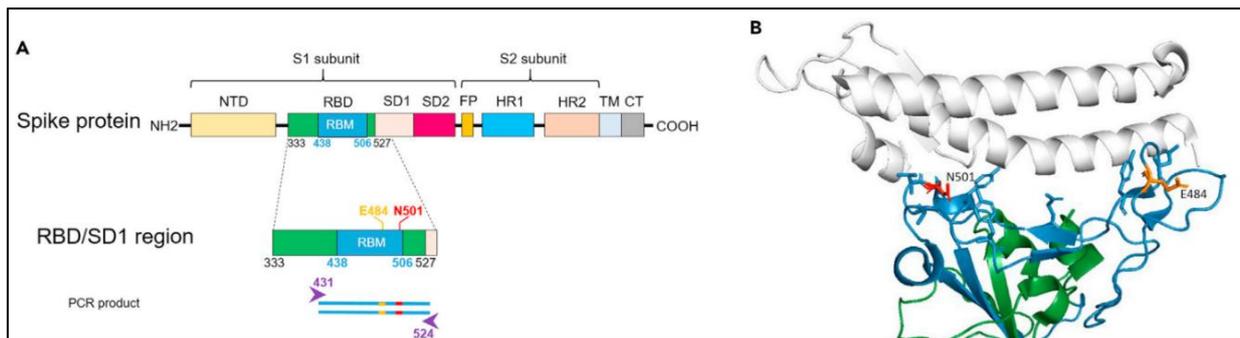
*Schéma généré avec BioRender.*

Parmi les kits commerciaux employant des endonucléases MM spécifiques et pouvant être utilisées à des fins de génotypage, on peut citer l'endonucléase Surveyor® Nuclease S. L'endonucléase intégrée dans le kit est extraite du céleri, et fait partie de la famille des nucléases CEL, endonucléases d'ADN végétales. Cette enzyme clive avec une grande spécificité du côté 3' de tout site de mésappariement dans les deux brins d'ADN, y compris toutes les substitutions de bases et les insertions/délétions jusqu'au moins 12 nucléotides (56). Cette endonucléase a été utilisée notamment pour cribler des gènes tels que BRCA1 et BRCA2 impliqués dans le cancer du sein héréditaire (57) ; pour la détection de mutations situées dans n'importe quel gène de l'ADN mitochondrial chez des enfants atteints d'encéphalomyopathie mitochondriale, permettant ainsi de réaliser des diagnostics prénataux (58), ou encore pour la surveillance de la dérive génétique du SARS-CoV-2 (59) (étude présentée dans ce travail).

### 3.1.2.1. Génotypage SARS-CoV-2

Cette partie de mon mémoire a fait l'objet d'une publication à laquelle j'ai contribué et été attribuée en deuxième auteure, intitulée « Endonuclease-based genotyping of the RBM as a method to track the emergence or evolution of SARS-CoV-2 variants », publiée dans le journal *iScience* en 2021 (59).

Depuis l'émergence du SARS-CoV-2, plusieurs mutations non synonymes (mutation induisant un changement d'acide aminé) ont été signalées dans la séquence codante de la protéine Spike, protéine clé pour l'entrée du virus dans les cellules hôtes. Certaines de ces mutations sont particulièrement surveillées parce qu'elles sont soupçonnées d'affecter les fonctions de la protéine et donc d'avoir un impact sur la biologie du virus. L'une de ces premières mutations fut une substitution (D614G) contribuant à l'augmentation de la charge virale dans les voies respiratoires supérieures avec une possible augmentation de la transmission (Plante et al., 2021). Plus tard, plusieurs variants définis comme des variants préoccupants (Variant of Concern, VoC) ont émergé et se sont propagés. Ces variants peuvent présenter une transmissibilité ou une gravité accrue de la maladie, une réduction de la séroneutralisation par les anticorps induits par une infection antérieure ou une vaccination, ou une résistance aux traitements thérapeutiques. Parmi les VoC, le premier - 501Y.V1 ou lignée B.1.1.7 - a été identifié au Royaume-Uni et a montré une transmission interhumaine accrue et une plus grande sévérité de la maladie par rapport à la souche d'origine (60). Ensuite, des variants des lignées B1.351 (501Y.V2), P.1 (501Y.V3) et B.617 ont été isolés et caractérisés en Afrique du Sud, au Brésil/Japon et en Inde, respectivement. Les variants B1.351 (501Y.V2) et P.1 (501Y.V3) présentent tous deux une résistance accrue à la neutralisation des anticorps (61–64). Les VoC ont en commun de présenter au moins une mutation non synonyme dans le motif de liaison au récepteur (Receptor Binding Motif, RBM). Le RBM est le sous-domaine du RBD (Receptor Binding Domain) contenant la plupart des résidus entrant en contact avec le récepteur cellulaire du virus, la protéine humaine ACE2 (hACE2), et se caractérise également par la présence d'épitopes cibler par les anticorps neutralisants (**Figure 34 A et B**). Certaines des mutations caractérisées sont N501Y pour 501Y.V1, E484K/N501Y pour 501Y.V2 et 501Y.V3. Ces mutations peuvent affecter la liaison de la protéine S à l'hACE2 et/ou l'efficacité des anticorps neutralisants (**Figure 34 B**) (63–67). L'évolution génétique de cette région fait l'objet d'un examen minutieux afin d'identifier d'éventuels nouveaux VoC. La détection rapide des VoCs est donc essentielle pour atténuer leur transmission et pour ajuster les thérapies afin d'éviter d'en réduire l'efficacité.



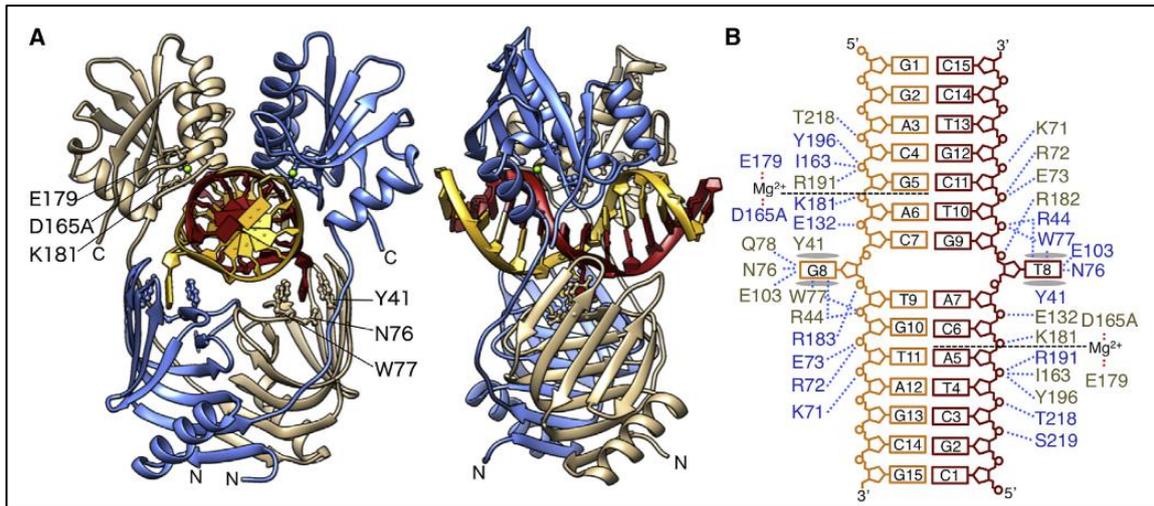
### Figure 34 : Le RBM : région clef pour l'entrée du virus et la séroneutralisation

**A)** Structure primaire de la protéine Spike du SARS-CoV-2. NTD: domaine N-terminal ; RBD: domaine de liaison au récepteur ; RBM: motif de liaison au récepteur ; SD1: sous-domaine 1 ; SD2: sous-domaine 2 ; FP: peptide de fusion ; HR1: répétition Heptad 1 ; HR2: répétition Heptad 2 ; TM: région transmembranaire ; CT: région cytoplasmique. Les positions des acides aminés du cadre de lecture de RBD et RBM sont écrites en noir et en bleu, respectivement. Les mutations fonctionnelles dans le RBM sont indiquées en jaune (E484K) et en rouge (N501Y). Les flèches violettes et les numéros correspondent à la région amplifiée par RT-PCR. **B)** Représentation de la structure tridimensionnelle montrant l'interaction entre le RBD du SARS-CoV-2 (vert) et le récepteur hACE2 (blanc) (PDB 6vw1). Le RBM est représenté en bleu. Les acides aminés E484 et N501 sont mises en évidence en orange et en rouge, respectivement. La figure est extraite de l'article (59)

Les deux méthodes les plus couramment utilisées pour la détection des VoC et la surveillance de l'évolution de la population du SARS-CoV-2 sont la RT-PCR en temps réel pour la recherche de mutations à des positions données et les campagnes massives de séquençage de nouvelle génération, qui contribuent toutes deux à la prise de décision en matière de santé publique (68). D'une part, la RT-PCR en temps réel est rapide et opérationnelle sur place, mais elle ne peut détecter que les mutations connues et ne permet pas de détecter de nouvelles mutations. D'autre part, le NGS peut détecter n'importe quelle mutation le long du génome, mais les délais d'obtention des résultats peuvent prendre plusieurs jours, ce qui retarde la prise de décision médicale à prendre dès l'identification de la séquence. De plus, tous les échantillons biologiques ne peuvent pas être séquençés et un échantillonnage en amont est obligatoire pour sélectionner les échantillons biologiques les plus pertinents à caractériser.

### 3.1.2.2. Les endonucléases comme outil de génotypage.

Depuis la fin 2021, l'endonucléase Surveyor® n'est plus commercialisée, et il n'existe à ce jour, selon notre connaissance, aucun kit similaire sur le marché. Dans la nature, il existe plusieurs endonucléases mis-match spécifiques (MM), qui participent principalement au maintien de l'intégrité génomique. On peut citer MutS, MutL et leurs homologues, qui sont des acteurs clés de la reconnaissance des mésappariements de bases et du déplacement du site mésapparenté par la synthèse d'un nouveau brin d'ADN chez les bactéries et les eucaryotes. Pour les archées, qui ne présentent pas d'homologue de MutS, les fréquences de mutations sont comparables ou inférieures à celles d'autres micro-organismes bactériens, ce qui suggère que les Archées maintiennent un niveau élevé de stabilité génomique grâce à d'autres protéines impliquées dans la réparation des mésappariements (69). De ce constat, Ishino *et al.* (70) ont identifié une nouvelle endonucléase appelée Endonucléase MS (EndoMS) coupant l'ADN double brin contenant une paire de bases mal appariées, provenant de l'archaebactérie *Pyrococcus furiosus*. La similarité de séquence suggère qu'EndoMS est un orthologue de la nucléase S (NucS) de *Thermococcus kodakarensis*. La structure de cette endonucléase a été résolue par cristallographie à rayon X (71) (**Figure 35A et B**), permettant l'identification du domaine de fixation de la protéine à l'ADN et de la région reconnaissant un MM. Cette endonucléase clive les mésappariements G/T, G/G, T/T, T/C et A/G, avec une préférence pour G/T, G/G et T/T, et n'a que peu ou pas d'effet sur les mésappariements C/C, A/C et A/A.



**Figure 35 : Structure de l'EndoMS issue de *Thermococcus kodakarensis***

**A)** Modèle du complexe TkoEndoMS-ADN double brin en vue frontale (à gauche) et latérale (à droite). Les résidus clés pour l'activité catalytique (D165A, E179, et K181) et la liaison des bases de mésappariement (Y41, N76, et W77) sont représentés sous forme de modèles en rubans. **B)** Schéma des interactions avec l'ADN double brin de TkoEndoMS. Les liaisons hydrogènes sont représentées par des lignes en pointillés bleus. Les lignes en pointillés noirs indiquent les positions de clivage. Les interactions avec les ions métalliques sont représentées par des lignes pointillées rouges. Celles qui passent par D165A sont hypothétiques et ne sont pas observées dans les structures cristallines parce que D165 a été muté. Figure extraite de l'article (71).

### 3.2. Enjeux du projet

L'objectif de cette deuxième partie de mon travail fut de faire la preuve de concept qu'une méthode alternative ou complémentaire au séquençage permettrait la surveillance de la dérive génétique du SARS-CoV-2 dans la région RBM où les mutations sont susceptibles d'affecter la transmission, la pathogénicité ou la résistance aux anticorps du virus. La technique, qui repose sur l'amplification de la séquence codante du RBM, suivie d'un test utilisant une endonucléase spécifique des mésappariements (Endonucléase commerciale, Surveyor® Nuclease S), devra être ensuite validée sur des échantillons biologiques pour démontrer sa faisabilité.

Comme mentionné précédemment, l'endonucléase Surveyor® Nuclease S n'est plus commercialisée. En s'appuyant sur l'étude de Ishino *et al.* (70) « *Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea* », l'objectif est de produire Endonucléase MS (EndoMS) et de démontrer son utilité dans un test de génotypage. L'enzyme sauvage étant spécifique de certains mésappariements, nous essaierons ensuite de réduire la spécificité de cette endonucléase en testant différents mutants d'EndoMS sélectionnés par une étude structurale pour améliorer les performances de l'enzyme.

### **3.3. Matériel et méthodes**

#### **3.3.1. Génotypage SARS-CoV-2**

Un test de génotypage des variants SARS-CoV-2 a été réalisé sur la séquence codante du RBM. Le développement du test a consisté à valider le système d'amplification et le système de détection avec une nucléase spécifique des mésappariements (cette partie n'est pas présentée dans ce mémoire car elle a été faite par une tierce personne). Afin de déterminer la linéarité du système, une RT-PCR en temps réel a été réalisée avec le kit QuantiTect SYBR Green RT-PCR Kit (Qiagen), sur un thermocycleur BioRad CFX96™. Le cycle thermique était le suivant : 50 °C pendant 30 min ; 95 °C pendant 15 min. ; 40 cycles de 95 °C pendant 15 sec. ; 50 °C pendant 30 sec. et 72 °C pendant 45 sec. Nous avons évalué la méthode sur un échantillon nasopharyngé d'une personne testée positive au COVID-19 par RT-qPCR. L'acide nucléique total a été extrait de 200 µL d'échantillons à l'aide de l'appareil EZ1 (Qiagen) avec le kit « virus mini kit v2.0 », et élué dans 60µl.

La RT-PCR a été réalisée avec le kit SuperScript II One-Step Platinum (ThermoFisher Scientific), sur un thermocycleur Biometra T3000. Les amorces ont été synthétisées et fournies par ThermoFisher Scientific. Le cycle thermique a été défini comme suit : 50 °C pendant 30 min. ; 94 °C pendant 2 min. ; 40 cycles de 94 °C pendant 15 sec. ; 55 °C pendant 20 sec et 68 °C pendant 20 sec. L'élongation finale a été réalisée à 72°C pendant 2 min.

La présence d'une mutation sur la cible amplifiée a été détectée à l'aide du kit de détection de mutation Surveyor® Mutation Detection Kit (Integrated DNA Technologies). Deux produits PCR (un produit de référence et un échantillon à analyser) ont été mélangés dans un volume final de 10µl. Un cycle de dénaturation lente et de ré-hybridation a été effectué selon les recommandations du fabricant (protocole en [annexe 11](#)). Puis, 1 µl de Surveyor® Enhancer S, 1 µl de Surveyor® Nuclease S et 1,2 µl de solution Surveyor® MgCl<sub>2</sub> ont été ajoutés au mélange et incubés pendant 1 heure à 42°C suivie de 10 min à 4°C selon les recommandations du fabricant. Le mélange a ensuite été chargé sur un appareil d'électrophorèse capillaire « Bioanalyser Agilent 2100 » avec le kit « Agilent high sensitivity DNA kit » selon le protocole du kit.

### 3.3.2. Production, caractérisation et optimisation des endonucléases EndoMS

#### 3.3.2.1. Production des endonucléases

Les séquences des inserts permettant la production des différentes endonucléases sont présentées en [Annexe 12](#). La séquence de l'endonucléase WT est celle de *Thermococcus Kodakarensis* KOD1 (NCBI Reference Sequence NC\_006624.1), et les mutants sont construits à partir de cette séquence. Les plasmides servant de matrice pour la production de ces endonucléases sont synthétisés par la société Thermo Fischer Scientific, dans des plasmides de type PMX, avec un gène de résistance antibiotique à la Kanamycine. Les plasmides sont resuspendus dans 50µl d'eau ultra pure pour obtenir une concentration finale à 0.1µg/µl.

Des bactéries *E.coli* T7 Expression Iq sont transformées avec 1µl de plasmide et sont étalées sur un mélange d'agarose, de milieu de culture Luria Bertani (LB) et Kanamycine (0,05 mg/ml), puis incubées toute une nuit à 37°C. Les précultures sont faites dans 25 ml de milieu LB + Kanamycine (0,05 mg/ml) et sont incubées toute une nuit à 37°C à 200 RPM. Les cultures sont faites dans 1L de milieu Terrific Broth (TB) (47,6 mg/ml), 8 ml de glycérol et Kanamycine (0,05 mg/ml). Les cultures sont incubées à 37°C, 200 RPM, la densité optique (DO<sub>600nm</sub>) est mesurée. L'induction de la protéine se fait avec 0,5mmol/L d'IPTG, puis est incubée toute une nuit à 17°C à 200 RPM. Les cultures sont arrêtées par centrifugation pendant 10 minutes à 4000g. Les bactéries sont ensuite lysées avec 50ml de tampon de lyse (50 mmol/l TRIS base, 300mmol/l NaCl, 10 mmol/L imidazole, 5% glycérol, 0,1% Triton, 10µg/ml DNase, 0,25 mg/ml lysozyme, 1 mmol/l PMSF, 1x BugBuster pH8) et 200µl d'inhibiteur de protéase P8849 (SIGMA-ALDRICH) et incubées pendant 30 minutes avec agitation. Le milieu est ensuite centrifugé pendant 30 minutes à 12000 RPM. Le surnageant est collecté, et incubé pendant 30 minutes avec ajout de 4ml de billes de Nickel.

La purification se fait par affinité sur colonne. Trois lavages sont faits, le premier avec 15ml de tampon « A » (50mmol/l TRIS base, 300 mmol/l NaCl, 10 mmol/l imidazole), le deuxième avec 15ml de tampon « A » + 1M NaCl et le troisième avec 14.25 ml de tampon « A » et 0,75ml de tampon « B » (50 mmol/l TRIS base, 300 mmol/l NaCl, 500 mmol/l imidazole). L'élution se fait avec 15ml d'un mélange en quantité égale de tampon « A » et « B », récolté en trois fractions de 5ml. Les protéines obtenues sont dosées par spectrophotométrie à 280nm. Les fractions contenant une quantité raisonnable de protéine (minimum 1mg/ml de protéine) sont concentrées dans un concentrateur de protéine « Merck Millipore Ensembles de filtration par

centrifugation Amicon™ Ultra 10K Da » jusqu'à obtenir un volume final de 5 ml. Le produit final est dosé par spectrophotométrie à 280nm.

Les protéines sont ensuite dialysées pour éliminer l'excédent d'imidazole dans des membranes de dialyses « ZelluTrans/ROTH T2 : Taille de pore 6000-8000, 32 mm », dans 2l de tampon de dialyse (50mM TRIS pH8, 0.5mM DTT, 0.5M NaCl) et sous agitation à 4°C toute la nuit. Les protéines dialysées sont dosées encore une fois par spectrophotométrie à 280nm.

#### 3.3.2.2. Tests d'activité

Pour réaliser les tests de digestion, nous avons commandé des amorces simples de 45 nucléotides, de la même séquence que dans l'article (70). Les séquences de ces amorces sont détaillées en [Annexe 13](#). Deux amorces « Reverse » sont taguées Cyanine 5 en 5' pour pouvoir les analyser par fluorescence. Les amorces sont fournies par EUROGENTEC.

Les amorces sont diluées pour avoir une concentration finale de 16,7 nM. 10µl d'amorce sens et 8µl d'amorce complémentaire taguée Cy-5 sont mélangées et hybridées selon le même cycle thermique que pour le génotypage SARS-CoV-2. L'endonucléase est ajoutée de manière à avoir une concentration finale d'enzyme à 10nM, ainsi qu'1µl de tampon de digestion (20mM TrisHCL pH8, 2mM MgCl<sub>2</sub>, 100mM NaCl, 0.1% Triton, 0.1% BSA). Le mélange est incubé pendant 15 min à 55°C. La réaction est stoppée avec 1µl d'EDTA à 500mM ainsi qu'1µl de protéinase K (Qiagen), puis incubée 30min à 50°C. 4µl de solution de bleu de bromophénol est ajoutée à chaque échantillon. L'analyse se fait sur gel 20% acrylamide / 7M Urée suivi d'une migration sous tension 300V pendant environ 1h30. La révélation se fait par fluorescence.

### 3.4. Résultats et discussion

#### 3.4.1. Génotypage SARS-CoV-2

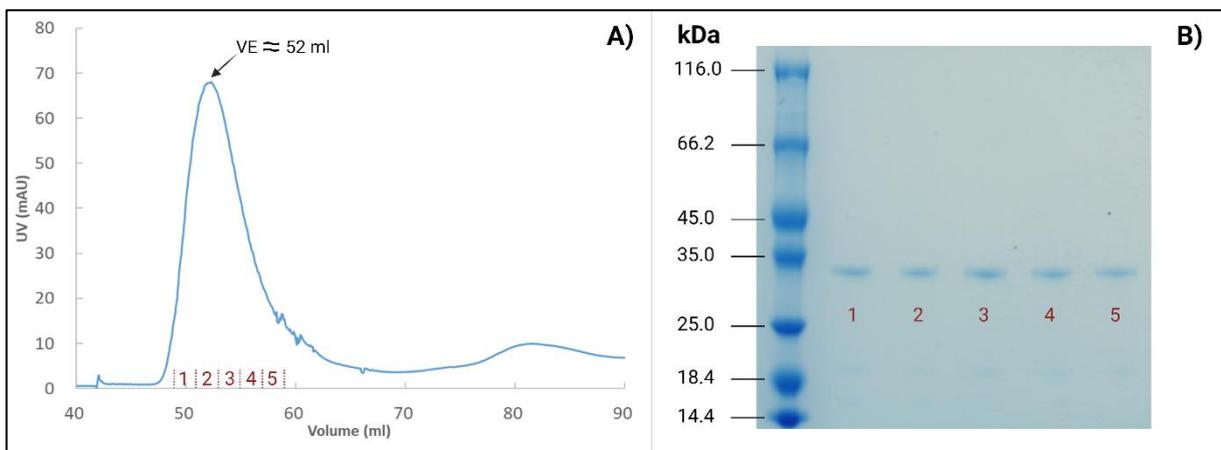
Ces résultats ont fait l'objet d'un article, disponible en [Annexe 14](#). Pour ce projet, ma contribution a consisté dans un premier à définir la sensibilité du test de génotypage par qPCR SYBR Green, afin d'appliquer un seuil minimal de copies/ $\mu$ l d'ARN utilisable pour ce test. Dans un second temps, j'ai contribué à la validation de cette technique en l'appliquant à des tests à grande échelle.

#### 3.4.2. Production d'une endonucléase recombinante.

##### 3.4.2.1. Résultats préliminaires.

La production dans *E. coli*, la purification et la caractérisation fonctionnelle de l'EndoMS ont dans un premier temps été réalisées sur l'EndoMS WT de *Thermococcus kodakarensis*

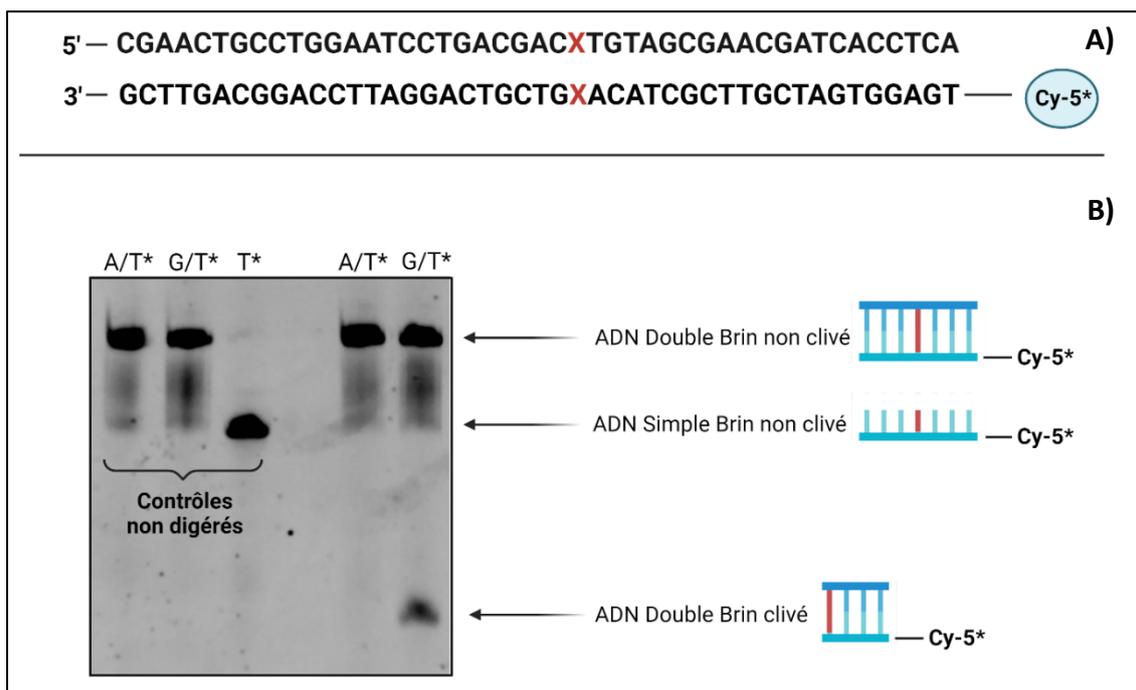
La production de l'EndoMS a été réalisée en culture bactérienne. Le lysat bactérien a été purifié par SEC afin de contrôler la pureté du mélange, et le cas échéant de ne collecter que les fractions correspondant à la protéine d'intérêt (**Figure 36 A**). Le volume d'élution de la protéine (pic indiqué par une flèche) est d'environ 52 ml, correspondant -selon la courbe de calibration de la colonne ([Annexe 15](#))- à une protéine de plus de 150 kDa, nous indiquant une possible tétramérisation de la protéine. Dans la littérature, l'EndoMS s'organise en dimère (71). Les fractions récoltées correspondant à la protéine d'intérêt ont été conservées et analysées par SDS-PAGE (**Figure 36 B**). Les bandes sur le gel permettent de confirmer que l'enzyme qui va être utilisée pour la suite des tests est pure (une seule bande sur le gel) et est au poids moléculaire attendu (30kDa).



**Figure 36 : Chromatographie d'exclusion de taille d'EndoMS.**

**A) :** Électrophorégramme obtenu après analyse par gel filtration de l'EndoMS WT. **VE :** Volume d'élution de la protéine. **B) :** SDS-PAGE coloré au bleu de Coomassie des fractions récoltées en SEC.

Afin d'identifier la spécificité de reconnaissance de l'EndoMS, nous avons mis en place un test dans lequel deux ADN (45pb) simples brins complémentaires entre eux à l'exception d'un nucléotide, sont hybridés, digérés par EndoMS puis séparés sur gel urée/acrylamide (**Figure 37 A**). L'EndoMS étant capable de reconnaître et de cliver des MM de type G/T (70), les premiers tests de digestion ont été réalisés sur des mésappariements de type G/T afin de tester la fonctionnalité de l'endonucléase synthétisée (**Figure 37 B**). Un ADN double brin ne présentant pas de MM (A/T\*) est utilisé comme contrôle négatif, et un ADN simple brin marqué Cy-5 est utilisé comme témoin d'hybridation. Un ADN MM G/T non digéré est utilisé comme contrôle négatif de digestion. Après révélation par fluorescence, nous observons une bande correspondant à un ADN simple brin clivé pour l'ADN présentant un MM : nous pouvons conclure de ce test que l'EndoMS synthétisée est fonctionnelle pour la reconnaissance et le clivage d'un MM G/T. Des tests de clivage de l'EndoMS WT avec d'autres mutations n'ont pas été réalisés en première intention, car elle est décrite comme non fonctionnelle dans la littérature (70).

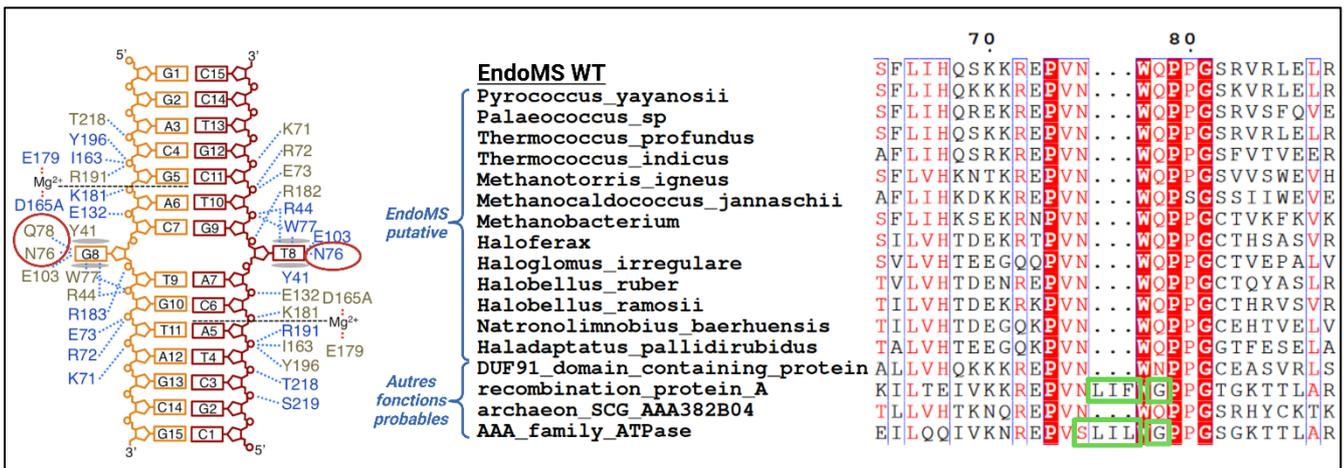


**Figure 37 : Fonctionnalité d'EndoMS**

**A)** Séquences nucléotidiques des ADNs servant de matrice pour les tests de digestion avec EndoMS. Les nucléotides notés « X » peuvent être A, T, C ou G, et correspondent donc à la zone où un mis-match est intégré. **B)** Analyse sur gel acrylamide/urée des produits de digestion d'un mismatch G/T. Les trois premiers points constituent les témoins négatifs de l'expérience, car ils n'ont pas été digérés par l'enzyme. A/T\* est un contrôle négatif de digestion. Les nucléotides notés « \* » correspondent aux brins fluorescents marqué avec une Cy-5. Les bandes sur le gel sont révélées par fluorescence.

### 3.4.2.2. Tentatives d'optimisation

En s'appuyant sur la structure d'EndoMS, et sur un alignement de séquences protéiques homologues chez d'autres procaryotes (**Figure 38**), nous avons défini quatre mutants d'EndoMS pouvant être de bon candidats pour réduire cette spécificité de clivage (**Figure 38**, panel gauche). Les acides aminés N76 et Q78 sont impliqués dans la reconnaissance des MM, et présentent des mutations chez des endonucléases homologues (zones encadrées en vert (**Figure 38**, alignement à droite). Une insertion de 3 acides aminés « LIL ou LIF » est également observée chez deux espèces.

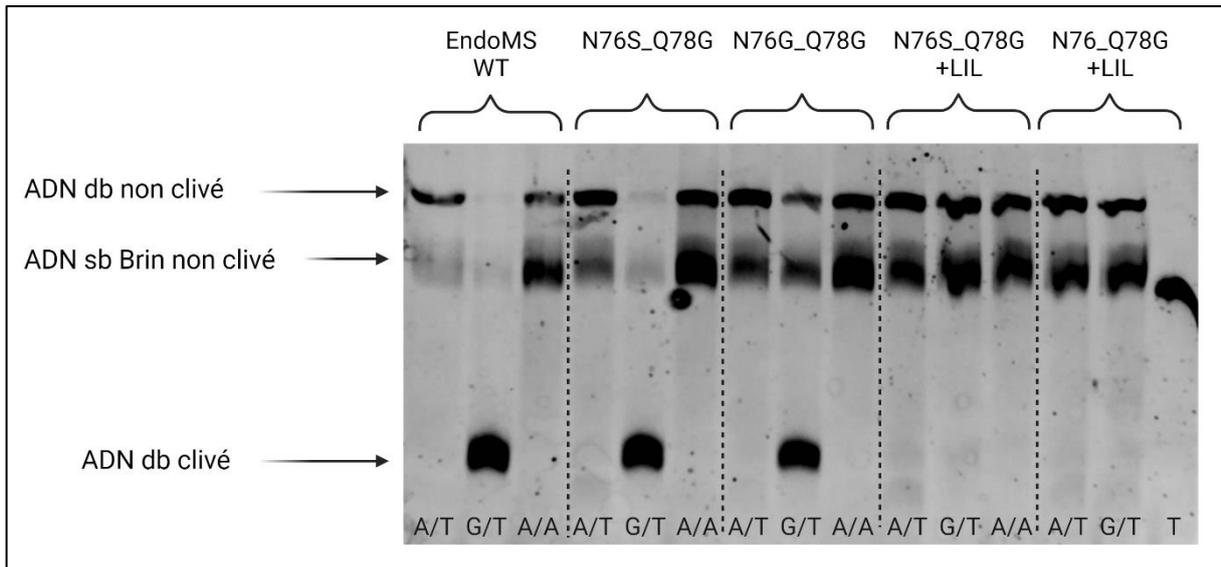


**Figure 38 : Identification des résidus impliqués dans l'interaction d'EndoMS avec les mésappariements**

Gauche : résidus impliqués dans l'interaction EndoMS-ADN. Les acides aminés entourés en rouge correspondent aux zones où des mutations sont observées. Droite : alignement multiple de séquences homologues au domaine de reconnaissance des mésappariements d'EndoMS. Les cases rouges représentent une identité stricte, et les caractères en rouge représentent les groupes présentant une similarité. Les zones encadrées correspondent aux mutations d'intérêt. Alignement réalisé avec Esript 3.0 (72).

De ce constat, quatre mutants d'EndoMS ont été générés, correspondant à des combinaisons de mutations : N76S\_Q78G ; N76G\_Q78G ; N76S\_Q78G insertion **LIL** ; N76\_Q78G insertion **LIF**. Ces mutants ont été synthétisés selon le même protocole que l'EndoMS WT, et ont été testés sur un mésappariement de type G/T, reconnu par l'endonucléase WT, pour contrôler que les mutations n'induisent pas une perte de reconnaissance de l'endonucléase. Ils ont également été testés sur un MM A/A, non reconnu par EndoMS (70), afin de savoir si les mutations ainsi générées permettent de réduire la spécificité de la reconnaissance à un MM (**Figure 39**). La production de ces enzymes ainsi que les tests de digestion ont été

réalisés de la même manière que pour l'EndoMS WT. Les rendements de production (données non présentées) de ces mutants sont identiques à ceux obtenus pour l'EndoMS WT.



**Figure 39: Influence des mutations sur l'efficacité et la spécificité de clivage d'EndoMS**

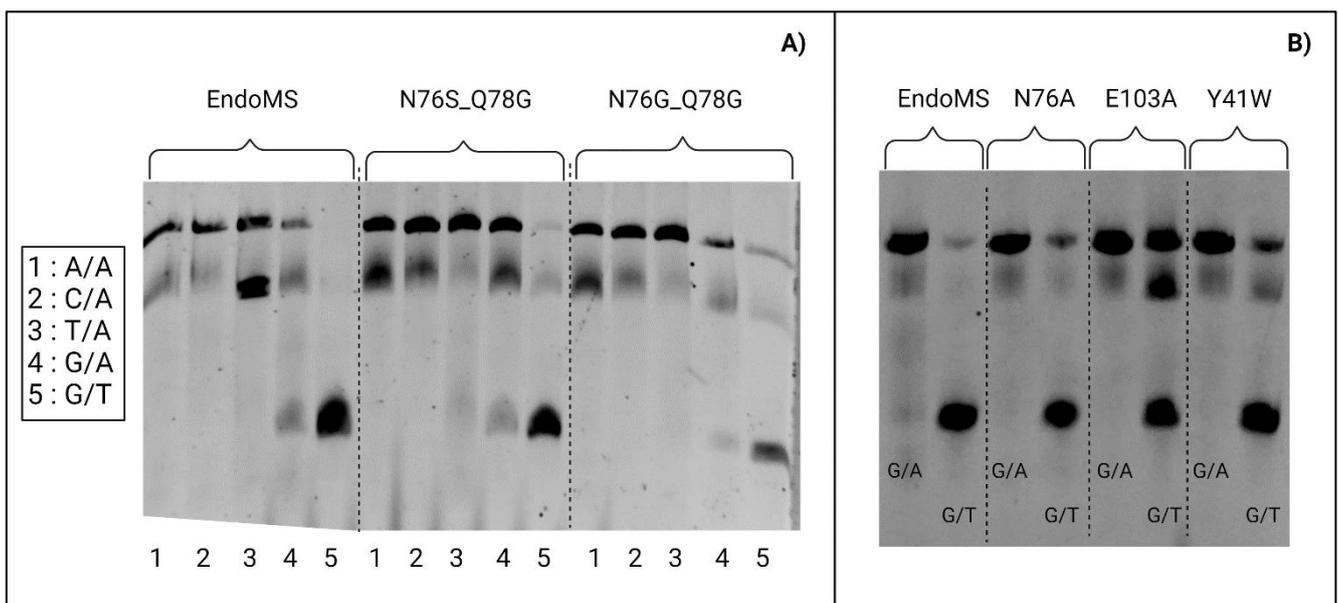
Analyse sur gel d'acrylamide/urée des produits de digestion des brins d'ADN digérés avec les différents mutants d'EndoMS. La révélation du gel se fait par fluorescence. db : double brin ; sb : simple brin

Ce test permet de conclure sur différents points. Dans un premier temps, une insertion, LIL ou LIF, en position 77 juste après l'asparagine (N), couplée à deux mutations (N76S\_Q78G ou N76G\_Q78G) induit une perte d'efficacité totale de l'endonucléase. Ce résultat permet la mise en évidence de l'importance de cette région dans la capacité de reconnaissance et/ou de clivage des mésappariements. Dans un second temps, la modification d'une Asparagine en Sérine en position 76, et d'une Glutamine en Glycine en position 78 (N76S\_Q78G), n'induit aucun changement d'efficacité de l'endonucléase. Cependant, le substrat n'est pas totalement consommé lors d'une digestion avec le mutant N76G\_Q78G : une modification de l'Asparagine en Glycine induit donc une légère perte d'efficacité sur un mésappariement G/T. Pour terminer, aucun de ces mutants n'est capable de reconnaître et/ou de cliver des mésappariements A/A.

Les mutants contenant une insertion LIL ou LIF sont retirés de l'étude, du fait de leur perte fonctionnelle sur le mésappariement G/T. Les deux autres mutants, ayant conservé leur activité sur G/T, ont été testés sur d'autres types de mésappariement : G/A et C/A (**Figure 40 A**). Selon l'article de Ishino et al.(70), les mésappariement G/A sont reconnus et clivés par l'EndoMS, à l'instar des C/A. En pratique, l'EndoMS que nous avons produit est moins efficace qu'attendu : pour un mismatch G/A, le substrat n'est pas totalement digéré par l'enzyme dans les conditions de notre test. Il en va de même pour les deux mutants,

avec toujours une efficacité moindre pour le mutant N76G\_Q78G (**Figure 40 A**).

L'étude fonctionnelle a mis en évidence que les mutations testées, non seulement ne permettent pas d'élargir le spectre de mésappariement de l'EndoMS, mais induisent pour certaines une perte fonctionnelle. Après un second tour d'analyse structurale, trois nouveaux mutants ont été testés: les mutants N76A, E103A et Y41W, ces trois positions étant impliquées dans la fixation du mésappariement, et la potentielle spécificité de reconnaissance à la base G ou T. Comme présenté sur la **Figure 40 B**, les trois mutants présentent une efficacité de reconnaissance et de clivage des mésappariement G/T. Cependant, nous observons une bande de forte intensité, correspondant à de l'ADN double brin non clivé, synonyme d'un clivage partiel. Cette bande est quasiment absente lors d'un clivage avec l'EndoMS WT : les mutants N76A, E103A et Y41W présentent une efficacité de clivage inférieure à l'EndoMS WT. De plus, nous n'observons pas de bande correspondant à un ADN clivé pour les MM de type G/A : aucunes de ces mutations ne confère à EndoMS une reconnaissance et un clivage



En conclusion, nous avons démontré qu'il est possible de produire cette endonucléase spécifique des mis-match de manière recombinante. Cette endonucléase est capable de reconnaître et de cliver des mésappariements de type G, avec un clivage total des mésappariement G/T et très partiel des G/A. Les différents mutants testés à ce jour ne permettent pas de réduire la spécificité de reconnaissance d'EndoMS. Cependant l'utilisation de l'EndoMS WT sera prochainement évalué sur un test de géotypage en comparaison avec nos reliquats du kit Surveyor® Nuclease S ; Par ailleurs, d'autres endonucléases sont actuellement en cours de test : les endonucléases SPI et CeII, extraient respectivement de l'épinard et du céleri, sont des composantes de l'endonucléase commerciale Surveyor® Nuclease S, et peuvent être de bon candidat pour une reconnaissance de mésappariement non spécifique en vue d'une application.

#### **4. Conclusion générale et perspectives**

Les tests de détection moléculaire, et notamment ceux dédiés aux virus pathogènes pour l'homme, nécessitent l'utilisation de contrôles et de matériels de référence qui garantiront la pertinence des résultats lors du développement ou de l'utilisation de ces tests.

Dans un premier temps la plateforme de diagnostic moléculaire a nécessité la rationalisation du développement, de la validation et de la production des contrôles positifs, IVT ou ARN encapsidés pour les virus à génome ARN, plasmides ou amplicons pour les virus à génome ADN. Pour cela, nous avons cherché à évaluer l'homogénéité des contrôles positifs IVT déjà disponibles au sein du laboratoire. Nous avons identifié des problèmes d'hétérogénéité de taille dans les contrôles de type transcrits *in vitro*, imposant des changements de protocoles de production. Ces travaux nous ont permis d'obtenir des IVT homogènes, permettant la mise en place de tests de processus de LoD (Limit of Detection, limite de détection) fiables par dilutions en série des IVT quantifiés. En parallèle, nous avons établi un protocole de production de témoins positifs encapsidés « Armored RNA » qui, en mimant une particule virale, permettent de contrôler toutes les étapes d'un processus diagnostique et permettent de s'affranchir des problèmes de stabilité de l'ARN. Les rendements de ce procédé – jusqu'à 2 mg/L de culture - permettent la production de centaines de milliers de tests positifs stables à température ambiante après lyophilisation. L'ensemble du processus de production et de distribution permet au laboratoire de les déployer à travers le monde. Certains de ces matériels sont disponibles sur site internet partenaire du laboratoire, EVA Global (<https://www.european-virus-archive.com/>) pour la communauté scientifique et les institutions de santé publique. Il s'agit notamment d'ArRNA cibles de virus à ARN, mais aussi MS2. En effet, dans le souci de maîtriser les technologies, nous avons substitué le phage MS2, utilisé comme contrôle interne lors de l'extraction virale, par un ArRNA MS2. Dans la majorité des cas, ces tests sont à destination des pays du Sud, par exemple sous forme d'EQA (External Quality Assesment, évaluation externes de qualité). Ces EQA sont des matériels de référence permettant de comparer les tests d'un laboratoire à une source extérieure, et permettent d'évaluer la performance diagnostic notamment des pays africains. Dans le cadre du projet international FIND (<https://www.finddx.org/>), nous avons développé, produit et envoyé plus de 30000 EQA à l'Institut Pasteur du Sénégal. Dans le cadre de la pandémie de SARS-CoV-2, différents ArRNA ciblant la région NTD+RBD du virus (régions du génome présentant les mutations les plus surveillées) ont été développés et mis à disposition de la communauté scientifique. Depuis le développement de ces ArRNA au laboratoire et dans un souci d'harmonisation des pratiques,

chaque construction plasmidique générée aujourd'hui est compatible pour la production de témoins sous forme IVT et ArRNA pour répondre à tous les besoins, sans modification de design préalable. A ce jour, les ArRNA sont utilisables uniquement dans un format RUO (Research Use Only), mais le processus de production devrait faire l'objet d'une certification ISO13485 afin d'être intégrés à des tests de diagnostic moléculaire.

La prochaine étape pour le développement de contrôle ARN est de mettre en place un procédé permettant l'encapsidation de séquences d'ARN double brin. En effet, la méthode de production d'ArRNA permet de générer des contrôles ARN simple brin uniquement. Nous ne disposons donc à ce jour d'aucun contrôle encapsidé pour la détection de virus à génome ARN double brin (groupe III de la classification de Baltimore), comme les Rotavirus, responsable de gastro-entérite chez l'enfant, entraînant le décès de centaines de milliers d'enfants de moins de 5 ans chaque année en Afrique (73). Afin d'utiliser la machinerie du phage MS2 pour faire un ArRNA double brin, l'insert plasmidique à encapsider contiendra la séquence cible du Rotavirus, ainsi que sa séquence inverse-complémentée, reliée par un « linker ». Cette technique permettrait de générer un ArRNA simple brin, qui se replierait naturellement grâce aux liaisons des nucléotides complémentaires afin de former un ArRNA double brin. La production de cet ArRNA Rotavirus est actuellement en cours de développement, même si nous nous attendons à des difficultés lors de la synthèse des gènes.

Afin d'étendre le panel des contrôles encapsidés au virus à génome ADN, nous avons aussi cherché à développer la production de contrôles encaspidés pour les ADN, afin de disposer de témoins d'extraction pour les virus à génome ADN. La preuve de concept de la faisabilité de cette technique a été réalisée avec une séquence cible du Virus de l'Hépatite B, et pourrait être transposable à n'importe quel type de virus à génome ADN simple brin. Nous avons mis en évidence l'encapsidation de l'ADN recombinant, mais notre protocole actuel n'est pas encore totalement mature pour les phases de production. En effet, les rendements de production actuels ne nous permettent pas le déploiement de ces tests à grande échelle. De plus, le processus de production d'ArDNA *via* l'utilisation du phage M13 ne permet d'encapsider que de l'ADN simple brin, et ne permet donc pas une mimer à l'identique une région du génome de virus à ADN double brin. Pour y remédier, nous avons essayé d'utiliser de capsides recombinantes de virophages, mais le processus relativement complexe d'encapsidation de ces modèles, pourtant parmi les plus simples chez les virus ADN, n'est pas encore maîtrisé.

Alors que les technologies de séquençage sont déployées dans le monde entier notamment pour l'identification des agents pathogènes, des besoins au lit du patient demeurent. Comme cela a

été montré pour les contrôles positifs avec les systèmes MS, les outils de biotechnologie peuvent pourvoir à ces besoins. Des enzymes participant à l'intégrité génomique des organismes peuvent en effet être utilisées pour l'identification de mutations. Dans ce cadre, nous avons développé un projet permettant la mise en évidence de variants du SARS-CoV-2 sur la région RBM du génome. Cette technique permet l'identification d'un variant déjà connu, mais également de détecter n'importe quelle autre variation dans cette région génomique du virus sans faire appel au séquençage en première ligne. Malheureusement, l'endonucléase de restriction nous ayant permis de réaliser ce projet, la Surveyor® Nuclease S, n'est plus commercialisée depuis juin 2022. Cette technique d'identification donnant des résultats intéressants, nous avons produit une endonucléase d'archaebactérie sous forme recombinante afin de pouvoir continuer à utiliser cette méthode de détection. Nous avons confirmé que cette enzyme était spécifique des mésappariements impliquant un G ou un T. Par mutagenèse, nous avons essayé d'éliminer cette spécificité afin que l'EndoMS puisse détecter n'importe quel type de mésappariement. Bien que non fructueuse, cette approche nous a cependant conforté dans l'utilisation de l'EndoMS WT dans la mesure où l'appariement d'un ADN de référence et d'un ADN échantillon conduit à la production d'un hétéroduplex impliquant systématiquement au moins un G ou un T. Cette technologie n'a probablement pas d'avenir pour le SARS-CoV-2, dans la mesure où la variabilité génétique sur le RBM est devenue importante du fait de la dimension de la pandémie, mais nous envisageons de la repositionner sur du génotypage d'arbovirus.

.

.

## Références

1. Lecoq H. 2001. Découverte du premier virus, le virus de la mosaïque du tabac : 1892 ou 1898 ? Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie 324:929–933.
2. Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. 2012. Human viruses: discovery and emergence. *Philos Trans R Soc Lond B Biol Sci* 367:2864–2871.
3. La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, Merchat M, Suzan-Monti M, Forterre P, Koonin E, Raoult D. 2008. The virophage as a unique parasite of the giant mimivirus. *Nature* 455:100–104.
4. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res* 46:D708–D717.
5. Baltimore classification ~ ViralZone. <https://viralzone.expasy.org/254>. Retrieved 14 September 2022.
6. Baltimore D. 1971. Expression of animal virus genomes. *Bacteriol Rev* 35:235–241.
7. Morse SS. 1996. *Emerging Viruses*. Oxford University Press.
8. Tomasello D, Schlagenhauf P. 2013. Chikungunya and dengue autochthonous cases in Europe, 2007–2012. *Travel Medicine and Infectious Disease* 11:274–284.
9. Cochet A, Calba C, Jourdain F, Grard G, Durand GA, Guinard A, Noël H, Paty M-C, Franke F, Auzet-Caillaud M, Aventini C, Borel A, Bruel C, Decoppet A, Gossa M, Krouk M, Laboureyras C, Mazza A, Mihoubi M, Muriel A, Peloux-Petiot F, Piétin C, Raibaut J,

Saintillan L, Schmitt C, Travanut M, Vassal C, Albert P, Bulmanski-Then L, Cahuzac C, Catala L, Cot A, Creton P-M, Dubois A, Dehecq JS, Estève-Moussion I, Gaillard F, Giraud C, Glass O, Lagarde V, Larrose A, Lecerf C, Ould Larabi R, Maes B, Peiffer G, Rico C, Rouvié-Laurie I, Roux N, Santana G, Sauthier N, Sellami S, Wagner S, Zumbo, B, Balle E, Coiplet L, Gaucher S, Leccia, G, Morel, D, Jouanthoua, F, Belkadi L, Broustal O, Durand C, Mouly D, Yubero G, Riondel A, Bekheira L, Kelly D, Mertens-Rondelart I, Sanchez-Ruiz M-A, Souares Y, Yanwou, N, Barbry A, Durand T, Ovize A, Soares, A, Roquebert B, Verdurme, L, Mansuy J-M, Aumaître, H, Chanaud L, L'Ambert G, Kervella, Y-M, Lacour G, Mignotte A, Tizon, C, Biancarelli A, Maestracci J-L, Pompa B, Santoni J-B. 2022. Autochthonous dengue in mainland France, 2022: geographical extension and incidence increase. *Euro Surveill* 27:2200818.

10. 2021. *Aedes albopictus* - current known distribution: March 2021. <https://www.ecdc.europa.eu/en/publications-data/aedes-albopictus-current-known-distribution-march-2021>. Retrieved 19 April 2023.
11. Muralidar S, Ambi SV, Sekaran S, Krishnan UM. 2020. The emergence of COVID-19 as a global pandemic: Understanding the epidemiology, immune response and potential therapeutic targets of SARS-CoV-2. *Biochimie* 179:85–100.
12. Spencer RC. 2003. *Bacillus anthracis*. *Journal of Clinical Pathology* 56:182–187.
13. Centers for Disease Control and Prevention (CDC). 2001. Update: Investigation of bioterrorism-related anthrax and interim guidelines for exposure management and antimicrobial therapy, October 2001. *MMWR Morb Mortal Wkly Rep* 50:909–919.
14. Tabatabaeizadeh S-A. 2021. Airborne transmission of COVID-19 and the role of face mask to prevent it: a systematic review and meta-analysis. *Eur J Med Res* 26:1.

15. WHO-2019-nCov-IPC\_Masks-2020.5-fre.pdf.
16. DICOM\_Jocelyne.M, DICOM\_Jocelyne.M. 2023. Comment se protéger de la dengue ? Ministère de la Santé et de la Prévention. <https://sante.gouv.fr/soins-et-maladies/maladies/maladies-infectieuses/article/comment-se-protoger-de-la-dengue>. Retrieved 22 March 2023.
17. Rougeole. <https://www.who.int/fr/news-room/fact-sheets/detail/measles>. Retrieved 22 March 2023.
18. Kerkhof K, Falconi-Agapito F, Van Esbroeck M, Talledo M, Ariën KK. 2020. Reliable Serological Diagnostic Tests for Arboviruses: Feasible or Utopia? *Trends Microbiol* 28:276–292.
19. Moalic V, Mercier B, Ferec C. 2004. Technologie Luminex<sup>TM</sup> : principe, applications, et perspectives. *Immuno-analyse & Biologie Spécialisée* 19:181–187.
20. DeBiasi RL, Tyler KL. 2004. Molecular Methods for Diagnosis of Viral Encephalitis. *Clin Microbiol Rev* 17:903–925.
21. Netto GJ, Saad RD, Dysert PA. 2003. Diagnostic molecular pathology: current techniques and clinical applications, part I. *Proc (Bayl Univ Med Cent)* 16:379–383.
22. Bustin SA, Mueller R. 2005. Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. *Clinical Science* 109:365–379.
23. Préparation et intervention d'urgence lors d'émergences de pathogènes d'intérêt médical et/ou vétérinaire. Agence nationale de la recherche. <https://anr.fr/Projet-ANR-20-SEBM-0004>. Retrieved 9 May 2023.

24. Badrick T. 2021. Integrating quality control and external quality assurance. *Clinical Biochemistry* 95:15–27.
25. Monteiro L, Bonnemaïson D, Vekris A, Petry KG, Bonnet J, Vidal R, Cabrita J, Mégraud F. 1997. Complex polysaccharides as PCR inhibitors in feces: *Helicobacter pylori* model. *J Clin Microbiol* 35:995–998.
26. Ninove L, Nougairède A, Gazin C, Thirion L, Delogu I, Zandotti C, Charrel RN, De Lamballerie X. 2011. RNA and DNA bacteriophages as molecular diagnosis controls in clinical virology: a comprehensive study of more than 45,000 routine PCR tests. *PLoS One* 6:e16142.
27. Vesper HW, Miller WG, Myers GL. 2007. Reference Materials and Commutability. *Clin Biochem Rev* 28:139–147.
28. Overview of External Quality Assessment (EQA). <https://www.who.int/publications/m/item/overview-of-external-quality-assessment-eqa>. Retrieved 20 September 2022.
29. Pasloske BL, Walkerpeach CR, Obermoeller RD, Winkler M, DuBois DB. 1998. Armored RNA Technology for Production of Ribonuclease-Resistant Viral RNA Controls and Standards. *J Clin Microbiol* 36:3590–3594.
30. Gholami M, Ravanshad M, Baesi K, Samiee SM, Rozbahani NH, Mohraz M. 2018. Preparation and Evaluation of Ribonuclease-Resistant Viral HIV RNA Standards Based on Armored RNA Technology. *Iran Biomed J* 22:394–400.
31. Fu Y, Zhang R, Wu Q, Zhang J, Bao L, Li J. 2018. Development and evaluation of armored RNA-based standards for quantification of BCR-ABL1 p210/p190 fusion gene transcripts. *J Clin Lab Anal* 32:e22612.

32. Atkins JF, Steitz JA, Anderson CW, Model P. 1979. Binding of mammalian ribosomes to MS2 phage RNA reveals an overlapping gene encoding a lysis function. *Cell* 18:247–256.
33. Kuzmanovic DA, Elashvili I, Wick C, O’Connell C, Krueger S. 2003. Bacteriophage MS2: molecular weight and spatial distribution of the protein and RNA components by small-angle neutron scattering and virus counting. *Structure* 11:1339–1348.
34. Shiba T, Suzuki Y. 1981. Localization of A protein in the RNA-A protein complex of RNA phage MS2. *Biochim Biophys Acta* 654:249–255.
35. Rolfsson Ó, Middleton S, Manfield IW, White SJ, Fan B, Vaughan R, Ranson NA, Dykeman E, Twarock R, Ford J, Cheng Kao C, Stockley PG. 2016. Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2. *J Mol Biol* 428:431–448.
36. Pickett GG, Peabody DS. 1993. Encapsidation of heterologous RNAs by bacteriophage MS2 coat protein. *Nucleic Acids Res* 21:4621–4626.
37. Huang Q, Cheng Y, Guo Q, Li Q. 2006. Preparation of a chimeric armored RNA as a versatile calibrator for multiple virus assays. *Clin Chem* 52:1446–1448.
38. Wei B, Wei Y, Zhang K, Yang C, Wang J, Xu R, Zhan S, Lin G, Wang W, Liu M, Wang L, Zhang R, Li J. 2008. Construction of armored RNA containing long-size chimeric RNA by increasing the number and affinity of the pac site in exogenous rna and sequence coding coat protein of the MS2 bacteriophage. *Intervirology* 51:144–150.
39. Lin G, Zhang K, Zhang D, Han Y, Xie J, Li J. 2017. Fast preparation of a long chimeric armored RNA as controls for external quality assessment for molecular detection of Zika virus. *Clin Chim Acta* 466:138–144.

40. Freystetter A, Paar C, Stekel H, Berg J. 2017. Armored DNA in recombinant Baculoviruses as controls in molecular genetic assays. *Appl Microbiol Biotechnol* 101:7259–7269.
41. Masker WE, Kuemmerle NB, Allison DP. 1978. In vitro packaging of bacteriophage T7 DNA synthesized in vitro. *J Virol* 27:149–163.
42. Shyamala V, Arcangel P, Cottrell J, Coit D, Medina-Selby A, McCoin C, Madriaga D, Chien D, Phelps B. 2004. Assessment of the target-capture PCR hepatitis B virus (HBV) DNA quantitative assay and comparison with commercial HBV DNA quantitative assays. *J Clin Microbiol* 42:5199–5204.
43. Gotsch A, Schubert A, Bombis A, Wiedmann M, Zauke M, Schorling S. 2007. Nuclease-resistant single-stranded DNA controls for nucleic acid amplification assays. *J Clin Microbiol* 45:2570–2574.
44. Zhang L, Sun Y, Chang L, Jia T, Wang G, Zhang R, Zhang K, Li J. 2015. A novel method to produce armored double-stranded DNA by encapsulation of MS2 viral capsids. *Appl Microbiol Biotechnol* 99:7047–7057.
45. Ledsgaard L, Kilstrup M, Karatt-Vellatt A, McCafferty J, Laustsen AH. 2018. Basics of Antibody Phage Display Technology. *Toxins (Basel)* 10:236.
46. Talk:Team:TCU Taiwan/M13Phage - 2014.igem.org. [https://2014.igem.org/Talk:Team:TCU\\_Taiwan/M13Phage](https://2014.igem.org/Talk:Team:TCU_Taiwan/M13Phage). Retrieved 6 June 2023.
47. Pardon E, Laeremans T, Triest S, Rasmussen SGF, Wohlkönig A, Ruf A, Muyldermans S, Hol WGJ, Kobilka BK, Steyaert J. 2014. A general protocol for the generation of Nanobodies for structural biology. *Nat Protoc* 9:674–693.

48. Mairhofer J, Wittwer A, Cserjan-Puschmann M, Striedner G. 2015. Preventing T7 RNA polymerase read-through transcription-A synthetic termination signal capable of improving bioprocess stability. *ACS Synth Biol* 4:265–273.
49. Lyakhov DL, He B, Zhang X, Studier FW, Dunn JJ, McAllister WT. 1998. Pausing and termination by bacteriophage T7 RNA polymerase. *J Mol Biol* 280:201–213.
50. Mikel P, Vasickova P, Kralik P. 2017. One-plasmid double-expression His-tag system for rapid production and easy purification of MS2 phage-like particles. *Sci Rep* 7:17501.
51. Hengen P. 1995. Purification of His-Tag fusion proteins from *Escherichia coli*. *Trends Biochem Sci* 20:285–286.
52. Al-Soud WA, Jönsson LJ, Rådström P. 2000. Identification and Characterization of Immunoglobulin G in Blood as a Major Inhibitor of Diagnostic PCR. *Journal of Clinical Microbiology* 38:345–350.
53. Fischer MG, Suttle CA. 2011. A virophage at the origin of large DNA transposons. *Science* 332:231–234.
54. Born D, Reuter L, Mersdorf U, Mueller M, Fischer MG, Meinhart A, Reinstein J. 2018. Capsid protein structure, self-assembly, and processing reveal morphogenesis of the marine virophage mavirus. *Proc Natl Acad Sci U S A* 115:7332–7337.
55. Xiao W, Jiang L, Wang W, Wang R, Fan J. 2018. Evaluation of rice tetratricopeptide domain-containing thioredoxin as a novel solubility-enhancing fusion tag in *Escherichia coli*. *J Biosci Bioeng* 125:160–167.
56. Qiu P, Shandilya H, D'Alessio JM, O'Connor K, Durocher J, Gerard GF. 2004. Mutation detection using Surveyor nuclease. *Biotechniques* 36:702–707.

57. Pilato B, De Summa S, Danza K, Papadimitriou S, Zaccagna P, Paradiso A, Tommasi S. 2012. DHPLC/SURVEYOR Nuclease: A Sensitive, Rapid and Affordable Method to Analyze BRCA1 and BRCA2 Mutations in Breast Cancer Families. *Mol Biotechnol* 52:8–15.
58. Pilch J, Asman M, Jamroz E, Kajor M, Kotrys-Puchalska E, Goss M, Krzak M, Witecka J, Gmiński J, Sieroń AL. 2010. Surveyor nuclease detection of mutations and polymorphisms of mtDNA in children. *Pediatr Neurol* 43:325–330.
59. Lopez E, Barthélémy M, Baronti C, Masse S, Falchi A, Durbesson F, Vincentelli R, de Lamballerie X, Charrel R, Coutard B. 2021. Endonuclease-based genotyping of the RBM as a method to track the emergence or evolution of SARS-CoV-2 variants. *iScience* 24:103329.
60. Davies H, Dicks E, Stephens P, Cox C, Teague J, Greenman C, Bignell G, O’Meara S, Edkins S, Parker A, Stevens C, Menzies A, Blow M, Bottomley B, Dronsfield M, Futreal PA, Stratton MR, Wooster R. 2006. High throughput DNA sequence variant detection by conformation sensitive capillary electrophoresis and automated peak comparison. *Genomics* 87:427–432.
61. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CAB, Russell TW, Tully DC, Washburne AD, Wenseleers T, Gimma A, Waites W, Wong KLM, van Zandvoort K, Silverman JD, CMMID COVID-19 Working Group, COVID-19 Genomics UK (COG-UK) Consortium, Diaz-Ordaz K, Keogh R, Eggo RM, Funk S, Jit M, Atkins KE, Edmunds WJ. 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372:eabg3055.

62. Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH. 2021. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* 593:270–274.
63. Liu Y, Liu J, Xia H, Zhang X, Fontes-Garfias CR, Swanson KA, Cai H, Sarkar R, Chen W, Cutler M, Cooper D, Weaver SC, Muik A, Sahin U, Jansen KU, Xie X, Dormitzer PR, Shi P-Y. 2021. Neutralizing Activity of BNT162b2-Elicited Serum. *N Engl J Med* NEJMc2102017.
64. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong PD, Graham BS, Mascola JR, Chang JY, Yin MT, Sobieszczyk M, Kyratsous CA, Shapiro L, Sheng Z, Huang Y, Ho DD. 2021. Antibody Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7. *bioRxiv* <https://doi.org/10.1101/2021.01.25.428137>.
65. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD. 2021. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* 29:463-476.e6.
66. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F. 2020. Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581:221–224.
67. Zahradník J, Marciano S, Shemesh M, Zoler E, Chiaravalli J, Meyer B, Rudich Y, Dym O, Elad N, Schreiber G. 2021. SARS-CoV-2 RBD in vitro evolution follows contagious mutation spread, yet generates an able infection inhibitor. *bioRxiv* <https://doi.org/10.1101/2021.01.06.425392>.
68. Oude Munnink BB, Nieuwenhuijse DF, Stein M, O’Toole Á, Haverkate M, Mollers M, Kamga SK, Schapendonk C, Pronk M, Lexmond P, van der Linden A, Bestebroer T,

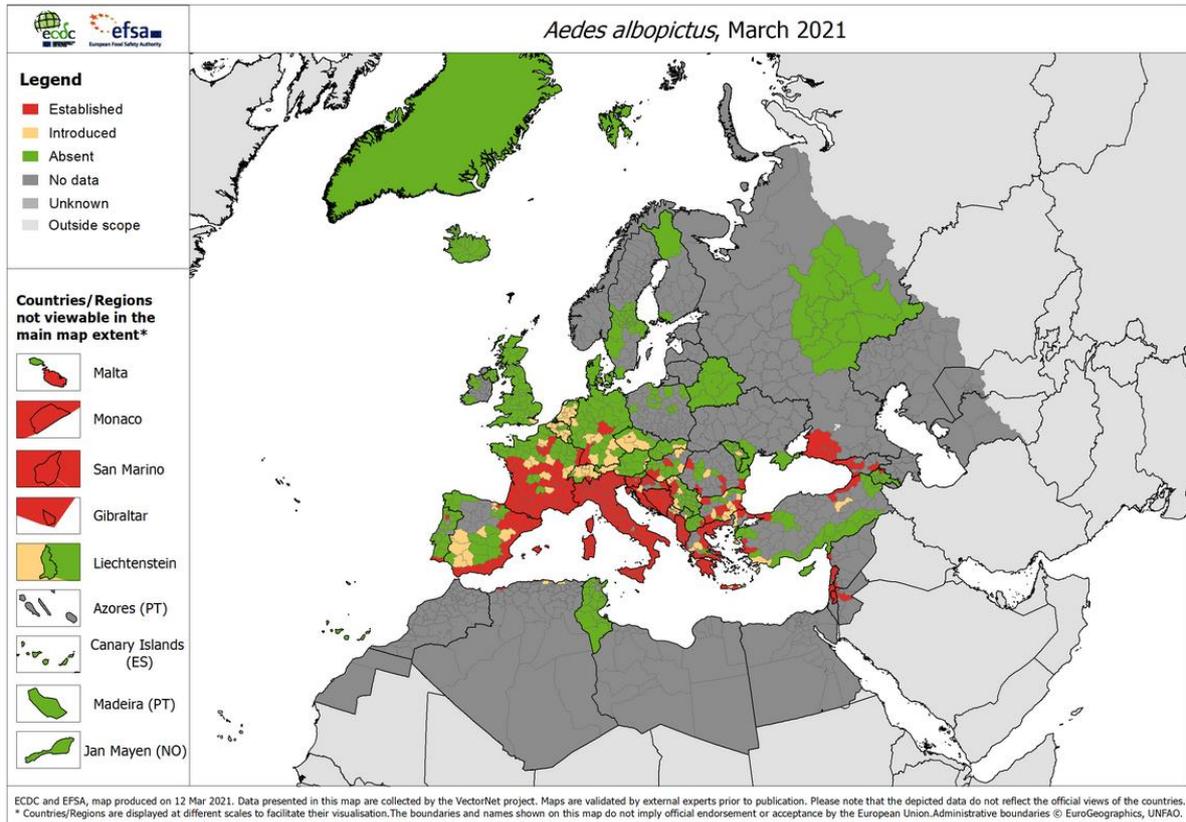
- Chestakova I, Overmars RJ, van Nieuwkoop S, Molenkamp R, van der Eijk AA, GeurtsvanKessel C, Vennema H, Meijer A, Rambaut A, van Dissel J, Sikkema RS, Timen A, Koopmans M. 2020. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. 9. *Nat Med* 26:1405–1410.
69. Grogan DW, Carver GT, Drake JW. 2001. Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. *Proc Natl Acad Sci U S A* 98:7928–7933.
70. Ishino S, Nishi Y, Oda S, Uemori T, Sagara T, Takatsu N, Yamagami T, Shirai T, Ishino Y. 2016. Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea. *Nucleic Acids Res* 44:2977–2986.
71. Nakae S, Hijikata A, Tsuji T, Yonezawa K, Kouyama K, Mayanagi K, Ishino S, Ishino Y, Shirai T. 2016. Structure of the EndoMS-DNA Complex as Mismatch Restriction Endonuclease. *Structure* 24:1960–1971.
72. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* 42:W320–W324.
73. Omatola CA, Olaniran AO. 2022. Rotaviruses: From Pathogenesis to Disease Control—A Critical Review. *Viruses* 14:875.

# ANNEXES

## Annexe 1

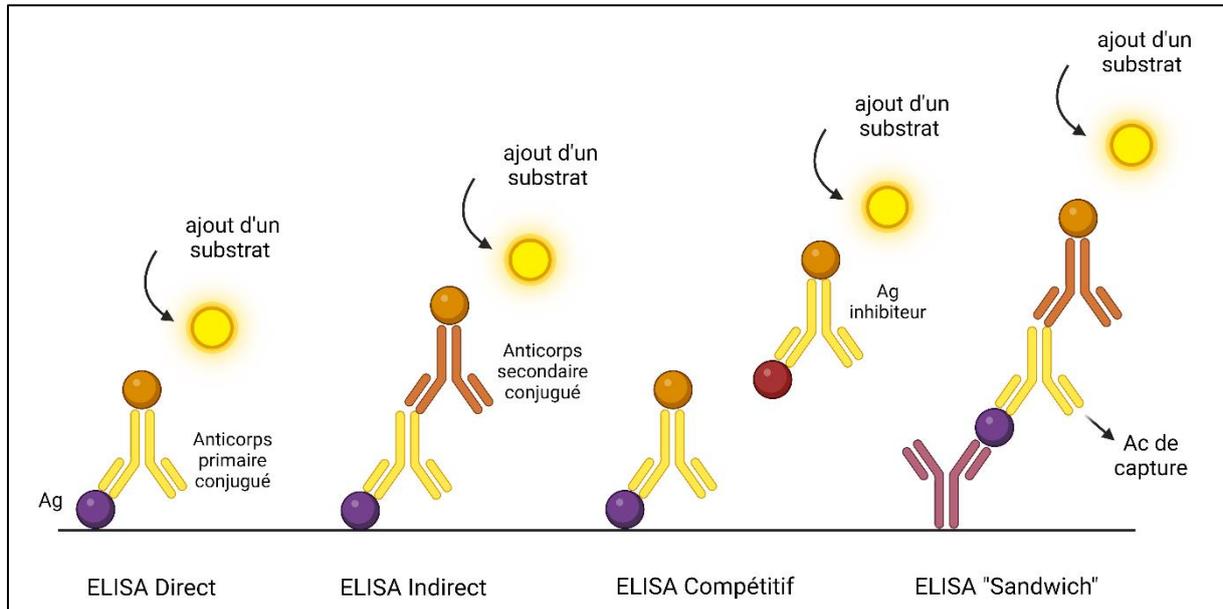
### Figure 4 A) complète

Figure extraite du site « European Centre for Disease Prevention and Control », consulté le 19 avril 2023.



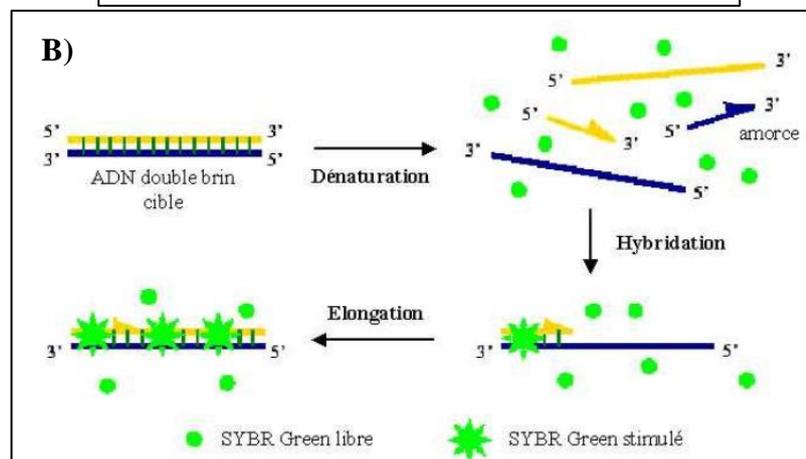
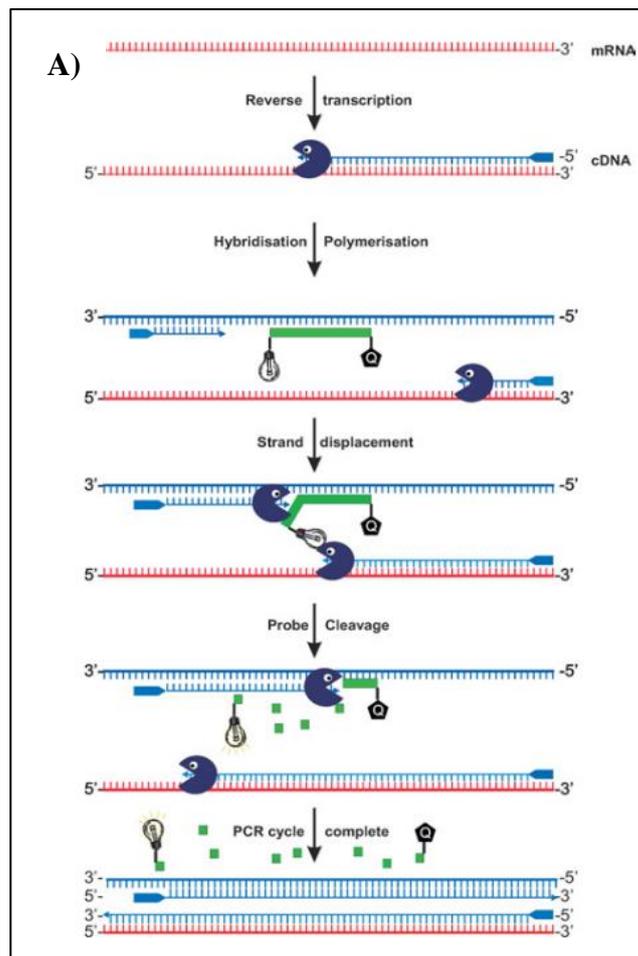
## Annexe 2

Représentation schématique des différents types d'ELISA. Figure générée avec BioRender



### Annexe 3

Présentation des techniques de détection de PCR en temps réelle, la technologie TaqMan (A) et la technologie SybrGreen (B)). La figure A) est extraite de l'article « *Bustin SA, Mueller R. 2005. Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. Clinical Science 109 :365-379.* La Figure B) est extraite de l'article « *(Bouladoux N.2003. Rôle de l'interféron-gamma dans l'immunité cellulaire anti-microsporidienne. Etude du modèle de la souris déficiente pour le récepteur à l'interféron-gamma, infectée oralement avec la microsporidie Encephalitozoon intestinalis. »* .



## Annexe 4

### Protocole de lyophilisation des témoins positifs

	Congélation	Lyophilisation 1 <sup>ère</sup> aire				Lyophilisation 2 <sup>ème</sup> aire	
Température (°C)	-30	-20	-20	-10	-10	20	20
Pression (mBar)		1,028	0,708	0,708	0,402	0,402	0,231
Temps	2h	15min	2h	1h	2h	1h	2h

## Annexe 5

Construction des gènes synthétique permettant la synthèse des protéines « fIMCP », « MCP<sub>ΔC</sub> » et « penton »

**AAA** : Thioredoxine (Solubilisation)

**AAA** : His<sub>6</sub>-Tag (récupération)

**AAA** : Site de clivage TEV protease (enlever le His<sub>6</sub>-Tag)

### **fIMCP :**

M**SDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA**  
**KLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSQKQLKEFLDANLAGSGSG**  
**H****MHHHHHH****G****GLENLYFQG****NTPPELDTVLQAPYA**YNWPTSKNVKIASRIGIPYSTFQTI  
QPVS DAPNNGIGQITFNQPLGNLTGGAPRLRVSFTA EIKNILADSSLKDQIGLKSFPVN  
RSIPVAVINMNGKTFTSYPAQLIKLHQYNADPLELALLSPCSDVDEYNKIKAVSMNNP  
YRQGTESTDSRMSRGLGCNYAYYIHPRAAGSTSVKIDFVVDEALVANPTQYKNIKDP  
VPFRNLNTFKVILDGQFKPENMIGIADDVKLVAGKADFEVDITGFKINMLVQNWVAP  
LEIGDIPKTIYNTPLISLEGNISSMCLNTKDPYGIPGERNKHILTTTHSMAMNNVPSMFA  
VMVSQETPTKKFAPDQLAGIIGLEIKVDSVDVGIFRELEQQQLYELSSSNGYNKRFSCFS  
GALANGLTVADPAVAAGNKFKEAIFGAGSVIFFRPSDLGLKDYNV MANANKSINMQ  
VQATFVTPEAAGTGAHYKLEVFSIRDNLTYSFEDGTFMDDLTYTPDQLRSPLKLT  
DNNKLMRVMGGSFMGDVMTNFNHMAAHPVTKTVTKLLRNAGPLKDYAGDGTMM  
GNIASVYGYGKKKTTTRKKKGGIVLLGSGKKGGKLSDKQLHDLRNL\*

### **MCPAC :**

M**MHHHHHH****G****GLENLYFQG****NTPPELDTVLQAPYA**YNWPTSKNVKIASRIGIPYSTFQTIQ  
PVSDAPNNGIGQITFNQPLGNLTGGAPRLRVSFTA EIKNILADSSLKDQIGLKSFPVNRS  
IPVAVINMNGKTFTSYPAQLIKLHQYNADPLELALLSPCSDVDEYNKIKAVSMNNPYR  
QGTESTDSRMSRGLGCNYAYYIHPRAAGSTSVKIDFVVDEALVANPTQYKNIKDPVP  
FRNLNTFKVILDGQFKPENMIGIADDVKLVAGKADFEVDITGFKINMLVQNWVAPLEI  
GDIPKTIYNTPLISLEGNISSMCLNTKDPYGIPGERNKHILTTTHSMAMNNVPSMFAVM  
VSQETPTKKFAPDQLAGIIGLEIKVDSVDVGIFRELEQQQLYELSSSNGYNKRFSCFSGA  
LANGLTVADPAVAAGNKFKEAIFGAGSVIFFRPSDLGLKDYNV MANANKSINMQVQ  
ATFVTPEAAGTGAHYKLEVFSIRDNLTYSFEDGTFMDDLTYTPDQLRSPLKLTDDN  
NKLMRVMGG\*

**Penton :**

MHHHHHGGLENLYFQKQYIWLNETIKSNKQLAGPRGSYKRPVSVDIFRSSTILDP  
DKNYLLIVEEFHLHKIRLPLFKPAGHDYQVGIFNRSTDEIMGVREVDVDFSTFVDEEDGYM  
YDYVDVGTAINETLAGLCDGIIGEEEDIPVFSFNKHSKKFEITTTENFRNGHFIMFNDDM  
RVDFNSFEFDDIDEEYSLVILNEDVETQDASTLEFLTPISHIVIESNDLPVSYELLPSISK  
NTTISDNTGVFLTNYKYLQNNQDYNLSILFRVENSSNKYHNILQTNFNRFNLSFTIYD  
YDNEKHPLTLLPQTVIQLKLLFESID\*

## Annexe 6

### Préparation des bactéries compétentes

**Isolation des colonies :** Les colonies sont isolées sur une boîte LB AGAR. La boîte est mise à l'étuve à 37°C over night

**Pré-culture :** Choisir une colonie isolée de la boîte Inoculer 2,5 ml (avec antibiotique) de milieu et incuber à 37°C over night à 200 rpm

**Culture :** Inoculer 100 ml de LB avec la préculture dans un erlen de 500 ml Laisser incuber à 37°C jusqu'à ce que la Do arrive à 0,4 - 0,6

**Compétence :** Centrifuger la culture dans 2 tubes **stériles** à 4000 rpm 10 min 4°C

Reprendre chaque culot dans 15 ml de solution TFB1 froide

Transvaser dans un seul tube de 50 ml Laisser 10 min dans la glace Centrifuger à 4000 rpm 10 min à 4°C

Reprendre le culot dans 5 ml de solution TFB2 froide

Laisser 30 dans la glace

Aliquoter à 50 µl les cellules dans des tubes **stériles** de 1,5 ml et stocker à -80°C jusqu'à transformation

Tester la contamination de la suspension par étalement sur plusieurs boîtes LB AGAR avec différents antibiotiques

#### **COMPOSITION DES TAMPONS**

##### **TFB1**

Acétate de potassium 30 mM

KCl 100 mM

CaCl<sub>2</sub> 10 mM

MnCl<sub>2</sub> 50 mM

Glycérol 15%

H<sub>2</sub>O qsp

##### **TFB2**

MOPS 10 mM

CaCl<sub>2</sub> 75 mM

KCl 10 mM

Glycérol 15%

H<sub>2</sub>O qsp

Les tampons sont stérilisés par filtration 0,22 µM **stérile** juste avant leur utilisation

## Annexe 7

**HINDIII**

pelB

**SfiI**

**NcoI**

**XhoI**

**BamHI**

**BstEII**

**NOTI**

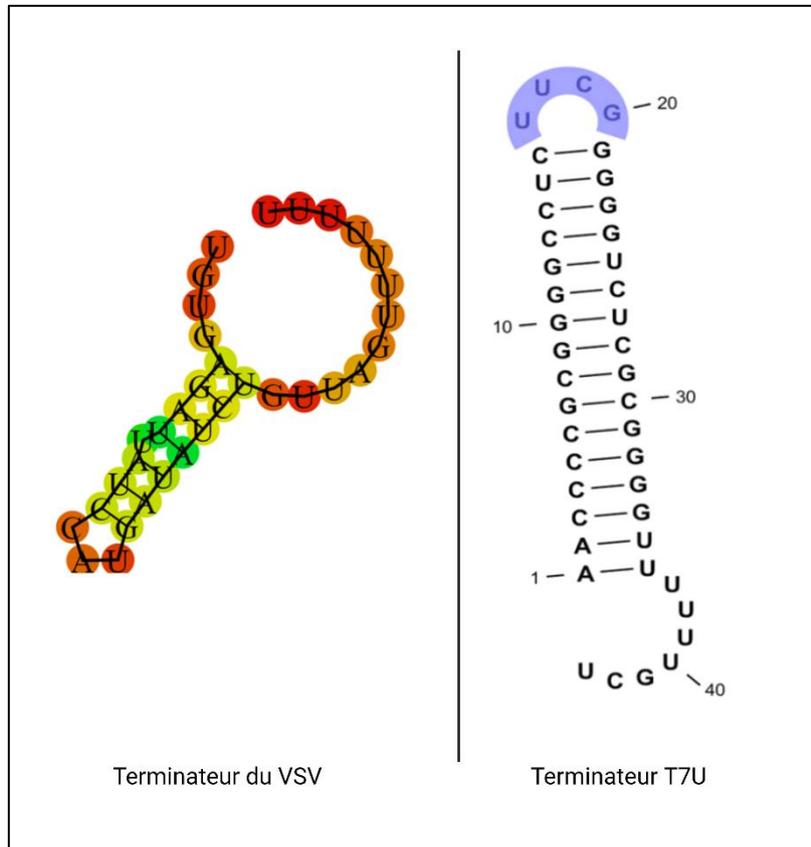
HA TAG

CACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCCCC  
GAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCG  
TATTGACGCCGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTG  
AGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGT  
GCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACC  
GAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGG  
AACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATG  
GCAACAACGTTGCGCAAACCTATTAACCTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATT  
AATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTG  
GCTGGTTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCA  
CTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAAC  
TATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAAC  
TGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTTTAATTTAAA  
AGGATCTAGGTGAAGATCCTTTTTTGATAATCTCATGACCAAATCCCTTAACGTGAGTTTTTC  
GTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTTC  
TGCGCGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTTGTTTGCCG  
GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAA  
TACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTA  
CATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTT  
ACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCCGGCTGAACGGGGG  
TTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTG  
AGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGC  
AGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCTGGTATCTTTATAG  
TCCTGTGCGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGC  
GGAGCCTATGGAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCT  
TTTGCTCACATGTTCTTTCCTGCGTTATCCCCGATTCTGTGGATAACCGTATTACCGCCTT  
TGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCCGAGCGCAGCGAGTCAGAAGCGCCC  
AATACGCAAACCGCCTCTCCCCGCGGTTGGCCGATTCAATTAATGCAGCTGGCACGACAGGT  
TTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAG  
GCACCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATA  
ACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGCATGCAAATCTATT  
TCAAGGAGACAGTCATAATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTACTC

GC GGCCCAGCCGG CCATGG CCCAGGTGCAGCTGCAGGAG CTCGAG GATCCG GTCACC GTCTC  
CA CCGGCCGC TACCCGTACGACGTTCCGGACTACGGTTCGGGCCGA GCATAGACTGTTGAAA  
GTTGTTTAGCAAACCTCATAACAGAAAATTCATTTACTAACGTCTGGAAAGACGACAAAAC  
TTAGATCGTTACGCTAACTATGAGGGCTGTCTGTGGAATGCTACAGGCGTTGTGGTTTGTAC  
TGGTGACGAAACTCAGTGTACGGTACATGGGTTCCCTATTGGGCTTGTATCCCTGAAAATG  
AGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTTACTAAA  
CCTCCTGAGTACGGTGATACACCTATTCCGGGCTATACTTATATCAACCTCTCGACGGCAC  
TTATCCGCCTGGTACTGAGCAAACCCCGCTAATCCTAATCCTTCTCTTGAGGAGTCTCAGC  
CTCTTAATACTTTCATGTTTCAGAATAATAGGTTCCGAAATAGGCAGGGTGCATTAACCTGTT  
TATACGGGCACTGTTACTCAAGGCACTGACCCCGTTAAACTTATTACCAGTACACTCCTGT  
ATCATCAAAGCCATGTATGACGTTACTGGAACGGTAAATTCAGAGACTGCGCTTTCATT  
CTGGCTTTAATGAGGATCCATTTCGTTTGTGAATATCAAGCCAATCGTCTGACCTGCCTCAA  
CCTCCTGTCAATGCTGGCGGCGGCTCTGGTGGTGGTCTGGTGGCGGCTCTGAGGGTGGCGG  
CTCTGAGGGTGGCGGTTCTGAGGGTGGCGGCTCTGAGGGTGGCGGTTCCGGTGGCGGCTCCG  
GTTCCGGTGATTTTGTATTATGAAAAATGGCAAACGCTAATAAGGGGGCTATGACCGAAAAT  
GCCGATGAAAACGCGCTACAGTCTGACGCTAAAGGCAAACCTGATTCTGTGCTACTGATTA  
CGGTGCTGCTATCGATGGTTTCATTGGTGACGTTTCCGGCCTTGCTAATGGTAATGGTGCTA  
CTGGTGATTTTGTGGCTCTAATTTCCCAAATGGCTCAAGTCGGTGACGGTGATAATTCACCT  
TTAATGAATAATTTCCGTCAATATTTACCTTCTTTGCCTCAGTCGGTTGAATGTGCCCCCTA  
TGTCTTTGGCGCTGGTAAACCATATGAATTTTCTATTGATTGTGACAAAATAAACTTATTCC  
GTGGTGTCTTTGCGTTTCTTTTATATGTTGCCACCTTTATGTATGTATTTTCGACGTTTGT  
AACATACTGCGTAATAAGGAGTCTTAATAA GAATTC ACTGGCCGTCGTTTTACAAC GTCGTG  
ACTGGGAAAACCTGGCGTATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCAC  
CGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTC  
TCCTTACGCATCTGTGCGGTATTTACACCCGCATACGTCAAAGCAACCATAGTACGCGCCCT  
GTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGC  
AGCGCCTTAGCGCCCGCTCCTTTCGCTTCTTCCCTTCTTCTCGCCACGTTCCGCCGGCTT  
TCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCGATTTAGTGCTTTACGGCACC  
TCGACCCCAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACG  
GTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACCTGG  
AACAACTCAACTCTATCTCGGGCTATTCTTTTGTATTATAAGGGATTTTGGCGATTTTCG  
TCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAAACAAAATATTA  
ACGTTTACAATTTTATGGTGCCTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCA  
GCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCG  
CTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTACCGTCATCA  
CCGAAACGCGGAGACGAAAGGGCCTCGTGATACGCCATTTTTTATAGGTTAATGTCATGAT  
AATAATGGTTTCTTAGACGTCAGGTGGCACTTTTTCGGGAAATGTGCGCGGAACCCCTATTT  
GTTTATTTTTCTAAATACA TTCAAATATGTATCCGCTCATGAGACAA TAACCCTGATAAATG  
CTTCAATAATATTGAAAAAGGAAGATATGAGTATTCAACATTTCCGTGTGCCCCCTTATTCC  
CTTTTTTGCGGCATTTTGCCTTCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAG  
ATGCTGAAGATCAGTTGGGTG

## Annexe 8

Structures secondaires des différents sites de terminaison de transcription testés. La structure secondaire du T7U a été extraite de l'article (48)



## Annexe 9

### “Insert” HBV

HIS TAG - (GS)<sub>2</sub> - GIIIp - TAA - Lambda T0 Term - XhoI - BamHI -  
BstEII - EcoRI

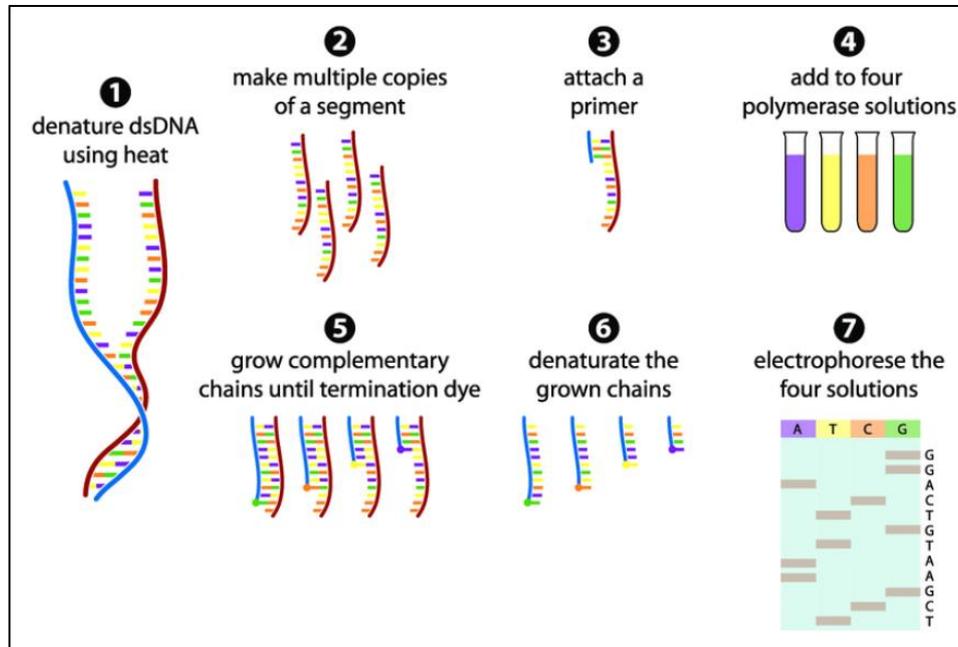
POCATCACCATCACCATCACCATCACCACGGTTCCTGGTTCGACTGTTGAAAGTTGTTTAGCAAACCTCATA  
GAAAATTCATTTACTAACGTCTGGAAAGACGACAAAACCTTTAGATCGTTACGCTAACTATGAGGGCTGTCTGTGG  
AATGCTACAGGCGTTGTGGTTTGTACTGGTGACGAAACTCAGTGTACGGTACATGGGTTCTATTGGGCTTGCT  
ATCCCTGAAAATGAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTA  
CCTCCTGAGTACGGTGATACACCTATTCCGGGCTATACTTATATCAACCTCTCGACGGCACTTATCCGCCTGGT  
ACTGAGCAAACCCCGCTAATCCTAATCCTTCTCTTGAGGAGTCTCAGCCTCTTAATACTTTTCATGTTTCAGAA  
AATAGGTTCCGAAATAGGCAGGGTGCATTAACGTGTTTATACGGGCACTGTTACTCAAGGCACTGACCCCGTTAA  
ACTTATTACCAGTACACTCCTGTATCATCAAAGCCATGTATGACGCTTACTGGAACGGTAAATTCAGAGACTGC  
GCTTTCCATTCTGGCTTTAATGAGGCTCAATTCGTTTGTGAATATCAAGGCCAATCGTCTGACCTGCCTCAACCT  
CCTGTCAATGCTGGCGGCGGCTCTGGTGGTGGTTCTGGTGGCGGCTCTGAGGGTGGCGGCTCTGAGGGTGGCGG  
TCTGAGGGTGGCGGCTCTGAGGGTGGCGGTTCCGGTGGCGGCTCCGGTTCGGTGATTTTGTATTATGAAAAATG  
GCAAACGCTAATAAGGGGGCTATGACCGAAAATGCCGATGAAAACGCGCTACAGTCTGACGCTAAAGGCAAACCT  
GATTCTGTGCTACTGATTACGGTGTCTATCGATGGTTTTCATTGGTGACGTTTCCGGCCTTGCTAATGGTAAT  
GGTGCTACTGGTGAATTTGCTGGCTCTAATCCCAAATGGCTCAAGTCGGTGACGGTGATAATTCACCTTTAATG  
AATAATTTCCGTCAATATTTACCTTCTTTGCCTCAGTCGGTTGAATGTCGCCCTTATGTCTTTGGCGCTGGTAAA  
CCATATGAATTTTCTATTGATTGTGACAAAATAAACTTATTCCGTGGTGTCTTTGCGTTTCTTTTATATGTTGCC  
ACCTTTATGTATGTATTTTTCGACGTTTGCTAACATACTGCGTAATAAAGGAGTCTTAATAAGACTCCTGTTGATAG  
ATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATGGTG  
AGAATCTCGAGGATCCGGTCACCGAATTC

Séquence cible HBV : Amorces - sonde

TaqMan GAATTCCTCTCGAGGATTGGGGACCTGCGCTGAACATGGAGAATCAGGATTCCTAGGACCC  
CTGCTCGTGTACAGGCGGGGTTTTCTTGTGACAAGAATCCTCACAATACCGCAGAGTCTAGACTCGTGGTGG  
ACTTCTCTCAATTTCTAGGGGGAACCTACCGTGTGTCTTGGCCAAAATTCGCAGTCCCAACCTCCAATCAC  
CCAACCTCCTGTCTCCAACCTTGTCTGGTTATTCGCTGGATGTGTCTGCGGCGTATTATAGCGGCCGCTTATTA  
TCCTGCTGCTATGCCTCATCTTCTTGTGGTTCTTCTGGACTATCAAGGATGTTGCCCGTTGTCTCTAATTC  
CAGGATCTTCGACCACCAGCACGGGACCCTGCAGAACCTGCACGACTGAATTC

## Annexe 10

Représentation schématique du fonctionnement de la méthode Sanger ((Gauthier, Michel. (2007). *Simulation of polymer translocation through small channels: A molecular dynamics study and a new Monte Carlo approach.*)



## Annexe 11

### Protocole de dénaturation et hybridation

Temperature	Time	Temperature ramp
95°C	10 min	
95°C to 85°C		(-2.0°C/sec)
85°C	1 min	
85°C to 75°C		(-0.3°C/sec)
75°C	1 min	
75°C to 65°C		(-0.3°C/sec)
65°C	1 min	
65°C to 55°C		(-0.3°C/sec)
55°C	1 min	
55°C to 45°C		(-0.3°C/sec)
45°C	1 min	
45°C to 35°C		(-0.3°C/sec)
35°C	1 min	
35°C to 25°C		(-0.3°C/sec)
25°C	1 min	
4°C	Hold ∞	

## Annexe 12

Construction des gènes synthétiques permettant la synthèse de différentes endonucléases de restriction

### WT

MASMTGGQQMGGA MSKDKVTVITSPSTEELVSLVNSALLEEAMLTIFARCKVHYDG  
RAKSELGSGDRVIIVKPDGSFLIHQSKKREPVNWQPPGSRVRLELRENPVLSIRRKPR  
ETLEVELEEYVMVSVFRAEDYEELALTGSEAEMAELIFENPEVIEPGFKPLFREKAIGT  
GIVDVLGRDSDGNIVVLELKRRAELHAVRQLKSYVEILREEYGDKVRGILVAPSLTS  
GAKRLEKEGLEFRKLEPPKRDSKKKGRQKTLF

### N76S + Q78G

MASMTGGQQMGGA MSKDKVTVITSPSTEELVSLVNSALLEEAMLTIFARCKVHYDG  
RAKSELGSGDRVIIVKPDGSFLIHQSKKREPVS<sup>W</sup>GPPGSRVRLELRENPVLSIRRKPR  
ETLEVELEEYVMVSVFRAEDYEELALTGSEAEMAELIFENPEVIEPGFKPLFREKAIGT  
GIVDVLGRDSDGNIVVLELKRRAELHAVRQLKSYVEILREEYGDKVRGILVAPSLTS  
GAKRLEKEGLEFRKLEPPKRDSKKKGRQKTLF

### N76G + Q78G

MASMTGGQQMGGA MSKDKVTVITSPSTEELVSLVNSALLEEAMLTIFARCKVHYDG  
RAKSELGSGDRVIIVKPDGSFLIHQSKKREPVS<sup>W</sup>G<sup>G</sup>PPGSRVRLELRENPVLSIRRKPR  
ETLEVELEEYVMVSVFRAEDYEELALTGSEAEMAELIFENPEVIEPGFKPLFREKAIGT  
GIVDVLGRDSDGNIVVLELKRRAELHAVRQLKSYVEILREEYGDKVRGILVAPSLTS  
GAKRLEKEGLEFRKLEPPKRDSKKKGRQKTLF

### N76S + Q78G +Insertion LIL

MASMTGGQQMGGA MSKDKVTVITSPSTEELVSLVNSALLEEAMLTIFARCKVHYDG  
RAKSELGSGDRVIIVKPDGSFLIHQSKKREPVS<sup>LIL</sup>W<sup>G</sup>PPGSRVRLELRENPVLSIRRK  
PRETLEVELEEYVMVSVFRAEDYEELALTGSEAEMAELIFENPEVIEPGFKPLFREKAI  
GTGIVDVLGRDSDGNIVVLELKRRAELHAVRQLKSYVEILREEYGDKVRGILVAPSL  
TSGAKRLEKEGLEFRKLEPPKRDSKKKGRQKTLF

### N76 + Q78G +Insertion LIF

MASMTGGQQMGGA MSKDKVTVITSPSTEELVSLVNSALLEEAMLTIFARCKVHYDG  
RAKSELGSGDRVIIVKPDGSFLIHQSKKREPVN<sup>LIF</sup>W<sup>G</sup>PPGSRVRLELRENPVLSIRR  
KPRETLEVELEEYVMVSVFRAEDYEELALTGSEAEMAELIFENPEVIEPGFKPLFREKA  
IGTGIVDVLGRDSDGNIVVLELKRRAELHAVRQLKSYVEILREEYGDKVRGILVAPS  
LTSGAKRLEKEGLEFRKLEPPKRDSKKKGRQKTLF

**GLU103Ala**

MASMTGGQQMGGA MSKDKVTVITSPSTEELVSLVNSALLEEAMLTIFARCKVHYDG  
RAKSELGSGDRVIIVKPDGSFLIHQSKKREPVNWQPPGSRVRLELRENPVLVSIRRKPR  
ATLEVELEEYVMVSVFRAEDYEELALTGSEAEMAELIFENPEVIEPGFKPLFREKAIGT  
GIVDVLGRDSDGNIVVLELKRRRAELHAVRQLKSYVEILREEYGDKVRGILVAPSLTS  
GAKRLEKEGLEFRKLEPPKRDSKKKGRQKTLF

**ASN76ALA**

MASMTGGQQMGGA MSKDKVTVITSPSTEELVSLVNSALLEEAMLTIFARCKVHYDG  
RAKSELGSGDRVIIVKPDGSFLIHQSKKREPVAWQPPGSRVRLELRENPVLVSIRRKPR  
ETLEVELEEYVMVSVFRAEDYEELALTGSEAEMAELIFENPEVIEPGFKPLFREKAIGT  
GIVDVLGRDSDGNIVVLELKRRRAELHAVRQLKSYVEILREEYGDKVRGILVAPSLTS  
GAKRLEKEGLEFRKLEPPKRDSKKKGRQKTLF

**Tyr41Trypto**

MASMTGGQQMGGA MSKDKVTVITSPSTEELVSLVNSALLEEAMLTIFARCKVHWVWV  
RAKSELGSGDRVIIVKPDGSFLIHQSKKREPVNWQPPGSRVRLELRENPVLVSIRRKPR  
ETLEVELEEYVMVSVFRAEDYEELALTGSEAEMAELIFENPEVIEPGFKPLFREKAIGT  
GIVDVLGRDSDGNIVVLELKRRRAELHAVRQLKSYVEILREEYGDKVRGILVAPSLTS  
GAKRLEKEGLEFRKLEPPKRDSKKKGRQKTLF

### Annexe 13

Construction des amorces utilisées pour tester l'efficacité des différentes endonucléases.

FWD\_Endo\_A : CGAACTGCCTGGAATCCTGACGACATGTAGCGAACGATCACCTCA

FWD\_Endo\_T : CGAACTGCCTGGAATCCTGACGACTTGTAGCGAACGATCACCTCA

FWD\_Endo\_C : CGAACTGCCTGGAATCCTGACGACCTGTAGCGAACGATCACCTCA

FWD\_Endo\_G : CGAACTGCCTGGAATCCTGACGACGTGTAGCGAACGATCACCTCA

REV\_Endo\_T : TGAGGTGATCGTTCGCTACATTGTCGTCAGGATTCCAGGCAGTTTCG

Rev\_Endo\_A : TGAGGTGATCGTTCGCTACAAGTCGTCAGGATTCCAGGCAGTTTCG

Rev\_Endo\_Cy5\_T : Cy5-

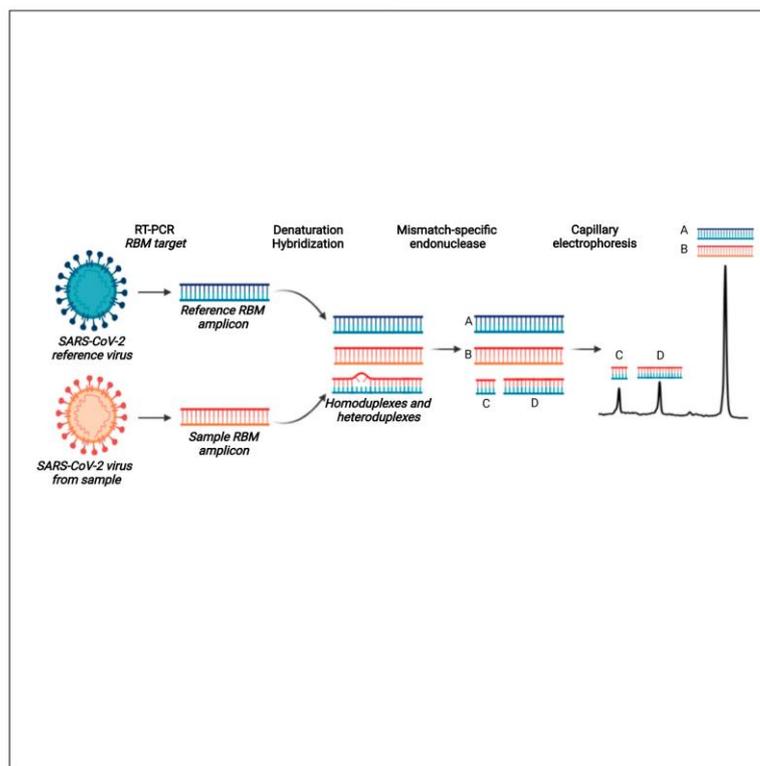
TGAGGTGATCGTTCGCTACATTGTCGTCAGGATTCCAGGCAGTTTCG

Rev\_Endo\_Cy5A : Cy5-

TGAGGTGATCGTTCGCTACAAGTCGTCAGGATTCCAGGCAGTTTCG

## Article

## Endonuclease-based genotyping of the RBM as a method to track the emergence or evolution of SARS-CoV-2 variants



Eva Lopez, Margot Barthélémy, Cécile Baronti, ..., Xavier de Lamballerie, Rémi Charrel, Bruno Coutard

bruno.coutard@univ-amu.fr

## Highlights

Mutations in SARS-CoV-2 RBM are important for virus adaptation

We developed a genotyping assay on the RBM coding sequence.

The genotyping is based on the use of a mismatch-specific endonuclease

The test can help monitor the emergence and circulation of VOCs

Lopez et al., iScience 24, 103329  
November 19, 2021 © 2021 The Authors.  
<https://doi.org/10.1016/j.isci.2021.103329>



## Article

## Endonuclease-based genotyping of the RBM as a method to track the emergence or evolution of SARS-CoV-2 variants

Eva Lopez,<sup>1</sup> Margot Barthélémy,<sup>1</sup> Cécile Baronti,<sup>1</sup> Shirley Masse,<sup>2</sup> Alessandra Falchi,<sup>2</sup> Fabien Durbesson,<sup>3</sup> Renaud Vincentelli,<sup>3</sup> Xavier de Lamballerie,<sup>1,4</sup> Rémi Charrel,<sup>1,4</sup> and Bruno Coutard<sup>1,5,\*</sup>

## SUMMARY

Since the beginning of the COVID-19 pandemics, variants have emerged. Some of them display increased transmissibility and/or resistance to immune response. Most of the mutations involved in the functional adaptation are found in the receptor-binding motif (RBM), close to the interface with the receptor ACE2. We thus developed a fast molecular assay to detect mutations in the RBM coding sequence. After amplification, the amplicon is heat-denatured and hybridized with an amplicon of reference. The presence of a mutation can be detected using a mismatch-specific endonuclease and the cleavage pattern is analyzed by capillary electrophoresis. The method was validated on RNA of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants produced *in vitro* before being implemented for clinical samples. The assay showed 97.8% sensitivity and 97.8% specificity. The procedure can be set up for high-throughput identification of the presence of mutations and serve as a first-line screening to select the samples for full genome sequencing.

## INTRODUCTION

In December 2019, individuals with pneumonia of unknown etiology were recorded in the city of Wuhan, China. The number of cases increased steadily in the following weeks including in the countries surrounding China (Taiwan, Thailand, Malaysia, etc.), and then throughout the world via aerial transport, triggering the WHO to announce a Public Health Emergency of International Concern (PHEIC) on the first of February 2020. Shortly after, the disease was named "COVID-19", for "Coronavirus disease 2019." The coronavirus in question, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is an enveloped, positive single-stranded RNA virus. The viral particle exposes at its surface the envelope glycoprotein spike (S). The S protein is a multi-domain protein (Figure 1A) involved in host cell recognition, in particular via its receptor-binding domain (RBD) which specifically binds the human angiotensin-2 converting enzyme (hACE2). The S protein is considered a major determinant of viral infectivity and antigenicity (Li et al., 2020; Walls et al., 2020), and mutations in the coding sequence of the S protein are susceptible to affect the biology of SARS-CoV-2.

Since the emergence of SARS-CoV-2, several non-synonymous mutations have been reported in the coding sequence of the S protein. Some of these non-synonymous mutations are particularly monitored because they are suspected to affect functions of the protein and thereby impact the biology of the virus. One of these first mutations led to D614G, a substitution contributing to the enhancement of viral loads in the upper respiratory tract with possible increased transmission (Plante et al., 2021). Later, several variants defined as a variant of concern (VoC) have emerged and are disseminated. Those variants may demonstrate increased transmissibility or severity of the disease, reduction of seroneutralization by antibodies induced by previous infection or vaccination, or resistance to therapeutic treatments. Among the VoCs, the first one—501Y.V1 or B.1.1.7 lineage—was identified in the United Kingdom and showed enhanced human-to-human transmission and increased disease severity (Davies et al., 2021a, 2021b). Then, variants of B.1.351 (501Y.V2), P.1 (501Y.V3), and B.617 lineages were isolated and characterized in South Africa, Brazil/Japan, and India, respectively. Both B.1.351 (501Y.V2) and P.1 (501Y.V3) variants show increased resistance to antibody neutralization (Davies et al., 2021; Liu et al., 2021a; Wang et al., 2021a). VoCs have in common to present at least one non-synonymous mutation in the spike receptor binding motif (RBM). RBM is the sub-domain of the RBD containing most of the hACE2-contacting residues and is also characterized by the presence of epitopes for neutralizing antibodies

<sup>1</sup>Unité des Virus Émergents (UVE: Aix-Marseille Univ-IRD 190-Inserm 1207), Marseille, France

<sup>2</sup>UR7310, Laboratoire de Virologie, Université de Corse-Inserm, 20250 Corte, France

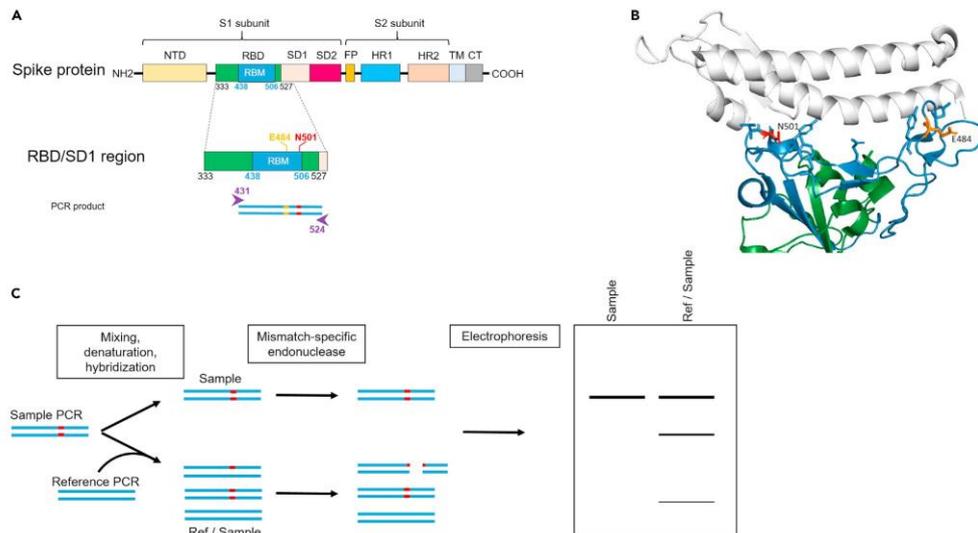
<sup>3</sup>Unité Mixte de Recherche (UMR) 7257, Centre National de la Recherche Scientifique (CNRS) Aix-Marseille Université, Architecture et Fonction des Macromolécules Biologiques (AFMB), Marseille, France

<sup>4</sup>Comité de Lutte contre les Infections Nosocomiales, Hôpitaux Universitaires de Marseille, AP-HM, Marseille, France

<sup>5</sup>Lead contact

\*Correspondence: bruno.coutard@univ-amu.fr  
<https://doi.org/10.1016/j.isci.2021.103329>





**Figure 1. RBM is a key region for viral entry and seroneutralization**

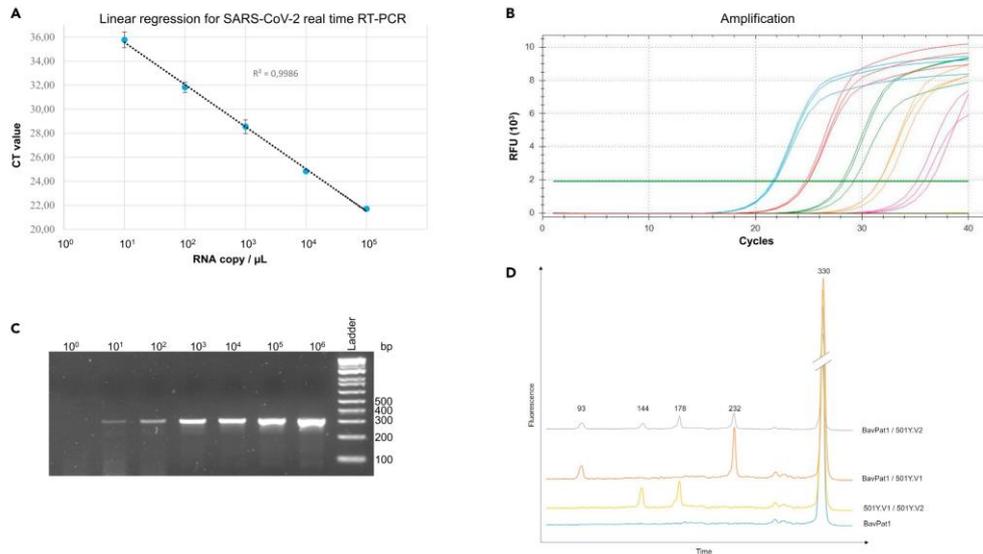
(A) Primary structure of the SARS-CoV-2 Spike protein. NTD: N-terminal domain; RBD: Receptor-binding domain; RBM: receptor-binding motif; SD1: subdomain 1; SD2: subdomain 2; FP: fusion peptide; HR1: Heptad repeat 1; HR2: Heptad repeat 2; TM: transmembrane region; CT: cytoplasmic region. Amino acid positions of the RBD and RBM boundaries are presented in black and blue numbers, respectively. Functional mutations in the RBM are shown in orange (E484K) and red (N501Y). Purple arrows and numbers correspond to the RT-PCR amplified region.

(B) Structure representation of the interaction between SARS-CoV-2 RBD core (green) and hACE2 receptor (white) (PDB 6ww1). RBM is shown in blue. E484 and N501 positions are highlighted in orange and red, respectively.

(C) Detection of mutation(s) in a sample compared with a reference sequence. Results of the digestion by a mismatch-specific endonuclease of an auto-control (amplicon of a sample alone) and mix (amplicons of reference and sample) are analyzed by electrophoresis and fragment patterns are detected.

(Figure 1B). Some of the characterized mutations are N501Y for 501Y.V1, E484K/N501Y for 501Y.V2 and 501Y.V3. By themselves, these mutations can affect the binding of the S protein to hACE2 and/or the potency of neutralizing antibodies (Greaney et al., 2021; Liu et al., 2021b; Shang et al., 2020; Wang et al., 2021b; Zahradnik et al., 2021). The genetic evolution of this region is specifically scrutinized to identify possible new VoCs. The rapid detection of VoCs is thus pivotal for mitigating transmission in hospital settings and for adjusting therapies to avoid lowering efficacy.

The detection of VoCs and surveillance of the evolution of SARS-CoV-2 population are currently surveyed by different approaches. Recently, researchers have developed methods such as RT-LAMP (LAMP SARS-CoV-2 variant detection panel, LaCAR), CRISPR-Cas9 or CRISPR-Cas13-based methods to specifically detect VoCs (Kumar et al., 2021; Wang et al., 2021c). However, the two most commonly used methods are the real-time RT-PCR for the search of mutations at given positions and massive campaigns of new-generation sequencing (NGS), both contributing to the public-health decision making (Oude Munnink et al., 2020). On one side, real-time RT-PCR is fast and operational on site, but it can detect only known mutations and does not address the newly emerging ones. On the other side, NGS can detect any mutation along the genome but the results are obtained in days rather than hours, delaying information required for medical decisions to be taken upon sequence identification. In addition, all the biological samples cannot be sequenced and upstream sampling is mandatory for the selection of the most relevant biological samples to characterize. Here we present the proof-of-concept for an alternative method allowing the surveillance of the genetic drift of SARS-CoV-2 in the RBM region where mutations are susceptible to affect the dissemination, pathogenicity or antibody-resistance of the virus. The technique, relying on the amplification of the RBM coding sequence followed by an assay using a mismatch-specific endonuclease, has been validated on biological samples demonstrating its feasibility.



**Figure 2. Evaluation of the target amplification and mismatch-specific nuclease assay on reference material**  
(A and B) Linearity of the real-time RT-PCR assay targeting the RBM coding sequence from  $10^5$  (blue curve),  $10^4$  (red curve),  $10^3$  (green curve),  $10^2$  (yellow curve), and  $10^1$  (pink curve) RNA copy/ $\mu\text{L}$ . The linear regression (dashed line) in Panel A has a correlation coefficient  $R^2=0.9986$ .  
(C) Endpoint analysis of PCR fragments on gel electrophoresis.  
(D) Electropherograms of the digestion products. Values above the peaks correspond to the size of the fragment. Peaks on the right (330bp) correspond to the homoduplexes created during the hybridization.

## RESULTS AND DISCUSSION

### Principle of the method

Some of the non-synonymous mutations occurring in the RBM coding sequence of SARS-CoV-2 are pivotal because they are susceptible to change the phenotype with possible influence on transmission pattern, increased pathogenesis, or immune escape; the latter can result in iterative infections, reduced vaccine efficacy, or resistance to mAb-based therapeutics. Accordingly, RBM genetic evolution has to be monitored for the surveillance of emerging variants. To detect mutations in the RBM region, we evaluated the SURVEYOR® Nuclease S, an endonuclease cleaving double-strand DNA where mismatches exist. The enzyme had already been used for genotyping for several purposes including the detection of mutations in *brca1* and *brca2* genes in the case of hereditary breast cancer (Davies et al., 2006; Pilato et al., 2012). The principle of the technique relies on the creation of heteroduplexes between an amplified DNA from a reference nucleic acid (*Ref*) and an amplified DNA from a sample to be evaluated (*Sample*). Both amplicons are mixed (Figure 1C). After denaturation/hybridization, a mixture of homoduplexes (*Ref/Ref* and *Sample/Sample*) and heteroduplexes (*Ref/Sample*) is produced. If the sequence from the *Sample* has mutations than the *Ref* sequence, mismatches happen in the heteroduplex. The latter is cleaved, resulting in cleavage products that can be evidenced by capillary electrophoresis, providing information on the number of SNPs and their approximate location(s). In contrast, if the sequences of *Ref* and *Sample* are identical, no cleavage is observed and only the full-length PCR products are visible by electrophoresis.

### Sensitivity of the reverse transcription-polymerase chain reaction assay targeting the receptor-binding motif coding sequence

Within the RBD region of the spike protein, the RBM contains amino acids (AA) subjected to mutations positions—for example, 452, 484, and 501—with functional relevance and observed in VoCs (Figure 1A). Its coding sequence was thus well-suited for a genotyping assay. Prerequisite was also that the targeted



**Table 1. Theoretical cleavage profiles for pairwise combinations of amplicons from BavPat1, 501Y.V1, and 501Y.V2 variants**

	0 BavPat1	N501Y 501.V1	E484K ; N501Y 501Y.V2
0	315	<u>84</u> / <u>231</u> / 315	<u>51</u> / <u>84</u> / <u>135</u> / <u>180</u> / <u>231</u> / 315
BavPat1			
N501Y	-	315	<u>135</u> / <u>180</u> / 315
501.V1			
E484K ; N501Y 501Y.V2	-	-	315

The amino acid residue and corresponding position in the sequence of the Spike protein for each variant are in italic. Values in the table correspond to the expected size of the fragments after treatment with the mismatch-specific endonuclease. Underlined: fragments resulting from the complete cleavage of the heteroduplexes; Bold: fragment resulting from an incomplete cleavage.

region is centered on the positions of interest and short enough so that it excludes the identification of mutations with low to no functional effect, that is, synonymous or non-synonymous with functional consequence. The size of the PCR product was thus set to 315-nt, allowing the detection of mutations in the region of the spike protein spanning AA positions 431 to 524 of the S protein.

To set up the assay, we first evaluated the sensitivity of the RT-PCR system in the range  $1-10^6$  of RNA copies/ $\mu\text{L}$ . The evaluation was conducted either by real-time RT-PCR with SYBR green (Figures 2A and 2B), or by endpoint analysis of the PCR products on agarose gel electrophoresis (Figure 2C). The amplification is linear from  $10$  to  $10^5$  copies of RNA/ $\mu\text{L}$ , with correlation coefficient  $R^2 = 0.9986$  (Figure 2A), and has a limit of detection on agarose gel up to 10 copies/ $\mu\text{L}$  (Figure 2C). However, given the quantity of material needed with the nuclease assay ( $>25$  ng/ $\mu\text{L}$ ), we arbitrarily applied the threshold at  $10^3/10^4$  copies/ $\mu\text{L}$ , which corresponds to samples with Ct values between 28 and 30 in reference detection systems (Pezzi et al., 2020). It should be noted that the SYBR green inhibits the endonuclease used for the detection of mismatches, likely by altering the structure of the DNA helix, thus rendering the resulting PCR product not suitable for subsequent capillary electrophoresis analysis (data not shown).

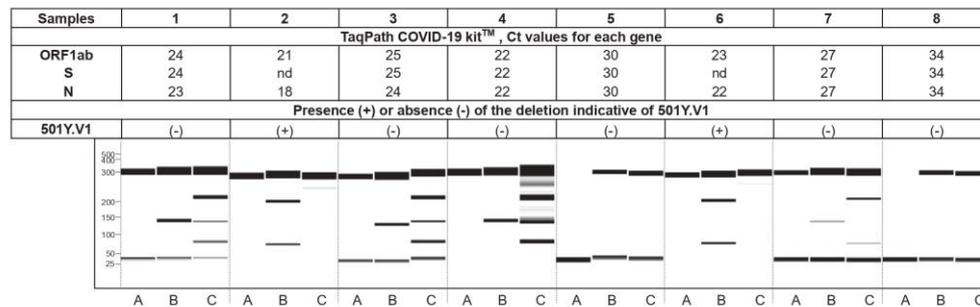
#### Detection and identification of mutations revealing in the receptor-binding motif

*Detection of mutations in reference material derived from severe acute respiratory syndrome coronavirus 2 isolates*

To establish the proof of concept, first experiments were conducted using viral RNA derived from cell cultures infected by three well-characterized variants: SARS-CoV-2 BavPat1, 501Y.V1 and 501Y.V2; (i) 501Y.V1 has N501Y (nt A1501T) mutation, (ii) and 501Y.V2 has E484K and N501Y (nt G1450A and A1501T, respectively) mutations by reference to the BavPat1, respectively. The theoretical cleavage profiles for pairwise combinations of the three variants are presented in Table 1. In practice, PCR products were mixed in pairs in approximate equimolar quantities before denaturation/hybridization prior to the mismatch-specific endonuclease assay. The results are presented in Figure 2D. The electropherogram corresponding to the hybridization of the BavPat1 amplicon with itself (blue curve), with no mismatch expected, resulted in a unique 330-nt long fragment. When mixed with one of 501Y.V1 (Figure 2D, orange curve), three DNA fragments were observed at 330, 232, and 93 bp, close to the anticipated profile (Table 1). For the identification of the 501Y.V2 (grey curve), five DNA fragments can be detected, corresponding to the products from both complete and partial cleavages of the mismatches at the expected positions. Finally, when mixing amplicons from the 501Y.V1 and 501Y.V2 samples (yellow curve), three DNA fragments were observed, where the two smaller being indicative of the difference at position 484 in the spike protein sequence of the two variants.

*Application to clinical samples containing severe acute respiratory syndrome coronavirus 2 ribonucleic acid as diagnosed by routine real-time reverse transcription-polymerase chain reaction assay*

To further assess the mutation detection method, 8 samples were tested after they were diagnosed as SARS-CoV-2 RNA positive using the routine diagnostic assay TaqPath real-time RT-PCR (Thermo Fisher



**Figure 3. Validation of the assay on SARS-CoV-2 positive clinical samples**

The First row indicates the number of the samples, from 1 to 8. Ct obtained from TaqPath COVID-19 kit™ (Thermo Fisher) real-time RT-PCR for ORF1ab, S and N genes are presented in the three next rows, respectively (nd: not determined, Ct>35), leading to the following classification: (+) refers to the presence of the deletion indicative of a 501Y.V1 variant, and (-) to a non-501Y.V1 sample. Below the table, the results of the mismatch-specific assay followed by capillary electrophoresis are presented in a "gel-like" format for the eight samples. The molecular weight ladder is indicated on the left. For each sample (1 to 8), the experiment was conducted in three conditions: The first lane is the self-hybridization (sample/sample; Lane A), the second one is sample/BavPat1 reference (lane B), and the third sample/501Y.V1 reference (Lane C). The fragment above 330 bp corresponds to uncleaved homoduplexes.

Scientific), known to discriminate the 501Y.V1 variant on the deletion observed in the S coding sequence, leading to the amplification of only two targets out the three of the test (Kidd et al., 2021). In order to make the analysis easy and rapid, it is advocated to test each sample against itself and against the selected reference with no prior quantification. In our case, the reference was the European lineage of SARS-CoV-2 (BavPat1 strain) since it was the dominant one at the time of the study. Clinical samples were identified as either SARS-CoV-2 positive for the three targets of the TaqPath assay™ (Figure 3, samples 1, 3, 4, 5, 7, and 8) or SARS-CoV-2/501Y.V1 positive (Figure 3 samples 2 and 6) with Ct ranging from 18 to 34. For samples 5 and 8, no fragment corresponding to the expected amplicon was visible in the self-hybridization assay (Figure 3, lanes 13 and 22), in line with the high Ct values obtained from initial real-time RT-PCR assays. By contrast, the amplification of the target region for the other sample was satisfactory (Figure 3, lanes 1, 4, 7, 10, 16, and 19). Overall, SARS-CoV-2 positive and 501Y.V1-putative samples (2 and 6) showed profiles matching with the expectations, that is, one mismatch with BavPat1 reference amplicon and no mismatch with 501Y.V1 reference amplicon (Figure 3, lanes 5, 6 and 17, 18, respectively).

For the other samples (Figure 3, samples 1, 3, 4, and 7), described as non-501Y.V1 according to the TaqPath assay, no cleavage was expected when compared with the BavPat1 reference amplicon, and two cleaved products with the 501Y.V1. However, at least one additional fragment was observed against both reference samples, of about 135 bp length (Figure 3, lanes 2, 3, 8, 9, 11, 12, 20, and 21). The PCR products were submitted to Sanger sequencing and a non-synonymous mutation yielding S477N substitution was detected, in agreement with the size of the cleaved products. This mutation had already been reported in viral populations circulating in Europe (<https://www.gisaid.org/>), and the corresponding variant has been shown to slightly increase the RBD's affinity for hACE2 (Hodcroft et al., 2020).

We next blindly evaluated the assay on 92 SARS-CoV-2 positive samples, for which Ct values were below 28 and NGS data available (Figure S1). The sensitivity, defined as the ability to generate an amplicon with yields compatible for the nuclease assay, is 97.83% as 2 samples out of 92 were not properly amplified. The specificity was defined as the ability to detect in the RT-PCR-positive samples a mutation compared with the BavPat1 strain or detect the lack of mutation for sequences identical to the reference. Compared with the sequence data, 2 samples out of 90 were inadequately identified, with a selectivity of 97.78%. Out of the 88 remaining profiles, 82 had the same profile as in Figure 2D. Yet unmet patterns were observed and confirmed by the sequence analysis with E471D mutation detected in two samples, E484K in one sample, L452R in one sample and both L452R / N501Y in 2 others.

Altogether, these results demonstrate that the mismatch-specific nuclease assay coupled with capillary electrophoresis is a suitable assay for the detection in clinical samples of already reported mutations.

The identification of atypical profiles, confirmed in this study by sequencing, also demonstrates that it is relevant for the discovery of yet unmet variants. With the incremental characterization of variants, the set of reference cleavage patterns can be updated to adapt the assay to circulating strains. As this technique can be dimensioned for 96/384 well devices and requires less than 4 h from the extracted RNA to production of individual results, it can be used to filter SARS-CoV-2 positive clinical samples and identify those for which virus isolation and complete genome sequencing is justified for surveillance purpose.

In conclusion, we developed a molecular assay dedicated to the surveillance of SARS-CoV-2 variants, specifically targeting the RBM coding sequence known to be involved in the functional adaptation of the virus. The assay is suitable to screen biological samples and identify the presence of new or emerging mutations.

#### Limitations of study

The technique we developed is based on RT-PCR amplification of the RBM coding sequence followed by mismatch-specific endonuclease assay and detection by DNA capillary electrophoresis and is made possible only for samples with RNA titers yielding Ct values better than 28 to produce enough amplified DNA material. The proof of concept has been validated but the method remains to be deployed on a large cohort to assess if it is amenable to high throughput genotyping or to individual surveillance of virus evolution during chronic infection by SARS-CoV-2, a situation favorable for virus adaptation (Kemp et al., 2021).

#### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAIL
  - Purified viral RNA from infected cell cultures
  - RNA from clinical samples
- METHOD DETAILS
  - Extraction of the RNA from infected cell cultures
  - Extraction of the RNA from clinical samples
  - Choice of the target sequence for amplification
  - RT-PCR protocol of the targeted sequence
  - Evaluation of the linearity in the designed system
  - Production of the reference PCR products
  - Production of the PCR products from biological samples
  - Mismatch-specific endonuclease assay and detection of cleavage

#### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.103329>.

#### ACKNOWLEDGMENTS

We thank Professor Bruno Lina and the Center National de Référence des virus des infections respiratoires for sharing extracted RNA of clinical samples and sequencing information. This study was funded by (i) the “European Virus Archive Global” (EVA-GLOBAL) project H2020-INFRAIA-2019 program, Project No 871029, (ii) the “Advanced Nanosensing platforms for Point of care global diagnostics and surveillance” (CONVAT), H2020, Project No 101003544. Eva Lopez is the recipient a DGA fellowship (Direction Générale de l’Armement).

#### AUTHOR CONTRIBUTIONS

B.C conceived the idea and supervised the work. E.L, M.B, C.B, F.D, and R.V carried out the experiments. S.M and A.L provided material on related information. E.L, R.C, and B.C wrote the article. All the authors discussed the results and commented on the manuscript.

**DECLARATION OF INTERESTS**

The authors declare no competing interests

Received: August 23, 2021

Revised: August 23, 2021

Accepted: October 19, 2021

Published: November 19, 2021

**REFERENCES**

- Bal, A., Destras, G., Gaymard, A., Stefic, K., Marlet, J., Eymieux, S., Regue, H., Semanas, Q., d'Aubarede, C., Billaud, G., et al. (2021). Two-step strategy for the identification of SARS-CoV-2 variant of concern 20212/01 and other variants with spike deletion H69-V70, France, August to December 2020. *Eurosurveillance* 26, 2100008. <https://doi.org/10.2807/1560-7917.ES.2021.26.3.2100008>.
- Davies, H., Dicks, E., Stephens, P., Cox, C., Teague, J., Greenman, C., Bignell, G., O'Meara, S., Edkins, S., Parker, A., et al. (2006). High throughput DNA sequence variant detection by conformation sensitive capillary electrophoresis and automated peak comparison. *Genomics* 87, 427–432. <https://doi.org/10.1016/j.ygeno.2005.11.008>.
- Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., Pearson, C.A.B., Russell, T.W., Tully, D.C., et al. (2021a). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*. <https://doi.org/10.1126/science.abcg3055>.
- Davies, N.G., Jarvis, C.I., CMMID COVID-19 Working Group, Edmunds, W.J., Jewell, N.P., Diaz-Ordaz, K., and Keogh, R.H. (2021b). Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature*. <https://doi.org/10.1038/s41586-021-03426-1>.
- Greaney, A.J., Loes, A.N., Crawford, K.H.D., Starr, T.N., Malone, K.D., Chu, H.Y., and Bloom, J.D. (2021). Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* 29, 463–476.e6. <https://doi.org/10.1016/j.chom.2021.02.003>.
- Hodcroft, E.B., Zuber, M., Nadeau, S., Crawford, K.H.D., Bloom, J.D., Veessler, D., Vaughan, T.G., Comas, I., Candelas, F.G., Stadler, T., and Neher, R.A. (2020). Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv*. <https://doi.org/10.1101/2020.10.25.20219063>.
- Kemp, S.A., Collier, D.A., Datir, R.P., Ferreira, I.A.T.M., Gayed, S., Jahun, A., Hosmillo, M., Rees-Spear, C., Micochova, P., Lumb, I.U., et al. (2021). SARS-CoV-2 evolution during treatment of chronic infection. *Nature* 592, 277–282. <https://doi.org/10.1038/s41586-021-03291-y>.
- Kidd, M., Richter, A., Best, A., Cumley, N., Mirza, J., Percival, B., Mayhew, M., Megram, O., Ashford, F., White, T., et al. (2021). S-variant SARS-CoV-2 lineage B.1.1.7 is associated with significantly higher viral loads in samples tested by ThermoFisher TaqPath RT-qPCR. *J. Infect. Dis.* <https://doi.org/10.1093/infdis/jiab082>.
- Kumar, M., Gulati, S., Ansari, A.H., Phutela, R., Acharya, S., Azhar, M., Murthy, J., Kathpalia, P., Kanakan, A., Maurya, R., et al. (2021). FnCas9-based CRISPR diagnostic for rapid and accurate detection of major SARS-CoV-2 variants on a paper strip. *eLife* 10, e67130. <https://doi.org/10.7554/eLife.67130>.
- Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., Zhao, C., Zhang, Q., Liu, H., Nie, L., et al. (2020). The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* 182, 1284–1294.e9. <https://doi.org/10.1016/j.cell.2020.07.012>.
- Liu, Y., Liu, J., Xia, H., Zhang, X., Fontes-Garfias, C.R., Swanson, K.A., Cai, H., Sarkar, R., Chen, W., Cutler, M., et al. (2021a). Neutralizing activity of BNT162b2-elicited serum. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMc2102017>.
- Liu, Z., VanBlargan, L.A., Bloyet, L.-M., Rothlauf, P.W., Chen, R.E., Stumpf, S., Zhao, H., Errico, J.M., Theel, E.S., Liebeskind, M.J., et al. (2021b). Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* 29, 477–488.e4. <https://doi.org/10.1016/j.chom.2021.01.014>.
- Oude Munnink, B.B., Nieuwenhuijse, D.F., Stein, M., O'Toole, A., Haverkate, M., Mollers, M., Kamsa, S.K., Schapendonk, C., Pronk, M., Lexmond, P., and Dutch-Covid-19 response team. (2020). Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat. Med.* 26, 1405–1410. <https://doi.org/10.1038/s41591-020-0997-y>.
- Pezzi, L., Charrel, R.N., Ninove, L., Nougairède, A., Molle, G., Coutard, B., Durand, G., Leparco-Goffart, I., de Lamballerie, X., and Thirion, L. (2020). Development and evaluation of a duo SARS-CoV-2 RT-qPCR assay combining two assays approved by the World Health Organization targeting the envelope and the RNA-dependent RNA polymerase (RdRp) coding regions. *Viruses* 12. <https://doi.org/10.3390/v12060686>.
- Pilato, B., De Summa, S., Danza, K., Papadimitriou, S., Zaccagna, P., Paradiso, A., and Tommasi, S. (2012). DHPLC/SURVEYOR nuclease: a sensitive, rapid and affordable method to analyze BRCA1 and BRCA2 mutations in breast cancer families. *Mol. Biotechnol.* 52, 8–15. <https://doi.org/10.1007/s12033-011-9468-5>.
- Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R., et al. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592, 116–121. <https://doi.org/10.1038/s41586-020-2895-3>.
- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581, 221–224. <https://doi.org/10.1038/s41586-020-2179-y>.
- Touret, F., Gilles, M., Barral, K., Nougairède, A., van Helden, J., Decroly, E., de Lamballerie, X., and Coutard, B. (2020). In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication. *Sci. Rep.* 10, 13093. <https://doi.org/10.1038/s41598-020-70143-6>.
- Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veessler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 281–292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>.
- Wang, P., Nair, M.S., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P.D., et al. (2021a). Increased resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7 to antibody neutralization. *bioRxiv Prepr. Serv. Biol.* <https://doi.org/10.1101/2021.01.25.428137>.
- Wang, P., Nair, M.S., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P.D., et al. (2021b). Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *bioRxiv*. <https://doi.org/10.1101/2021.01.25.428137>.
- Wang, Y., Zhang, Y., Chen, J., Wang, M., Zhang, T., Luo, W., Li, Y., Wu, Y., Zeng, B., Zhang, K., et al. (2021c). Detection of SARS-CoV-2 and its mutated variants via CRISPR-Cas13-based transcription amplification. *Anal. Chem.* <https://doi.org/10.1021/acs.analchem.0c04303>.
- Zahradnik, J., Marciano, S., Shemesh, M., Zoler, E., Chiaravalli, J., Meyer, B., Rudich, Y., Dym, O., Elad, N., and Schreiber, G. (2021). SARS-CoV-2 RBD in vitro evolution follows contagious mutation spread, yet generates an able infection inhibitor. *bioRxiv*. <https://doi.org/10.1101/2021.01.06.425392>.



## STAR METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Bacterial and virus strains</b>		
SARS-CoV-2 BavPat1 strain	Pr. C. Drosten	N/A
SARS-CoV-2 501Y.V1 strain	EVA-G	001V-04044
SARS-CoV-2 501Y.V1 strain	EVA-G	001V-04067
<b>Biological samples</b>		
Nasopharyngeal samples for the proof of concept	A. Falchi (Corsica)	N/A
Nasopharyngeal samples for the validation of the study	B. Lina (CNR Lyon)	N/A
<b>Chemicals, peptides, and recombinant proteins</b>		
Minimal essential medium (MEM)	Thermo Fischer	#21090022
Fetal calf serum	Thermo Fischer	#10270098
Penicillin-streptomycin	Thermo Fischer	#15070063
HEPES	Sigma Aldrich	#15630056
<b>Critical commercial assays</b>		
QIAmp 96 DNA virus QIAcube HT kit	Qiagen	#51331
Express one step superscript qRT-PCR kit	Thermo Fischer	#12574-035
TaqPath COVID-19 kit	Thermo Fischer	#A48067
Superscript II one-step platinum kit	Thermo Fischer	#10928042
QuantiTect SYBR green RT-PCR kit	Qiagen	#204243
Molecular size ladder DNA 1kb plus	Thermo Fischer	#10787026
Surveyor mutation detection kit	IDT	#706021
<b>Experimental models: Cell lines</b>		
VeroE6/TMPRSS2+	NIBSC	CFAR #100978
<b>Oligonucleotides</b>		
<i>Primers and probe for SARS-CoV-2 N gene</i>	This paper	N/A
Forward : 5'-GGCCGCAAATTGCACAAT-3'		
Reverse : 5'-CCAATGCGCGACATTCC-3'		
Probe : FAM-CCCCAGCGCTTCAGGTTCT-BHQ1		
<i>Primers for SARS-CoV-2 RBM site</i>	This paper	N/A
Forward : 5'-AGACTTTTATAGGTCCACAAAC-3'		
Reverse : 5'-TTACCAGATGATTTTACAGGC-3'		
<b>Other</b>		
Fragment analyzer 5200	Agilent	N/A
LabChip GXII electrophoresis	Perkin	N/A

### RESOURCE AVAILABILITY

#### Lead contact

Requests for further information should be directed to Bruno Coutard ([bruno.coutard@univ-amu.fr](mailto:bruno.coutard@univ-amu.fr)).

#### Materials availability

This study did not generate new unique material or reagents.

**Data and code availability**

- All data reported in this paper will be shared by the lead contact upon request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

**EXPERIMENTAL MODEL AND SUBJECT DETAIL****Purified viral RNA from infected cell cultures**

VeroE6/TMPRSS2+ (CFAR#100978) cells were grown in minimal essential medium (MEM) (Thermo Fisher Scientific) with 7.5% heat-inactivated fetal calf serum (FCS; Thermo Fisher Scientific), with 1% penicillin/streptomycin (PS, 5000 U.mL<sup>-1</sup> and 5000 µg.mL<sup>-1</sup> respectively; Thermo Fisher Scientific), supplemented with 1% non-essential amino acids (Thermo Fisher Scientific) and L-Glutamine (Thermo Fisher Scientific), at 37 °C with 5% CO<sub>2</sub>. SARS-CoV-2 strain BavPat1 was obtained from Pr. C. Drosten through EVA GLOBAL (<https://www.european-virus-archive.com/>). SARS-CoV-2 strain 2021/FR/7b-ex UK (EVA-G ref: 001V-04044) belonging to lineage 501Y.V1 and strain 2021/FR/1299-ex SA (EVA-G ref: 001V-04067) belonging to 501Y.V2 were isolated from Human nasopharyngeal swabs. The virus stocks were produced on VeroE6/TMPRSS2+ cells. Briefly, a 25 cm<sup>2</sup> flask of sub-confluent cells was inoculated with each strain at MOI=0.001. Cells were incubated at 37 °C overnight, after which the medium was changed, and incubation was continued for 24 h. The supernatant was then collected, clarified by spinning at 1500 g for 10 min, supplemented with 25 mM HEPES (Sigma-Aldrich) and stored at -80 °C in several aliquots. All infectious experiments were conducted in a biosafety level 3 laboratory (BSL3).

**RNA from clinical samples**

For the first set of 8 clinical samples, we collected anonymised residual nasopharyngeal (NP) samples of COVID-19 patients confirmed by real time RT-PCR from clinical laboratories based in Corsica in mid-January 2021. No nominative nor sensitive data on participant people have been collected. This study falls within the scope of the French Reference Methodology MR-004 according to 2016-41 law dated 26 January 2016 on the modernization of the French health system. The ethics committee of University of Corsica Pascal Paoli (IRB UCPP 2020-01) approved this study.

The collection of 92 SARS-CoV-2 positive clinical samples used in the second evaluation phase was kindly provided by the National Reference Center for Respiratory Viral Infections (CNR Lyon, France). The SARS-CoV-2 genomes from the 92 samples were fully sequenced by the CNR for the surveillance of the variants.

**METHOD DETAILS****Extraction of the RNA from infected cell cultures**

RNA extraction was performed using the QIAamp 96 DNA kit and the Qiacube HT plasticware kit on the Qiacube HT automate (Qiagen). Viral RNA obtained was quantified by real-time RT-PCR (EXPRESS One-Step Superscript qRT-PCR Kit (Thermo Fisher Scientific) using 3.5 µL of RNA and 6.5 µL of qRT-PCR mix and standard fast cycling parameters, i.e., 10 min at 50 °C, 2 min at 95 °C, and 40 amplification cycles (95 °C for 3 s followed by 30 s at 60 °C) (Touret et al., 2020). The quantification was provided by four log 2 serial dilutions of an appropriate T7-generated synthetic RNA standard of known quantities (102 to 108 copies). Real-time RT-PCR reactions were performed on QuantStudio 12K Flex Real-Time PCR System (APPLIED BIOSYSTEMS, Waltham, USA) and analysed using QuantStudio 12 K Flex Applied Biosystems software v1.2.3. Primers and probe targeting SARS-CoV-2 N gene, are: Forward: 5'-GGCCGCAAATTGCA CAAT-3'; Reverse: 5'-CCAATGCGCGACATTCC-3'; Probe: FAM-CCCCAGCGCTTCAGCGTTCT-BHQ1.

**Extraction of the RNA from clinical samples**

Total nucleic acids were extracted from 200 µL of NP samples using the QIACUBE processing system with the QIAamp 96 Virus QIACube HT Kit (Qiagen) and eluted to 100 µL of total nucleic acid. Real-time RT-PCR was performed for each sample with TaqPath COVID-19 Kit<sup>TM</sup> (Thermo Fisher Scientific). The TaqPath COVID-19 Kit<sup>TM</sup> is a multiplex real time RT-PCR diagnostic assay targeting three regions of the SARS-CoV-2 genome (N, S, and ORF1ab) which was approved by the French National Center of respiratory diseases for the detection of SARS-CoV-2. TaqPath RT-PCR<sup>TM</sup> assay is a useful tool enabling a rapid screening

of SARS-CoV-2 as an S-gene target failure was observed in variants with a deletion at positions 69–70 ( $\Delta$ HV69-70) whereas ORF1ab and N targets are correctly amplified (Bal et al., 2021).

#### Choice of the target sequence for amplification

The target of the SARS-CoV-2 Spike coding sequence ranges from nt position 1273 to 1587, including the primers, and is 315 nucleotide-long. The corresponding amino acid sequence (amino acid 425 to 529) encompasses the RBM module (Figure 1A).

#### RT-PCR protocol of the targeted sequence

RT-PCR was performed with SuperScript II One-Step Platinum Kit (#10928042, Thermo Fisher Scientific), on a Biometra T3000 thermal cycler. Primers were synthesized and provided by Thermo Fisher Scientific. The forward and reverse primer sequences are: 5'-TTACCAGATGATTTTACAGGC-3' and 5'-AGACTTTT TAGGTCCACAAAC-3', respectively. The cycling conditions were defined as follows: 50 °C for 30 min; 94 °C for 2 min; 40 cycles of 94 °C for 15 s; 55 °C for 20 s and 68 °C for 20 s. The final elongation was performed at 72 °C for 2 min.

#### Evaluation of the linearity in the designed system

To determine the linearity of the designed system, real time RT-PCR was performed with QuantiTect SYBR Green RT-PCR Kit (#204243, Qiagen), on a Bio-Rad CFX96 thermal cycler, software version 3.0 (Bio-Rad Laboratories). The cycling conditions were: 50 °C for 30 min; 95 °C for 15 min; 40 cycles of 95 °C for 15 s; 50 °C for 30 s and 72 °C for 45 s. PCR products were then loaded on a 2% agarose gel. The size of the fragment size is estimated using a molecular size ladder (DNA 1kb Plus #10787026, Thermo Fisher Scientific).

#### Production of the reference PCR products

SARS-CoV-2 BavPat1 strain was selected as the reference for this study. A large scale production of amplicons was done for this strain, following the RT-PCR protocol defined above.

#### Production of the PCR products from biological samples

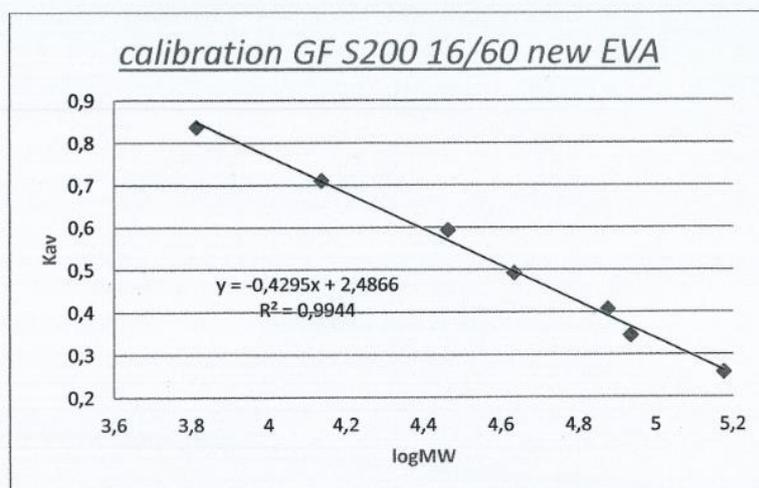
Each sample used in this study has been amplified following the RT-PCR protocol described above.

#### Mismatch-specific endonuclease assay and detection of cleavage

The presence of a mutation in the amplified target sequence was detected using the Surveyor Mutation Detection Kit (#706021, Integrated DNA Technologies). Two PCR products (a reference product and a sample to be analysed) were mixed in a 10  $\mu$ L final volume and the endonuclease mismatch specific cleavage was performed following the manufacturer recommendations. The mixture was then loaded on capillary electrophoresis system (Fragment Analyser 5200, Agilent or GXII, PerkinElmer) prior to analysis of the cleavage profile. When needed for confirmation of the presence of mutations, the PCR products were sequenced using the Sanger method with the forward and reverse primers used for the RT-PCR (Genewiz).

## Annexe 15

Courbe de calibration de la colonne de chromatographie d'exclusion de taille



*$V_0 = 46\text{mL}$  et  $V_t = 120\text{mL}$*

	ribonucléase A	carbonic anhydrase	ovalbumine	conalbumine	Dimère ova	Aprotinin	Dimère con
MW (Da)	13700	29000	43000	75000	86000	6500	150000
log MW	4,137	4,462	4,633	4,875	4,934	3,812	5,176
Ve (mL)	98.6	90	82.5	76.2	71.6	108	65.2
Kav	0,711	0,594	0,493	0,408	0,346	0,837	0,259

## DÉVELOPPEMENT DE TECHNIQUES DIAGNOSTIQUES POUR LES MALADIES VIRALES ÉMERGENTES ET RÉ-ÉMERGENTES

BARTHELEMY Margot

### RÉSUMÉ

Dans un contexte où l'émergence et la réémergence de pathogènes viraux ne cessent d'augmenter, la détection et la caractérisation de ces agents constituent un intérêt majeur de santé publique. Le développement de tests diagnostiques adaptés est donc primordial. Ils s'appuient en général sur la détection directe du matériel génétique viral, ARN ou ADN selon le type de virus. La fiabilité de ces tests doit être assurée, en utilisant des contrôles adaptés destinés à garantir le bon fonctionnement d'un test et le rendu du résultat. Lors de ce projet EPHE, nous avons développé et produit un panel de contrôles positifs utilisé à l'Unité des Virus Émergents en support à la plateforme hospitalo-universitaire de diagnostic: les Transcrits *In Vitro* (IVT) et les témoins positifs encapsidés. Ces derniers, appelés « Armored RNA » ou « Armored DNA » selon la nature de l'acide nucléique encapsidé, miment une particule infectieuse sans le caractère infectieux. Le manuscrit présente les étapes d'optimisation des contrôles IVT et ArRNA ainsi que le développement des ArDNA.

Dans ce même contexte d'émergence, au-delà de la détection des agents pathogènes, la caractérisation génétique des virus en circulation est également primordiale pour la prise de décision médicale suite à une infection virale. Le séquençage est aujourd'hui la méthode de caractérisation la plus utilisée à travers le monde, mais peut être coûteuse et peut présenter des délais d'obtention des résultats relativement longs. Nous présentons ici une méthode de génotypage par endonucléase spécifique des mésappariements (endonucléase « Surveyor® Nuclease S », commerciale) adaptée au suivi des premiers variants SARS-CoV-2. Cette méthode permet l'obtention de résultats de manière rapide, fiable et peu coûteuse. L'enzyme commerciale ayant été retirée du marché, nous avons cherché à développer notre propre endonucléase de manière recombinante, pour la caractériser fonctionnellement et évaluer son utilisation dans des tests de génotypage.

**Mots-Clés** : Diagnostic moléculaire, contrôles positifs, génotypage, virus émergents