

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LILLE

École doctorale Biologie – Santé de Lille

Pour l'obtention du grade de :

Docteur de l'Université de Lille

Thèse soutenue le 16 octobre 2023

N-glycosylation des protéines de membrane des coronavirus humains hautement pathogènes MERS-CoV et SARS-CoV-2

Centre d'infection et d'immunité de Lille (CIIL), INSERM U1019, CNRS UMR9017,
Institut Pasteur de Lille, Laboratoire de Virologie Moléculaire & Cellulaire

Directrice de thèse : **Dr. Sandrine Belouzard**

Présentée et soutenue publiquement par :

Dylan Juckel

Composition du jury :

Dr. Yann Guérardel	Président
Pr. Sophie Le Poder	Rapporteuse
Pr. Bernard Delmas	Rapporteur
Pr. François Helle	Examineur
Dr. Sandrine Belouzard	Directrice de thèse

REMERCIEMENTS

Après des pages et des pages entières d'écriture qui m'ont occupées pendant de nombreux mois, me voici revenu au tout début de ce manuscrit pour écrire ces derniers mots. Ça sent bon la fin de cette « grande aventure » !!! (Esther Martin de F. et al. 2023). Oui, la rédaction de ce manuscrit m'a appris à citer chacun de mes propos 😊

Avant tout, je souhaite remercier chaleureusement le Dr Jean Dubuisson pour m'avoir accueilli dans son laboratoire pendant presque 4 ans, pour réaliser mon stage de Master 2 et pour préparer ma thèse de doctorat.

Ensuite, je remercie grandement ma directrice de thèse, Sandrine Belouzard, pour ton encadrement tout au long de ces années. Aussi, je te remercie pour avoir toujours réussi à trouver le temps nécessaire pour discuter des manips, malgré ton emploi du temps très chargé, surtout au moment de la période chaude du COVID. Je suis aussi très reconnaissant de tout ton travail de correction de ce manuscrit, dans des délais relativement courts sur la fin, mais toujours réalisé soigneusement. Je n'oublie pas également ton aide importante à la paillasse lors des dernières manips. Enfin, grâce à toi, j'ai pu voyager pendant ma thèse, de Marseille à Atlanta aux États-Unis, avec un passage à... Cambrai en octobre... Ceci dit, j'ai pu voir de mes propres yeux Martin et Martine !

Je remercie également les membres de mes comités de suivi individuel de thèse, Pr Bernard Delmas et Dr Yann Guérardel, pour votre bienveillance lors de ces réunions. Vous avez fait de ces réunions des moments d'échanges scientifiques constructifs. Je vous remercie également de faire parti de mon jury de thèse, aux côtés du Pr François Helle et Pr Sophie Lepoder, que je remercie tout autant.

Un très grand merci à la Fondation pour la recherche médicale (FRM) qui a soutenu ce projet dans sa globalité.

Je dis un grand merci général à tous les membres du laboratoire : Muriel, Nathalie, Karin, Yves, Lowiese, Audrey, Laurence, Cécile-Marie, Claire, Virginie, Peggy, Anne et Sophana.

Je remercie également tous les étudiants du labo, Esther, Imelda, Malak, Martin, et Orféas. Force à vous, c'est bientôt la fin !!

Je voudrais remercier tout particulièrement Adeline, pour les nombreux cafés qu'on a pris ensemble, mais surtout pour tout ton travail. Tu m'as été d'une aide incontestable pendant toute ma thèse, tu y es beaucoup dans ce travail que je présente aujourd'hui. Je t'en suis extrêmement reconnaissant. Merci beaucoup !!! Tu peux désormais arrêter de rêver de « la figure 4 » ☺

Je remercie également DOCTEURE Cyrine Bentaleb (il est important le préfixe) pour tous tes conseils sur la thèse et pour ta bonne humeur au quotidien. Attention, tu devras bientôt m'appeler docteur également !!

Je remercie également d'anciens membres du laboratoire, Karoline, Charline, Kévin, Nathan, pour nos débriefs foot le lundi matin, et Thomas pour tes histoires les plus improbables les unes que les autres ☺

On en arrive au plus important ! Je remercie infiniment mes proches, ma mère et ma sœur. Vous avez toujours cru en moi, vous m'avez toujours soutenu, et sans vous, je ne serais pas là où j'en suis aujourd'hui. Maman, je suis très reconnaissant pour tous les sacrifices que tu as fait pour tes enfants, tu es une maman formidable. Je sais que tu es fier de moi, mais moi je suis encore plus fière de toi !

J'ai une grosse pensée pour mon papa. J'espère que de là où tu es, tu es fier de ce que deviennent tes enfants.

Merci à ma chérie, Hélène, qui m'as soutenue quotidiennement pendant cette thèse, et ça n'a pas tous les jours été facile... Tu as toujours réussi à trouver les mots positifs quand j'en ressentais le besoin. Tu es une personne formidable ! Je t'aime très fort !

Je garde cette dernière ligne pour ma fille, Albane, qui pointerà le bout de son petit nez quelques jours après ma soutenance (j'ai bien dit après, hein, stp ☺) . Tu seras mon plus beau cadeau, et qui sait, plus grande tu t'entraîneras peut-être à la lecture avec ma thèse.

*La science, mon garçon, est faite d'erreurs,
mais d'erreurs qu'il est bon de commettre,
car elles mènent peu à peu à la vérité.*

Jules Verne

TABLE DES MATIÈRES

ABRÉVIATIONS	1
RÉSUMÉ EN FRANÇAIS.....	4
RÉSUMÉ EN ANGLAIS	6
1 INTRODUCTION	8
1.1 <i>Les coronavirus.....</i>	8
1.2 <i>Les coronavirus humains.....</i>	10
1.2.1 <i>A pathogénicité modérée.....</i>	10
1.2.1.1 <i>HCoV-229E</i>	10
1.2.1.2 <i>HCoV-OC43</i>	11
1.2.1.3 <i>HCoV-NL63.....</i>	11
1.2.1.4 <i>HCoV-HKU1.....</i>	12
1.2.2 <i>A pathogénicité sévère.....</i>	12
1.2.2.1 <i>SARS-CoV.....</i>	12
1.2.2.2 <i>MERS-CoV</i>	13
1.2.2.3 <i>SARS-CoV-2</i>	15
1.2.2.4 <i>Les variants du SARS-CoV-2.....</i>	17
1.3 <i>Origines et transmission</i>	21
1.3.1 <i>L'origine des coronavirus</i>	21
1.3.2 <i>Les voies de transmission.....</i>	24
1.4 <i>Traitements et vaccins.....</i>	27
1.4.1 <i>Les traitements curatifs.....</i>	27
1.4.2 <i>Les vaccins.....</i>	29
1.5 <i>La biologie des coronavirus.....</i>	32
1.5.1 <i>La particule virale</i>	32
1.5.2 <i>Le génome.....</i>	33
1.5.3 <i>Les protéines non-structurales</i>	36
1.5.4 <i>Les protéines structurales.....</i>	38
1.5.4.1 <i>Protéine spike (S)</i>	38
1.5.4.2 <i>Protéine d'enveloppe (E).....</i>	40
1.5.4.3 <i>Protéine de nucléocapside (N).....</i>	42
1.5.4.4 <i>Protéine hémagglutinine-estérase (HE).....</i>	42
1.5.4.5 <i>Protéine de membrane (M).....</i>	43

1.5.5	Les protéines accessoires	47
1.6	<i>Le cycle viral</i>	50
1.6.1	Attachement à la cellule cible.....	51
1.6.2	La fusion	54
1.6.3	Réplication et transcription du génome.....	57
1.6.4	Encapsidation, bourgeonnement et sécrétion.....	58
1.6.4.1	Encapsidation du génome.....	58
1.6.4.2	Bourgeonnement	59
1.6.4.3	Sécrétion.....	62
1.7	<i>La N-glycosylation</i>	63
1.7.1	Généralités.....	63
1.7.2	La voie de biosynthèse des N-glycanes.....	65
1.7.3	Les polylactosaminoglycanes.....	68
1.7.4	Fonctions des N-glycanes chez les coronavirus.....	70
1.7.4.1	Glycanes de S	70
1.7.4.2	Glycanes de M.....	74
1.7.4.3	N-glycanes de nsp3/nsp4.....	76
2	OBJECTIFS DE THESE	78
3	MATERIELS ET METHODES	81
3.1	<i>Culture cellulaire</i>	81
3.2	<i>Plasmides</i>	81
3.3	<i>Génération des lignées KO</i>	84
3.4	<i>Traitements aux glycosidases</i>	85
3.5	<i>Précipitations par les lectines</i>	86
3.6	<i>Infection</i>	86
3.7	<i>Immunofluorescence</i>	87
3.8	<i>Perméabilisations sélectives</i>	87
4	RESULTATS	90
4.1	<i>Les protéines M du MERS-CoV et du SARS-CoV-2 sont N-glycosylées</i>	90
4.2	<i>SARS-CoV-2-M et MERS-CoV-M sont modifiées par des chaines de polylactosamines</i> .	92
4.3	<i>Les résidus acides proches du site de glycosylation sont importants pour la N-glycosylation de la protéine M</i>	97
4.4	<i>Les mutants E9D18A-M et E11,12,18A-M conservent la topologie N-exo/C-endo</i>	101
4.5	<i>Rôle du trafic intracellulaire et du premier segment transmembranaire de M pour l'ajout des chaines de polylactosamines</i>	104
4.6	<i>La glycosylation de la protéine M est dépendante du complexe OST-B</i>	109

4.7	<i>Analyse de la protéine M en cellules infectées et incorporée aux particules virales ...</i>	111
5	DISCUSSION ET PERSPECTIVES	114
	BIBLIOGRAPHIE.....	125
	ANNEXES	158

LISTE DES FIGURES

Figure 1. Répartition des cas de MERS-CoV à travers le monde	14
Figure 2. Répartition mondiale des cas confirmés de COVID-19	17
Figure 3. Localisation des mutations de la protéine S selon les variants	18
Figure 4. Circulation des variants préoccupants du SARS-CoV-2	20
Figure 5. Mécanismes de transmission des coronavirus SARS-CoV et MERS-CoV	23
Figure 6. Modes de transmission du SARS-CoV-2	26
Figure 7. Représentation schématique de la particule des coronavirus	33
Figure 8. Architecture génomique des coronavirus	35
Figure 9. Structure de la protéine spike du coronavirus MERS-CoV	39
Figure 10. Structure de la protéine M du SARS-CoV-2	44
Figure 11. Comparaison structurale des protéines M et ORF3a du SARS-CoV-2	45
Figure 12. Les étapes du cycle viral des coronavirus	50
Figure 13. Mécanismes d'attachement et de fusion du SARS-CoV-2	56
Figure 14. Représentations schématiques des N-glycanes riches en mannose, complexes et hybrides.....	64
Figure 15. Voie de glycosylation des protéines en cellules eucaryotes	67
Figure 16. Structure cryo-EM des complexes oligosaccharidiques (OST) A et B	68
Figure 17. Mécanismes de synthèse des chaînes de polylectosamines	69
Figure 18. Profil de migration de la protéine M du MERS-CoV	79
Figure 19. La protéine M du MERS-CoV et du SARS-CoV-2 est N-glycosylée	91
Figure 20. MERS-CoV-M et SARS-CoV-2-M sont modifiées par des chaînes de polylectosamines	96
Figure 21. La mutation des résidus acides de MERS-CoV-M et SARS-CoV-2-M perturbe la glycosylation	98
Figure 22. Les mutants sont localisés dans le compartiment TGN	100
Figure 23. L'extrémité N-terminale des mutants est orientée côté luminal des compartiments intracellulaires	102
Figure 24. La N-glycosylation d'une séquence linker insérée entre le second et le troisième segment transmembranaire de MERS-CoV-M n'est pas dépendante des résidus acides du domaine N-terminal	103

Figure 25. Le premier site de glycosylation de la protéine HCV-E1 est modifié par des chaînes de polylactosamines lorsqu'il est transféré sur SARS-CoV-2-M	106
Figure 26. Le premier segment transmembranaire de SARS-CoV-2-M est requis pour la glycosylation de la chimère M/ORF3a	108
Figure 27. La glycosylation de la protéine M du MERS-CoV et du SARS-CoV-2 est dépendante de la machinerie OST-B	110
Figure 28. Les protéines M du MERS-CoV et du SARS-CoV-2 sont modifiées par des chaînes de polylactosamines en cellules infectées mais cette forme est absente des particules sub-virales	113
Figure 29. Alignements de séquences en acide-aminé du domaine N-terminal de plusieurs protéines M de coronavirus modifiées par des N-glycanes	120

LISTE DES TABLEAUX

TABLEAU 1. LISTE NON-EXHAUSTIVES DES CORONAVIRUS.....	9
TABLEAU 2. OLIGONUCLEOTIDES UTILISES POUR LE PREMIER CYCLE PCR.....	88
TABLEAU 3. ANTICORPS PRIMAIRES ET SECONDAIRES UTILISES POUR LES EXPERIENCES D'IMMUNOFLUORESCENCE.....	89

ABRÉVIATIONS

ACE2 : Angiotensin-converting enzyme 2
ADN : Acide désoxyribonucléique
ADNdb : Acide désoxyribonucléique double brin
AMM : Autorisation de mise sur le marché
APN : Aminopeptidase N
ARN : Acide ribonucléique
ARNdb : Acide ribonucléique double brin
ARNg : Acide ribonucléique génomique
ARNm : Acide ribonucléique messenger
ARNss : Acide ribonucléique simple brin
ARNsg : Acide ribonucléique sous-génomique
BCoV : Bovin coronavirus
CCoV : Canine coronavirus
CEACAM1 : Carcinoembryonic antigen-cell adhesion molecule 1
CNX : Calnexine
COPI : Coatomer protein I
COVID-19 : Coronavirus disease-19
CRYO-EM : Cryo-electronic microscopy
DMEM : Dulbecco's Modified Eagle Medium
DMV : Double-membrane vesicule
DPP4 : Dipeptidyl peptidase-4
DSL : *Datura stramonium* lectin
ECoV : Equine coronavirus
Endo-H : Endoglycosidase H
ERGIC : ER-Golgi intermediate compartment
ExoN : Exoribonucléase
FIPV : Feline infectious peritonitis virus
Fuc : Fucose
Gal : Galactose
GlcNAC : N-acétylglucosamine
HCoV : Human coronavirus

HCV : Hepatitis C virus
HRSV : Human respiratory syncytial virus
IBV : Infectious bronchitis virus
IFN : Interféron
KDA : kilodalton
LEL : *Lycopersicon esculentum* lectin
Man : Mannose
MBL : Mannose-binding lectin
MERS-CoV : Middle east respiratory syndrome coronavirus
MHV : Mouse hepatitis virus
MOI : Multiplicity of infection
MVA : Modified virus Ankara
NSP : Non-structural protein
OMS : Organisation mondiale de la santé
ORF : Open reading frame
OST : Oligosaccharyltransférase
PAGE : Polyacrylamide gel
PCR : Polymerase chain reaction
PEDV : Porcine epidemic diarrhea virus
pH : Potentiel hydrogène
PNGase F : N-glycosidase F
Poly-LacNAc : Chaines de polylactosamines
RBD : Domaine de liaison au récepteur
RCoV : Rat coronavirus
RdRp : RNA-dependent RNA polymerase
RE : Reticulum endoplasmique
RLR : Rig-I-like receptor
RPM : Rotation par minute
RTC : Replicase/transcriptase complexe
RT-PCR : Real-time polymerase chain reaction
SARS-CoV : Severe acute respiratory syndrome coronavirus
SARS-CoV-2 : Severe acute respiratory syndrome coronavirus-2
SDRA : Syndrome de détresse respiratoire aigu
SDS : Sodium dodecyl sulfate

Sia : Acide sialique

SNC : Système nerveux central

SPF : Specific pathogen free

SRAS : Syndrôme respiratoire aigu sévère

TGEV : Transmissible gastroenteritis virus

TGN : *Trans*-golgi network

TM : Transmembrane segment

TMPRSS2 : Type II transmembrane serine protease

UTR : Untranslated region

VIH : Virus de l'immunodéficience humaine

VLP : Virus-like particle

VSV : Vesicular stomatitis virus

RÉSUMÉ EN FRANÇAIS

Les coronavirus sont des virus à ARN simple brin positif enveloppés d'une bicouche lipidique qui dérive de la cellule hôte et dans laquelle sont enchâssées les protéines S, E et M. La protéine M est une protéine transmembranaire de type III, composée d'un court ectodomaine et d'un long endodomaine. L'ectodomaine de M a la particularité de porter des N- ou O-glycanes, et ces modifications sont conservées au cours de l'évolution des coronavirus, suggérant un rôle dans le cycle infectieux. Cependant, la structure et la fonction de ces glycanes restent mal caractérisées. Nos travaux ont permis de mieux comprendre les mécanismes de glycosylation des protéines M du MERS-CoV et du SARS-CoV-2, deux coronavirus hautement pathogènes pour l'Homme.

Ces deux protéines contiennent un seul site de N-glycosylation dans leur domaine N-terminal et présentent un profil de migration particulier en western-blot, avec 3 bandes dont une bande très diffuse. Des tests de digestion par des N-glycosidases ont indiqué que les sites de N-glycosylation portaient des N-glycanes riches en mannose et complexes. De plus, les N-glycanes complexes de MERS-CoV-M et SARS-CoV-2-M correspondent très probablement à des chaînes de polylactosamines (poly-LacNAc).

Les mécanismes régulant l'addition des poly-LacNAc sont peu connus. Ainsi, nous avons inséré des mutations dans le domaine N-terminal de MERS-CoV-M et SARS-CoV-2-M dans le but d'identifier une séquence proche du site de N-glycosylation régulant l'ajout de ces chaînes. Cependant, nous n'avons pas identifié une telle séquence. En revanche, nous avons observé que les résidus acides du domaine N-terminal sont des éléments indispensables pour la glycosylation de M. En effet, la mutation des résidus E9 et D18 de MERS-CoV-M et E11, E12 et E18 de SARS-CoV-2-M en alanine a totalement aboli la glycosylation. Nous nous sommes assurés de la qualité du repliement protéique en analysant le trafic intracellulaire des mutants, ainsi que de leur topologie membranaire (N-exo/C-endo) par (i) des expériences de perméabilisations sélectives des membranes cellulaires et par (ii) l'insertion d'une séquence linker contenant un site de glycosylation dans la boucle reliant les TM2 et TM3 de M.

Nous avons ensuite confirmé l'absence de séquence importante pour l'ajout de chaînes de poly-LacNAc dans la protéine M en réalisant des chimères de la protéine M du SARS-CoV-2 soit avec la glycoprotéine E1 du virus de l'hépatite C soit avec la protéine accessoire ORF3a du SARS-CoV-2 dont la structure est très similaire à celle de la protéine M. L'ensemble de nos résultats suggèrent que le trafic intracellulaire dans le Golgi plutôt qu'une séquence pourrait être important pour la modification des protéines M par des chaînes de poly-LacNAc.

La glycosylation des protéines eucaryotes dépend de deux machineries moléculaires multimériques, OST-A et OST-B. Alors que STT3A est la sous-unité catalytique d'OST-A, STT3B est celle d'OST-B. La construction de lignées Huh-7 *STT3A-KO* et/ou *STT3B-KO* a permis d'identifier la machinerie OST-B comme étant impliquée dans la glycosylation de M.

Enfin, nous avons analysé les formes de M incorporées dans les particules virales à l'aide d'un système de production de particules subvirales non infectieuses. Nos résultats indiquent qu'essentiellement la forme conjuguée à des N-glycanes riches en mannose est incorporée, ce qui est en accord avec le bourgeonnement des coronavirus dans le compartiment ERGIC.

Mots clés : Coronavirus – MERS-CoV – SARS-CoV-2 – Protéine M – Glycosylation – Polylactosamines

RÉSUMÉ EN ANGLAIS

Coronaviruses are single-stranded positive RNA viruses surrounded by a host-derived lipid bilayer in which structural proteins S, E and M are anchored. The M protein is a triple-spanning transmembrane protein, constituted of a short ectodomain and a long endodomain. The ectodomain of M carries N- or O-linked carbohydrates, which are conserved through coronaviruses evolution, indicating a function for the viral life cycle. Nevertheless, structure and function of these glycans are poorly characterized. Our work brings further knowledge about glycosylation mechanisms of the M protein of MERS-CoV and SARS-CoV-2, two highly pathogenic coronaviruses for humans.

These two proteins contain a unique N-glycosylation site in their N-terminal region and show a particular migration profile in western-blot, with 3 bands amongst which one migrating as a smear. Digestion assays with N-glycosidases revealed N-glycosylation sites were modified by high-mannose and complex N-glycans. Furthermore, complex N-glycans of MERS-CoV-M and SARS-CoV-2-M could be polylactosamine chains (poly-LacNAc).

Mechanisms regulating poly-LacNAc addition are not well understood. Therefore, we inserted mutations within the N-terminal domain of MERS-CoV-M and SARS-CoV-2-M in order to identify a sequence close to N-glycosylation site that could regulate poly-LacNAc addition. However, we failed to identify such a sequence. But, we observed acidic residues of the N-terminal domain as important element for glycosylation of M. Indeed, substitution of residues E9 and D18 of MERS-CoV-M, and E11, E12 and E18 of SARS-CoV-2-M totally disrupted glycosylation. We checked mutant protein folding by analyzing intracellular trafficking, as well as their membrane topology (N-exo/C-endo) by (i) selective permeabilizations of cellular membranes and (ii) by inserting a linker sequence with N-glycosylation site in the loop tethering TM2 and TM3 of M.

Then, we confirmed absence of a sequence leading poly-LacNAc addition by generating chimera of SARS-CoV-2-M, either with glycoprotein E1 of hepatitis C virus, or with accessory protein ORF3a of SARS-CoV-2, whose structure is really close to the one of

M. Our results suggest intracellular trafficking in the Golgi is more important for poly-LacNAc addition on M than a specific sequence.

Eukaryotic proteins are glycosylated by two multimeric machineries, OST-A and OST-B. While STT3A is the catalytic subunit of OST-A, STT3B is the one of OST-B. Generation of Huh-7 *STT3A*-KO and/or *STT3B*-KO cells allowed us to identify the OST-B complex as the M protein glycosylation machinery.

Keywords: Coronaviruses – MERS-CoV – SARS-CoV-2 – M protein – Glycosylation – Polylactosamines

1 Introduction

1.1 Les coronavirus

Les coronavirus font partis de l'ordre des *Nidovirales*, de la famille des *Coronaviridae*, de la sous-famille des *Orthocoronavirinae*, qui se décompose en 4 genres : les *alphacoronavirus*, les *betacoronavirus*, les *gammacoronavirus* et les *deltacoronavirus*. Ces 4 genres se divisent en davantage de sous-genres (**Tableau 1**). Ce sont des virus principalement respiratoires et entériques, infectant un large spectre d'oiseaux et de mammifères. Les *alpha*- et *betacoronavirus* sont connus pour être des virus infectant les mammifères essentiellement, alors que les *gamma*- et *deltacoronavirus* sont majoritairement des coronavirus aviaires, avec quelques exceptions comme par exemple le coronavirus de porc HKU15 et le coronavirus du béluga SW1. Le terme coronavirus provient du mot latin « corona » signifiant couronne, en référence à l'aspect de l'enveloppe virale en microscopie électronique.

Le premier coronavirus, du genre *gammacoronavirus*, a été détecté aux Etats-Unis pendant les années 1930, connu sous le nom du virus de la bronchite infectieuse (IBV). IBV infecte les jeunes poulets et se transmet rapidement au sein d'un élevage. Aucun cas d'infection à IBV chez l'Homme n'a été recensé à ce jour. Bien d'autres coronavirus ont été découverts par la suite, comme le virus de la gastroentérite transmissible (TGEV) en 1946, un *alphacoronavirus* responsable d'épidémie de gastroentérite chez les porcelets. Le virus TGEV peut être extrêmement mortel puisqu'on estime à 100 % son taux de mortalité chez les très jeunes porcelets, lié à une forte déshydratation des animaux infectés. Trois décennies plus tard, sur le continent européen, un second *alphacoronavirus* entérique a été observé chez les porcs, nommé le virus de la diarrhée épidémique porcine (PEDV). L'émergence de variants de PEDV au cours des années 2010 a été à l'origine d'épidémies sévères en Chine sur la période 2010-2012, ainsi qu'aux États-Unis et au Mexique entre 2013 et 2014, responsables de lourdes pertes économiques, notamment en Amérique du Nord, où 10 % de la population américaine de porcs domestiques a été perdue (Jung et Saif 2015). Les souris sont également sujettes à des infections par coronavirus, puisque le virus de l'hépatite murine (MHV), un *betacoronavirus*, a été identifié par une équipe

londonienne en 1951. Des coronavirus existent aussi chez les chiens et chats, avec notamment un coronavirus félin responsable de péritonites chez les chats (FIPV), observé pour la première fois en 1970 (Ward 1970), ainsi qu'un coronavirus canin (CCoV) responsable de troubles gastro-intestinaux.

Tableau 1. Liste non-exhaustives des coronavirus.

Genre	Sous-genre	Espèce
Alpha coronavirus	Colacovirus	Coronavirus de chauve-souris CDPHE15
	Decacovirus	Coronavirus de chauve-souris HKU10
		Alphacoronavirus de chauve-souris Grand rhinolophe (HuB-2013)
	Duvinacovirus	Coronavirus humain HCoV-229E
	Luchacovirus	Coronavirus de rat
	Minacovirus	Coronavirus de vison et de furet
	Minunacovirus	Coronavirus de chauve-souris <i>Miniopterus</i> (HKU8)
	Myotacovirus	Coronavirus de chauve-souris <i>Myotis ricketti</i> Sax-2011
	Nyctacovirus	Coronavirus de chauve-souris <i>Nyctalus velutinus</i> SC-2013
	Pedacovirus	Virus de la diarrhée épidémique porcine (PEDV)
		Coronavirus de chauve-souris <i>Scotophilus</i> 512
	Rhinacovirus	Coronavirus de chauve-souris <i>Rhinolophus</i> HKU2
	Setracovirus	Coronavirus humain NL63
Coronavirus de chauve-souris proche de NL63 (BtKYNL63-9b)		
Tegacovirus	Alphacoronavirus 1 (virus félin et canin)	
	Virus de la gastroentérite transmissible (TGEV, virus de porc)	
Beta coronavirus	Embecovirus	Betacoronavirus 1 (HCoV-OC43 et coronavirus bovin)
		Coronavirus humain HKU1
		Coronavirus murin MHV
		China <i>Rattus</i> coronavirus HKU24
		Bat Hp-betacoronavirus/Zhejiang2013
	Hibecovirus	Coronavirus de hérisson 1
	Merbecovirus	MERS-CoV et virus proches (humain, dromadaire, chauve-souris)
		Coronavirus de chauve-souris <i>Pipistrellus</i> HKU5
		Coronavirus de chauve-souris <i>Tylonycteris</i> HKU4
	Nobecovirus	Coronavirus de chauve-souris roussettes (HKU9 et GCCDC1)
Sarbecovirus	SARS-CoV et SARS-CoV2 (et virus proches)	
Gamma coronavirus	Cegacovirus	Coronavirus du béluga (SW1)
	Igacovirus	Coronavirus aviaires
Delta coronavirus	Andecovirus	Coronavirus de canard (HKU20)
	Buldecovirus	Coronavirus de porc HKU15
		Coronavirus de passereaux (HKU13 et HKU16)
	Herdecovirus	Coronavirus des bihoreaux (HKU19)
Moordecovirus	Coronavirus des poules d'eau (HKU21)	

Adapté de (Juckel et al. 2020).

1.2 Les coronavirus humains

Les coronavirus ne sont pas des virus exclusifs aux animaux sauvages et domestiques, puisque sept coronavirus en capacité d'infecter l'Homme sont connus à ce jour, classés en fonction de leur degré de pathogénicité. HCoV-229E, HCoV-OC43, HCoV-NL63 et HCoV HKU1 sont classés parmi les coronavirus humains à pathogénicité modérée, en raison de la faible dangerosité des symptômes qu'ils induisent chez les patients. En revanche, le SARS-CoV, le MERS-CoV et le SARS-CoV-2 sont classés parmi les coronavirus humains hautement pathogènes, à cause des pathologies plus sévères qu'ils induisent et le taux de mortalité associé.

1.2.1 A pathogénicité modérée

1.2.1.1 *HCoV-229E*

Le coronavirus humain HCoV-229E a été découvert en 1966 à partir de prélèvements nasopharyngés de personnes atteintes de rhumes (Hamre et Procknow 1966; McIntosh et al. 1967). Il est aujourd'hui distribué globalement. Le temps d'incubation de ce virus se situe entre 2 et 5 jours avant l'apparition des premiers symptômes. Il provoque majoritairement des symptômes bénins, comme des maux de tête, des écoulements nasaux, des maux de gorge, des éternuements et de la toux, pouvant s'accompagner de fièvre (Tyrrell et al. 1993). Des cas de co-infection avec d'autres virus respiratoires comme le virus respiratoire syncytial humain (HRSV) ont été observés. Selon une étude clinique, les primo-infections à HCoV-229E ont lieu dès le plus jeune âge, avec une séroconversion moyenne estimée à 3 ans et demi (Dijkman et al. 2008). En effet, sur 42 sérums d'enfants âgés de moins de 2 mois, 24 (soit 57,1 %) étaient séropositifs au HCoV-229E, alors que les pourcentages de séropositifs diminuent globalement chez des sujets plus âgés (Shao et al. 2007). Ces données montrent alors la forte prévalence de ce virus chez les sujets très jeunes.

1.2.1.2 *HCoV-OC43*

Identifié en 1967 à partir de prélèvements nasopharyngés de patients ayant contracté un rhume, HCoV-OC43 est le second coronavirus humain décrit. Les symptômes induits sont comparables à ceux observés avec le virus HCoV-229E, décrit précédemment. Cependant, des cas de cyanoses ont été observés à Sao Paulo, au Brésil, uniquement chez des patients infectés par HCoV-OC43 et présentant des comorbidités (Cabeça et al. 2013). Également, ce virus semble avoir les capacités à atteindre le système nerveux central (SNC) de souris BALB/c, à partir du bulbe olfactif, et d'induire des encéphalites (Jacomy et al. 2006). Du virus HCoV-OC43 a aussi été détecté dans le liquide cébrospinal d'un enfant, soupçonné d'être à l'origine d'encéphalomyélites, ce qui laisse présager le potentiel neuro-invasif de ce virus (Ann Yeh et al. 2004). HCoV-OC43 est un virus saisonnier, dont le pic épidémique a lieu durant l'hiver et son temps d'incubation est estimé entre 2 et 5 jours. Une étude menée en Grèce, indique un âge moyen de primo-infection de 3 ans, ainsi que des cas de co-infection avec le virus HRSV (Papa et al. 2007). HCoV-OC43 s'est aujourd'hui répandu dans le monde entier.

1.2.1.3 *HCoV-NL63*

Le coronavirus HCoV-NL63 est le troisième coronavirus humain faiblement pathogène identifié. Il a été isolé aux Pays-Bas en 2005 chez un nourrisson de 7 mois atteint d'une bronchiolite et d'une conjonctivite, et admis à l'hôpital (Fouchier et al. 2004; van der Hoek et al. 2006). Sept autres patients âgés de 1 à 67 ans, dont au moins 5 souffrants de pneumopathies, et admis dans ce même hôpital, ont été contaminés par HCoV-NL63, démontrant le haut potentiel infectieux de ce virus. Rapidement, des cas d'infection à HCoV-NL63 ont été recensés en France, en Belgique, aux Etats-Unis, au Canada, au Japon, en Australie, en Corée du Sud et à Hong-Kong (Arden et al. 2005; Bastien et al. 2005; Chiu et al. 2005; Ebihara et al. 2005; Moës et al. 2005; Vabret et al. 2005; Han et al. 2007b; Su et al. 2016). HCoV-NL63 est répandu dans le monde entier. En Australie, 16 prélèvements nasopharyngés sur 840, provenant de 766 patients, étaient positifs à HCoV-NL63. De plus, 81 % des personnes positives présentaient des infections des voies respiratoires profondes (Arden et al. 2005). En Corée du Sud, sur 872 prélèvements de patients de moins de 16 ans, 14 ont été déclarés positifs à HCoV-NL63, dont 64,2 %

souffraient de croup et 21,4% de bronchiolites (Han et al. 2007b). Ces analyses montrent ainsi la capacité de ce virus à se propager au-delà des tissus respiratoires supérieurs et à engendrer des pathologies plus sévères, pouvant nécessiter une prise en charge médicale.

1.2.1.4 HCoV-HKU1

HCoV-HKU1 a été détecté pour la première fois en janvier 2005 à Hong Kong chez un patient de 71 ans, atteint d'une inflammation chronique des bronches, et hospitalisé suite au développement d'une pneumonie (Woo et al. 2005). Ce coronavirus a été aujourd'hui repéré au-delà du territoire de Hong-Kong et présente une distribution globale (Sloots et al. 2006; Vabret et al. 2006; Su et al. 2016). Virus circulant majoritairement pendant la période hivernale, il provoque principalement des infections des voies respiratoires supérieures, qui se manifestent par des écoulements nasaux, de la toux et de la fièvre modérée. Cependant, la présence de comorbidités chez certains patients peut être la cause d'infection des voies respiratoires profondes, induisant des pneumopathies de type pneumonies, bronchiolites et crises asthmatiques (Lau et al. 2006). Aux États-Unis, HCoV-HKU1 circule activement chez les enfants de moins de 5 ans, comme l'indiquent certaines études (Esper et al. 2006; Talbot et al. 2009).

1.2.2 A pathogénicité sévère

Le début du XXIème siècle a connu l'émergence d'une nouvelle classe de coronavirus chez l'Homme, caractérisés par leur forte pathogénicité. Trois nouveaux coronavirus ont émergé en l'espace de 20 ans. Il y figure le SARS-CoV (Severe acute respiratory syndrome coronavirus) à l'origine du syndrome respiratoire aigu sévère (SRAS), le MERS-CoV responsable du syndrome respiratoire du Moyen-Orient, et le SARS-CoV-2 responsable de la pandémie de la COVID-19 (Coronavirus disease-19).

1.2.2.1 SARS-CoV

A la fin de l'année 2002, une épidémie de pneumonies atypiques s'est déclarée dans la région du Guangdong au Sud-est de la Chine (Zhong et al. 2003), causée par un nouveau

coronavirus humain, le SARS-CoV. Il s'est répandu rapidement et localement dans un premier temps, notamment chez le personnel hospitalier. Le taux de reproduction (R_0), estimé à 2-3, a révélé le potentiel épidémique de ce virus. La mobilité des personnes, principalement par les voies aériennes, a permis au SARS-CoV de se répandre dans 37 pays du monde, avec un pic épidémique en mai 2003.

Grâce à l'absence de contagiosité des personnes infectées avant l'apparition des premiers symptômes, des mesures de quarantaine ont suffi à stopper la propagation du virus. A la fin de l'épidémie en juillet 2003, environ 8000 cas de SRAS confirmés avaient été déclarés à l'OMS (organisation mondiale de la santé), qui ont conduit au décès de 800 personnes (taux de mortalité $\sim 10\%$). L'OMS estime que ce taux de mortalité s'élève à plus de 50 % chez les personnes de 65 ans et plus.

Après un temps d'incubation se situant en moyenne entre 2 et 11 jours, les premiers symptômes se manifestent généralement par une forte fièvre ($>38\text{ }^\circ\text{C}$), des maux de tête, des courbatures et un malaise général. D'autres symptômes, comme des nausées, des vomissements et de la diarrhée ont été observés dans de plus rares cas. Entre 2 et 7 jours suivant les premiers symptômes, certains patients, souvent atteints de comorbidités, présentaient des signes d'hypoxie. Il est estimé qu'entre 20 et 30 % des personnes infectées par le SARS-CoV ont nécessité une ventilation artificielle, suite au développement de pneumonie.

1.2.2.2 *MERS-CoV*

En juin 2012, un homme âgé de 60 ans a été hospitalisé à Djeddah en Arabie Saoudite, manifestant des symptômes fiévreux depuis plusieurs jours, une toux, des expectorations et des difficultés respiratoires (Zaki et al. 2012). Des analyses par RT-PCR ont permis d'identifier un nouveau coronavirus, d'abord appelé HCoV-EMC (HCoV-Erasmus Medical Center), puis renommé MERS-CoV par le Comité International de la Taxonomie des Virus (ICTV). Cette nouvelle émergence montre que les coronavirus sont une famille de virus émergents.

Le MERS-CoV est toujours en circulation à ce jour, mais sa diffusion reste globalement restreinte au Moyen-Orient. Sur la période d'Avril 2012 à Janvier 2023, l'OMS a recensé 2 603 cas confirmés de MERS, dont 935 décès. La mortalité liée à ce coronavirus est estimée à 36 % environ. Des cas de MERS ont été détectés dans 27 pays, essentiellement chez des personnes ayant voyagé dans la Péninsule Arabique (**Figure 1**). En France, deux patients hospitalisés au CHU de Lille ont été diagnostiqués positifs au MERS-CoV en 2013.

Le temps d'incubation du MERS-CoV est estimé entre 2 et 13 jours, avec une moyenne à 5 jours avant l'apparition des symptômes. Les pathologies associées à une infection par MERS-CoV sont variées, avec des formes asymptomatiques de la maladie, bénignes ou sévères. Fréquemment, les personnes positives au MERS-CoV vont être atteintes d'une fièvre, d'essoufflements, et de toux. Dans un tiers des cas, les patients peuvent également développer des troubles gastrointestinaux. Comme pour le SARS-CoV, le MERS-CoV se propage très majoritairement au sein des tissus respiratoires profonds, déclenchant des pathologies bien plus sévères, comme des pneumonies aiguës. Comme pour beaucoup de coronavirus, le passage de formes bénignes à sévères s'effectue principalement chez des sujets avec des comorbidités ou immunodéprimés.

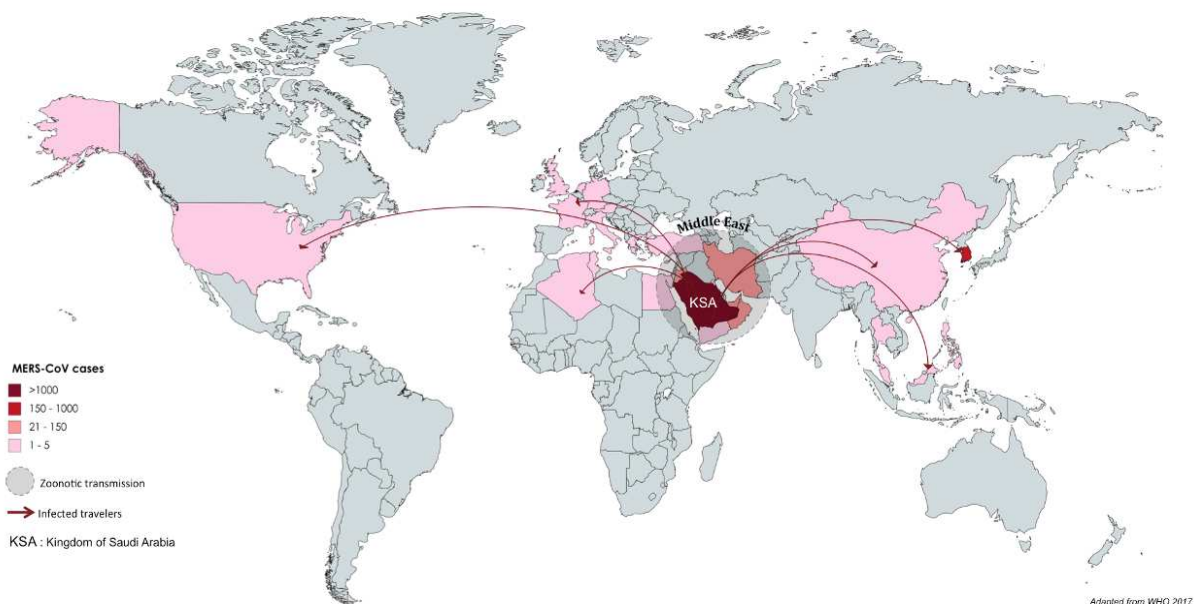


Figure 1. Répartition des cas de MERS-CoV à travers le monde.

Le nombre de cas d'infection confirmée par MERS-CoV à travers le monde en 2017 est représenté selon un gradient de couleur. Les personnes ayant contracté la maladie ont toutes voyagées dans la Péninsule Arabique les jours précédents leur retour dans leur pays d'origine (flèche rouge).

Source : OMS.

1.2.2.3 SARS-CoV-2

A la fin de l'année 2019, le dernier coronavirus humain connu à ce jour a été identifié après que plusieurs personnes atteintes de pneumonies atypiques sévères, aient été hospitalisées à Wuhan, dans la province de Hubei en Chine (Zhu et al. 2020). D'abord appelé nCoV-19 (new coronavirus-19), il a été rebaptisé SARS-CoV-2 par l'ICTV. Au niveau génomique, le SARS-CoV-2 partage environ 80 % de son génome avec celui du SARS-CoV de 2002-2003, contre seulement 50 % avec le MERS-CoV (Lu et al. 2020). Ce virus est à l'origine de la pandémie mondiale de la COVID-19, déclarée officiellement le 11 mars 2020 par l'OMS. Après plusieurs vagues épidémiques, plus de 770 000 000 de personnes ont été diagnostiquées positives à travers le monde, dont plus de 7 000 000 d'entre-elles ont succombées (chiffres actualisés au 16 août 2023). La répartition de ces cas (pour 100 000 habitants) est représentée en **Figure 2**. Il est important de préciser que ces chiffres sont vraisemblablement sous-estimés. En effet, une méta-analyse suggère que sur la période du 1^{er} janvier 2020 au 31 décembre 2021, 18,2 millions (intervalle de confiance à 95 %) de personnes à travers le monde seraient décédés directement ou indirectement (surmortalité) de la COVID-19, contre 5,94 millions officiellement déclarés sur la même période (Wang et al. 2022).

Les premiers cas de COVID-19 en France ont été identifiés le 24 janvier 2020. Aujourd'hui, plus de 39 000 000 de personnes ont été contaminées sur le territoire, dont au moins 168 000 d'entre-elles en sont décédées (chiffres actualisés au 16 août 2023). Le taux de mortalité associé au SARS-CoV-2 est difficile à déterminer avec précision. En début d'épidémie, il a été estimé à moins de 1 %.

Il a été estimé à 4-5 jours le temps d'incubation du SARS-CoV-2 avant les premiers signes cliniques. Nombreux sont les cas asymptomatiques de COVID-19, mais d'autres développeront des formes bénignes ou graves. Bien que leur fréquence et leur nature varient au cours de l'épidémie, les symptômes communs sont de la toux, de la fièvre, des maux de tête et de gorge, des myalgies, de la fatigue, des éternuements, des écoulements nasaux, etc. Des cas très fréquents de perte de goût et d'odorat (anosmie) ont été déclarés au début de l'épidémie, liés à la capacité du SARS-CoV-2 à se propager dans l'appareil sensoriel olfactif de la cavité nasale (de Melo et al. 2021). Le développement de

neuropathologies chez de nombreux patients atteints de la COVID-19 indique que le SARS-CoV-2 serait en mesure d'atteindre le SNC. Ses capacités répliquatives en cellules neuronales non-permissives ont été démontrées *in vitro*, grâce à l'utilisation de nanotubes riches en actine, reliant les cellules neuronales à des cellules Vero E6 permissives co-cultivées (Pepe et al. 2022). La présence d'ARN et de protéines virales du SARS-CoV-2 dans le cerveau de patients confirme les capacités neuro-invasives de ce virus (Meinhardt et al. 2021). *In vivo*, il est possible que le bulbe olfactif constitue une des voies d'entrée du virus vers le SNC. Cependant, d'autres mécanismes d'entrée doivent co-exister, car de l'ARN du SARS-CoV-2 a été détecté dans des régions du cerveau non-relées directement au bulbe olfactif, comme le cervelet.

Les formes graves de la COVID-19 se manifestent en général 7 jours suivant l'apparition des premiers symptômes, majoritairement chez des sujets atteints de comorbidités ou immunodéprimés. Il semblerait également que les hommes soient plus prédisposés à déclarer une forme grave de la COVID-19 (Williamson et al. 2020b). Les personnes développant une forme grave de la maladie se voient atteintes de difficultés respiratoires, évoluant vers un syndrome de détresse respiratoire aigu (SDRA), nécessitant une assistance respiratoire.

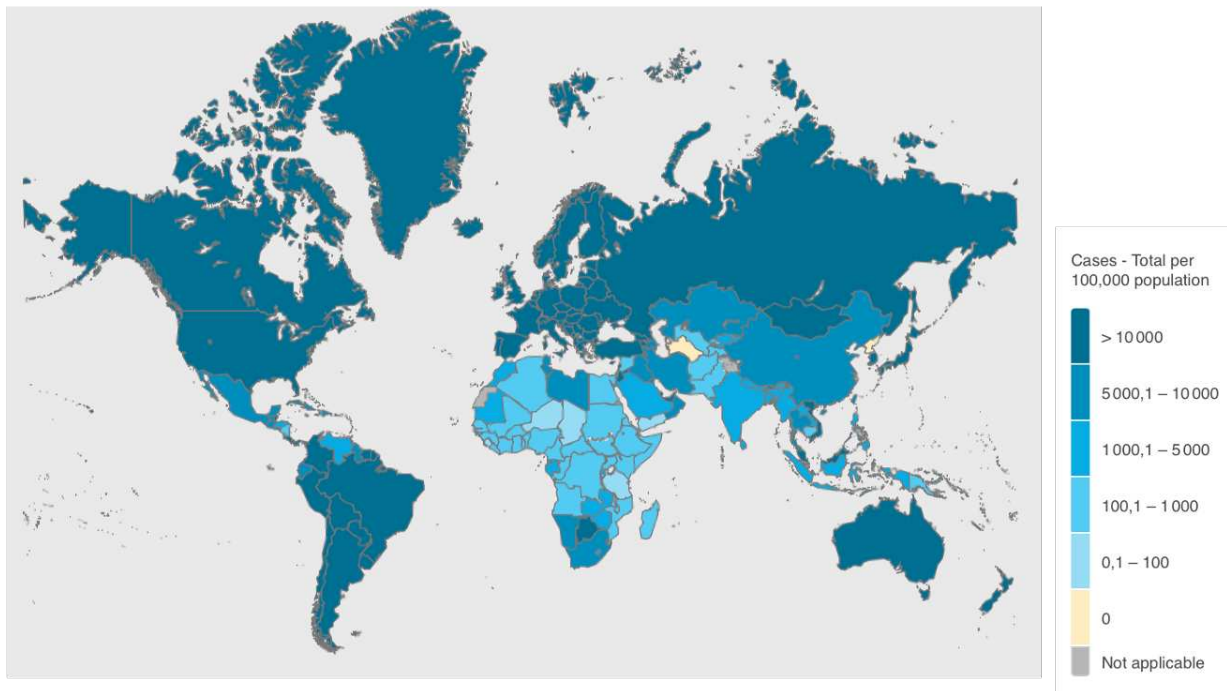


Figure 2. Répartition mondiale des cas-confirmés de COVID-19.

Les cas confirmés de COVID-19 pour 100 000 habitants sont représentés pour chaque pays selon un gradient de couleur. Données mises à jour le 16 août 2023.

Source : WHO COVID-19 Dashboard. Geneva: World Health Organization, 2020. Available online: <https://covid19.who.int/>.

1.2.2.4 Les variants du SARS-CoV-2

Les coronavirus sont des virus peu mutagènes car ils produisent une enzyme de relecture, ce qui permet la correction des erreurs lors de la réplication du génome, un caractère propre aux Nidovirus. Ainsi, les coronavirus mutent deux fois moins que les virus *Influenza* et quatre fois moins que le VIH, par exemple (Callaway 2020). Cependant, la circulation active du SARS-CoV-2 au sein de la population humaine a permis au virus d'acquérir des mutations dans son génome, favorables pour le virus, puisqu'elles ont permis d'augmenter drastiquement les capacités infectieuses du SARS-CoV-2. En effet, nombreuses de ces mutations sont localisées dans la protéine spike (**Figure 3**). L'accumulation de ces mutations a donné naissance à des souches variantes.

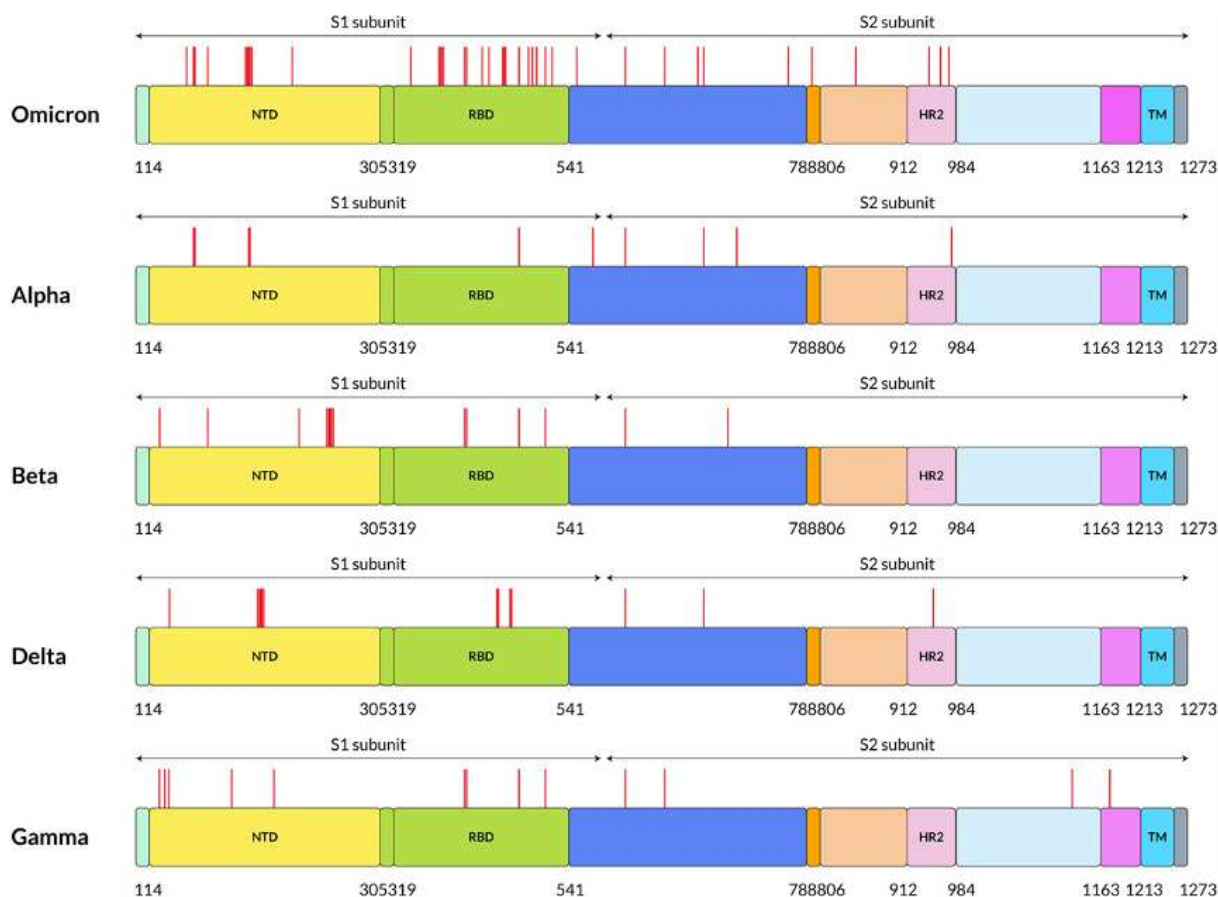


Figure 3. Localisation des mutations de la protéine S selon les variants.

Les mutations de la protéine S sont représentées par des traits rouges. Les domaines de S sont indiqués par des couleurs différentes. La majorité des mutations ont lieu dans le domaine S1, plus particulièrement dans le domaine de liaison au récepteur (RBD), notamment avec le variant Omicron. Figure adaptée de Araf *et al.*, 2022.

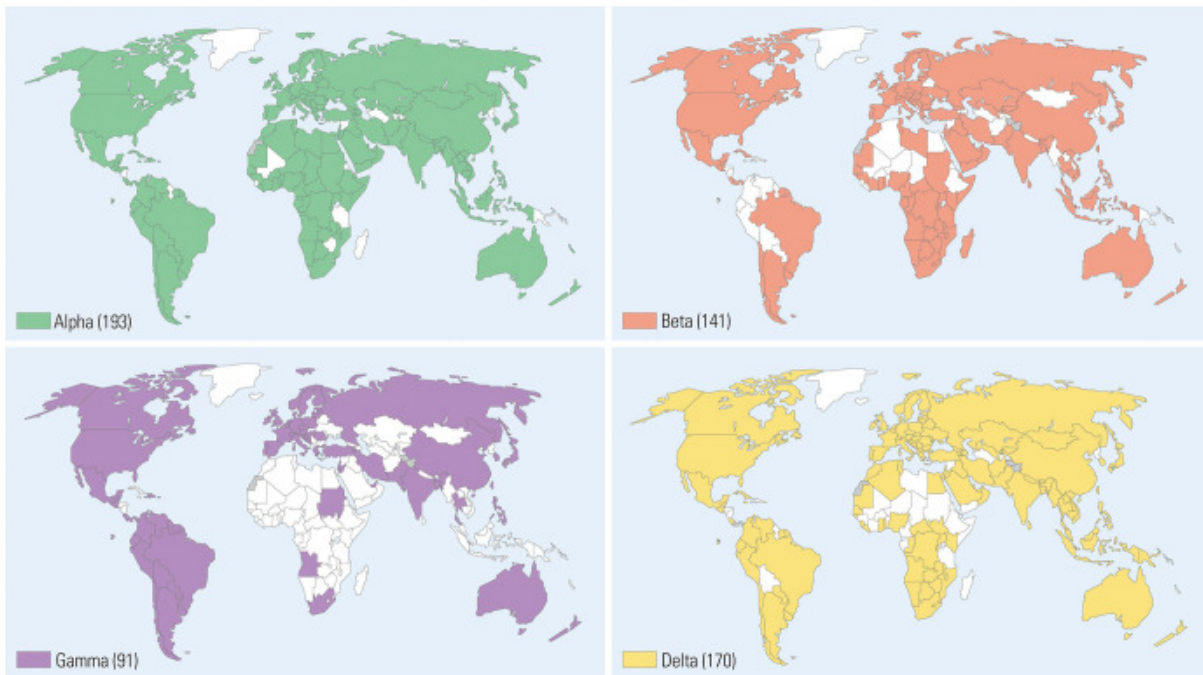
Les variants sont classés en trois catégories : les variants préoccupants (VOC, *variants of concern*), les variants d'intérêt (VOI, *variants of interest*), et les variants en cours d'évaluation (VUM, *variants under monitoring*). Depuis 2020, 5 VOC du SARS-CoV-2 apparus au cours du temps ont été recensés par l'OMS.

Le premier VOC est le variant **Alpha** (lignée B.1.1.7), détecté pour la première fois à la fin de l'année 2020 au Royaume-Uni. On dénombre 17 mutations génétiques (14 mutations non-synonymes et 3 délétions), comprenant la fameuse mutation D614G, qui augmente de façon importante les capacités infectieuses du virus (Korber et al. 2020). Ce variant s'est diffusé dans pas moins de 190 pays (**Figure 4A**) (Choi et Smith 2021). Le second VOC est le variant **Beta** (lignée B.1.351), dont le premier cas a été rapporté fin 2020 en Afrique du Sud. Ce variant présente 9 mutations dans la protéine S (L18F, D80A,

D215G, R246I, K417N, E484K, N501Y, D614G, et A701V). Selon la revue de Choi et Smith, cette souche a circulé dans 141 pays au cours de l'épidémie. Le variant **Gamma** (lignée P.1) est le troisième VOC lié à l'épidémie de COVID-19. Les premiers cas ont été rapportés au Brésil chez 4 voyageurs japonais au début de l'année 2021. Ce variant contient 10 mutations dans la protéine S (L18F, T20N, P26S, D138Y, R190S, H655Y, T1027I V1176, K417T, E484K, et N501Y), dont plusieurs sont communes avec les variants circulants précédemment. Des cas d'infection par ce variant ont été recensés dans 91 pays. Le quatrième VOC est le variant **Delta** (lignée B.1.617) qui se décompose en 3 sous-lignées et dont la principale est B.1.617.2. Ce variant a été détecté pour la première fois fin 2020-début 2021 en Inde, et est devenu très rapidement le variant dominant dans le monde. Les principales mutations dans la protéine S sont L452R, P681R, D614G, E484Q et G142D, toutes sous-lignées confondues. Notamment, la mutation L452R a été décrite comme une mutation accentuant l'infectiosité (Motozono et al. 2021). Le dernier VOC en date, et le variant **Omicron** (lignée BA.1), dont plusieurs sous-lignées ont émergé, nommés BA.2, BA.3, BA.4 et BA.5. Omicron a émergé en fin d'année 2021, et est aujourd'hui le variant dominant dans le monde (**Figure 4B**). Il contient de nombreuses mutations, dont 30 dans la protéine S (Araf et al. 2022).

Concernant les signes cliniques, peu de changements ont été rapportés pour les variants Alpha et Beta comparé à la souche originelle. En revanche, les variants Gamma et Delta se sont avérés être plus virulents que les précédents variants. En effet, des patients plus jeunes qu'auparavant étaient admis en soins intensifs. Cependant, le variant Omicron infecte davantage les voies respiratoires supérieures, et provoque donc des symptômes bénins, principalement, comme des écoulements nasaux, maux de tête, maux de gorge et éternuements (Menni et al. 2022).

A.



B.

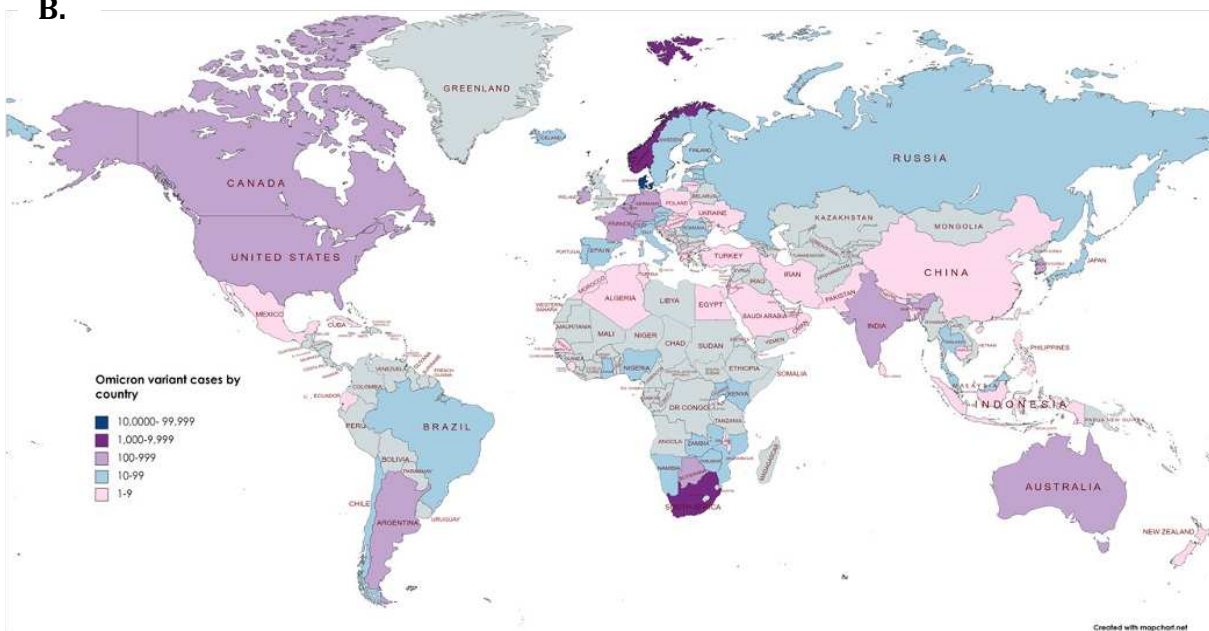


Figure 4. Circulation des variants préoccupants du SARS-CoV-2.

(A). Répartition géographique des variants Alpha, Beta, Gamma et Delta du SARS-CoV-2 au 31 août 2021. Figure adaptée de Choi et Smith, 2021.

(B). Répartition géographique du variant Omicron du SARS-CoV-2. Un code couleur représente le nombre de cas confirmés d'infection par ce variant au 15 décembre 2021. Figure adaptée de Araf *et al.*, 2022.

1.3 Origines et transmission

1.3.1 L'origine des coronavirus

Les coronavirus sont des virus zoonotiques, c'est-à-dire qu'ils sont à l'origine, des virus circulant exclusivement chez les animaux. Cependant, certains d'entre eux sont parvenus à s'adapter, leur permettant de franchir la barrière des espèces et à infecter d'autres hôtes. Nombreuses sont les études qui cherchent à percer la question de l'origine des coronavirus, notamment depuis l'émergence du SARS-CoV-2. Il est primordial de déceler les mécanismes de transmission des virus afin de diminuer le risque de nouvelles émergences ou de réémergences virales.

Les premières personnes infectées par le SARS-CoV ont toutes côtoyées dans les jours précédents les premiers symptômes, un marché d'animaux vivants dans la province du Guangdong, en Chine, favorisant l'hypothèse de l'origine zoonotique de ce virus. Au sein de ce marché, de nombreuses espèces d'animaux sauvages sont maintenues en captivité dans des conditions sanitaires déplorables. Des analyses sérologiques ont permis d'identifier des anticorps neutralisants SARS-CoV chez 80 % des civettes palmistes à masque (*Paguma larvata*) testées et vendues sur ce marché (Tu et al. 2004). La civette palmiste à masque est un mammifère carnivore de la famille des *Viverridae*, et dont la viande est consommée en Chine. L'absence d'anticorps neutralisant chez des civettes sauvages et d'élevage, suggère que ce mammifère a joué un rôle d'hôte intermédiaire dans le processus d'émergence (**Figure 5**). De plus, la souche virale circulant chez la civette (SARS-CiCoV) et chez l'humain présente >99 % d'identité nucléotidique. D'autres espèces ont également été testées positives, comme des blaireau-furets de Chine (*Melogale moschata*) et des chiens viverrins (*Nyctereutes procyonoides*). Sur ce même marché animalier y figurait des chauves-souris fer à cheval, de la famille des *Rhinolophidae*, également consommées en Chine. Des souches proches du SARS-CoV ont été isolées chez ces chauves-souris, ainsi que dans la population sauvage. Ainsi, il est possible que sur ce marché, ces chauves-souris aient contaminé des civettes, au sein desquelles, le virus s'est adapté à l'Homme.

Concernant le MERS-CoV, des anticorps neutralisants ont été observés chez des dromadaires (*Camelus dromedarius*) du Moyen-Orient, du Maghreb et d'Afrique de l'est (Reusken et al. 2013), suggérant la fonction d'hôte intermédiaire de cette espèce. Les souches de virus MERS-CoV circulant chez les dromadaires présentent une homologie de séquence >99 % avec la souche isolée chez l'Homme, ce qui conforte le rôle d'hôte intermédiaire des dromadaires (**Figure 5**) (de Wit et al. 2016). De plus, les souches circulant chez les dromadaires et l'Homme utilisent le même récepteur cellulaire (DPP4) pour pénétrer à l'intérieur de la cellule hôte. En revanche, seules les souches de MERS-CoV circulant chez les dromadaires du Moyen-Orient peuvent se transmettre à l'Homme. Ce sont des chauves-souris de la famille des *Vespertilionidae* et des *Nycteridae* qui sont les hôtes réservoirs, après l'identification de plusieurs souches similaires, comme HKU4 et HKU5, les plus proches parents du MERS-CoV sur le plan phylogénétique. De plus, des données sérologiques ont permis de dater à plus de 30 ans la transmission du virus de la chauve-souris aux dromadaires. De manière intéressante, des souches virales proches du HCoV-229E ont été détectées chez des dromadaires en Arabie Saoudite, avec une prévalence importante (Sabir et al. 2016). En revanche, il n'existe aucune donnée certifiant la transmission de ce virus à l'Homme par l'intermédiaire des dromadaires.

L'origine du SARS-CoV-2 est encore aujourd'hui largement débattue. Le marché d'animaux vivants de la ville de Wuhan est encore à ce jour considéré comme l'épicentre de la COVID-19, malgré les contestations émises par les autorités chinoises. L'étude menée par Worobey et son équipe montre que les premiers cas de COVID-19 recensés dans la ville ont tous un lien avec le marché d'animaux vivants de la ville, puisqu'ils se situaient tous dans un périmètre de 5 km environ autour du marché (Worobey et al. 2022). Des prélèvements environnementaux réalisés sur des cages, des chariots et des congélateurs du marché de Wuhan ont été effectués. Il s'avère que les prélèvements positifs au SARS-CoV-2 ont tous été réalisés dans la zone sud-ouest du marché, zone dans laquelle sont vendus des animaux vivants tels que des renards roux, des chiens viverrins et des blaireaux. Il est alors très probable que les premiers cas de contamination de l'Homme par le SARS-CoV-2 se soient déroulés dans cette zone du marché. Des souches virales proches du SARS-CoV-2 ont été isolés chez des chauves-souris fer à cheval (*Rhinolophidae*), comme par exemple le *betacoronavirus* RaTG13, qui partage 96 % de son génome avec celui du SARS-CoV-2. Cela suggère que les chauves-

souris puissent à nouveau être le réservoir naturel du virus ancestral. Cela reste encore aujourd'hui à confirmer, ainsi que les modalités de passage du virus chez l'Homme.

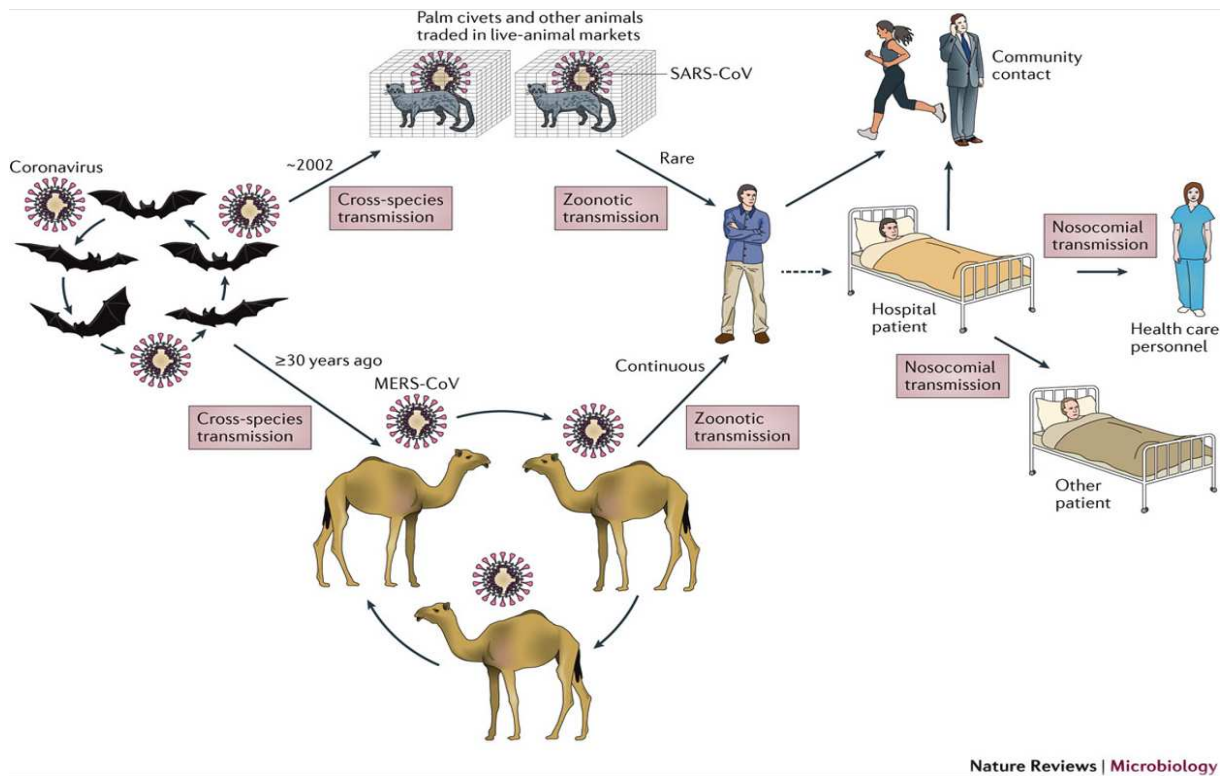


Figure 5. Mécanismes de transmission des coronavirus SARS-CoV et MERS-CoV.

Le SARS-CoV et le MERS-CoV ont très probablement émergé à partir de virus de chauves-souris de la famille des *Rhinolophidae* (SARS-CoV) ou des familles *Vespertilionidae* et *Nycteridae* (MERS-CoV). La proximité de l'hôte réservoir avec d'autres espèces animales a permis au virus le franchissement de la barrière des espèces vers un hôte intermédiaire dans lequel le virus s'est adapté pour infecter l'Homme. Les civettes palmistes à masque (*Paguma larvata*) et les dromadaires (*Camelus dromedarius*) sont les hôtes intermédiaires respectifs du SARS-CoV et du MERS-CoV. Les marchés d'animaux vivants ainsi que les élevages sont à l'origine de contacts rapprochés entre des animaux infectés et l'Homme. Ces conditions peuvent être favorables à une transmission du virus à l'Homme. Le MERS-CoV étant peu adapté aux cellules humaines, l'infection de l'Homme par ce virus se fait essentiellement par des contacts avec des dromadaires infectés ou par transmission nosocomiale. Le SARS-CoV s'est avéré être mieux adapté à l'Homme, permettant une transmission inter-humaine. Figure adaptée de De Wit *et al.*, 2016.

1.3.2 Les voies de transmission

Les coronavirus humains sont des virus respiratoires. Un des moyens de transmission s'effectue via les gouttelettes respiratoires émises par des individus infectés, lors de la toux, des éternuements ou de la parole. Ces gouttelettes, de diamètre supérieur à 5 μm , sont chargées en particules virales, et sont projetées sur une distance d'environ 1 mètre. Elles peuvent atteindre des zones d'entrée d'un individu sain, comme la bouche, le nez ou les yeux. Cependant, au regard de leur diamètre, ces gouttelettes respiratoires ne restent que peu de temps en suspension dans l'air. On estime qu'un individu sain s'expose à une contamination via ces gouttelettes respiratoires s'il reste exposé à moins d'1 mètre d'une personne malade, et sur une durée supérieure à 15 min. La transmission du virus via ces gouttelettes est également possible si elles sont manuportées.

Le SARS-CoV-2, notamment, se transmet principalement via les aérosols que nous émettons lors de la respiration (**Figure 6**) (Hu et al. 2021). Mais l'émission de ces aérosols est plus importante encore lorsque nous parlons (10 fois plus) ou crions (50 fois plus). Ils sont caractérisés par une taille plus petite que les gouttelettes décrites précédemment. En effet, leur diamètre est inférieur à 5 μm , leur permettant, ainsi, de rester plus longtemps en suspension dans l'air environnant. De plus, une étude démontre que les particules de SARS-CoV-2 présentes dans ces aérosols sont viables, puisqu'elles sont toujours en capacité d'infecter des cultures de cellules Vero-E6 (Lednicky et al. 2020). La transmission du SARS-CoV et du MERS-CoV via ces aérosols a également été étudiée. Contrairement au MERS-CoV, les particules de SARS-CoV ne possédaient plus leurs capacités infectieuses (Yu et al. 2004; Booth et al. 2005; Kim et al. 2016).

La transmission des coronavirus via les gouttelettes respiratoires ou les aérosols est possible si le virus est suffisamment bien adapté à l'Homme. Dans le cas contraire, les transmissions interhumaines sont limitées. Comparé à la pandémie de la COVID-19 provoquée par le SARS-CoV-2, les virus du SARS-CoV et du MERS-CoV n'ont été responsables que d'un nombre restreint de contamination. Pour le SARS-CoV, cela est dû à l'absence de contagiosité des personnes infectées avant l'apparition des premiers symptômes, ce qui a permis de les placer en quarantaine rapidement après leur diagnostic, limitant ainsi la propagation du virus au sein de la population. Le SARS-CoV s'est alors principalement transmis d'Homme à Homme en milieu hospitalier (transmission

nosocomiale), où se concentre des patients affaiblis immunologiquement. En revanche, pour le MERS-CoV, la faible adaptation du virus aux cellules humaines limite très fortement la transmission interhumaine. Le taux de reproduction (R_0) du MERS-CoV, c'est-à-dire le nombre moyen de personne pouvant être infectée par une personne malade, est vraisemblablement inférieur à 1. Ainsi, les chaînes de transmission interhumaines du MERS-CoV sont très courtes.

Une autre voie de transmission possible s'effectuerait par l'alimentation, notamment pour le MERS-CoV. En effet, l'OMS met en garde quant à la consommation de produits d'origine animale crus ou mal cuits, comme le lait et la viande caméline, pouvant entraîner un risque d'infection par le virus ainsi que par d'autres agents pathogènes.

L'expression du récepteur ACE2 (Angiotensin-converting enzyme 2) (utilisé par le SARS-CoV et le SARS-CoV-2) et de la protéase à sérine TMPRSS2 (Type II transmembrane serine protease) au niveau des épithéliums respiratoires et entériques, ne laisse pas présager que les coronavirus ne sont uniquement que des virus respiratoires. En effet, ils sont également en capacité de se répliquer dans le système gastro-intestinal. Plusieurs études ont permis de montrer que la transmission des coronavirus par la voie oro-fécale était possible (**Figure 6**). Par exemple, des particules virales viables de SARS-CoV et de MERS-CoV ont été retrouvées sur du matériel souillé et sur des surfaces. Aussi, de l'ARN de SARS-CoV-2 est détectable dans des selles de patients, et cela encore plusieurs jours après la fin des symptômes (Xiao et al. 2020; Zhang et al. 2020).

En conclusion, ces données permettent de mettre en évidence que les coronavirus se transmettent par voie respiratoire principalement. Cependant, d'autres données suggèrent que ces virus pourraient se transmettre par d'autres routes, comme par la voie oro-fécale. Il est donc nécessaire de considérer chacune de ces études, afin de rédiger des protocoles sanitaires les plus efficaces pour freiner la propagation virale.

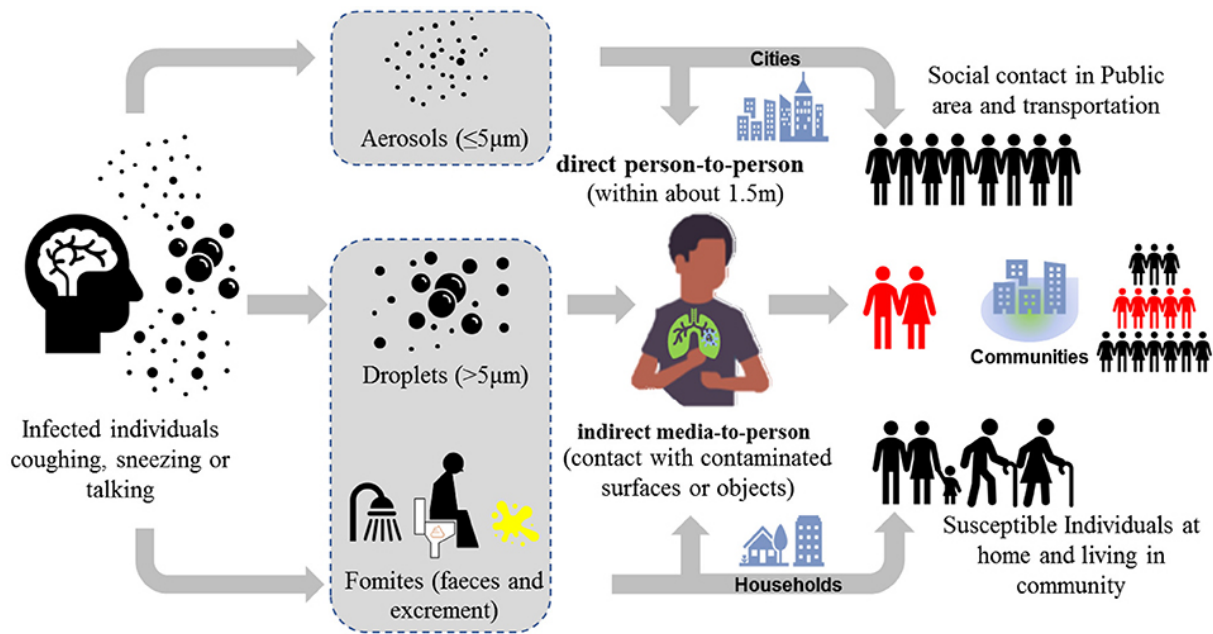


Figure 6. Modes de transmission du SARS-CoV-2.

Les coronavirus sont des virus infectant principalement les voies respiratoires supérieures et profondes. Ainsi, une personne contaminée projette des aérosols ($< 5\ \mu\text{m}$) et des gouttelettes ($> 5\ \mu\text{m}$) lors de la parole, de la toux ou des éternuements. Une transmission inter-humaine directe est possible via les aérosols grâce à leur capacité à rester en suspension dans l'air. Une transmission inter-humaine indirecte est également possible par l'intermédiaire des gouttelettes respiratoires, qui tombent rapidement au sol, contaminant des surfaces et objets. La voie oro-fécale semblerait également être un moyen de transmission du SARS-CoV-2. Figure adaptée de Hu *et al.*, 2021.

1.4 Traitements et vaccins

Il existe deux stratégies de traitement des infections. On parle (i) de traitement prophylactique quand l'intérêt est de prévenir une infection, et (ii) de traitement curatif, qui vise à guérir les patients déjà infectés. Cependant, avant la pandémie de la COVID-19, aucune thérapie approuvée contre les coronavirus n'était disponible. Néanmoins, plusieurs groupes ont cherché à mettre au point des traitements contre le SARS-CoV et le MERS-CoV. Ces études n'ont pas été sans intérêt dans le développement rapide de traitements contre le SARS-CoV-2.

1.4.1 Les traitements curatifs

Face à l'émergence soudaine du SARS-CoV en 2002, nombreuses sont les molécules à avoir été administrées aux patients, sans n'avoir jamais évalué leurs effets contre les coronavirus auparavant. C'est notamment le cas de la ribavirine, un analogue nucléosidique de la guanine. Les analogues nucléosidiques font figure d'antiviraux à large spectre, puisqu'ils visent à bloquer l'étape de réplication virale. L'activité antivirale de la ribavirine était connue *in vitro* contre plusieurs virus, comme les virus de l'herpès, les poxvirus, les adénovirus ou encore le virus respiratoire syncytial. Malheureusement, son utilisation contre l'infection à SARS-CoV a été stoppée suite à l'observation d'un certain degré de toxicité de cette molécule (hémolyse, bradycardie, taux anormalement élevé des transaminases) pour la moitié de la cohorte de patients (49 % - 71 patients) (Booth 2003). La combinaison ribavirine/corticostéroïdes s'est également montrée inefficace, avec un pic de la charge virale toujours présent à 10 jours après l'apparition des premiers signes cliniques (Peiris et al. 2003). La combinaison antirétrovirale Lopinavir/ritonavir, initialement utilisée dans le traitement contre le VIH, a démontré son efficacité à la fois *in vitro* et *in vivo* avec une diminution progressive de la charge virale, une augmentation précoce du nombre de lymphocyte et une diminution de transmissions nosocomiales. Dans autre étude, l'association Lopinavir/ritonavir avec la ribavirine et des corticostéroïdes s'est aussi avérée être efficace, avec une réduction du taux de mortalité (2,3 % contre 15,6 %) et d'intubation (0 % contre 11 %) des patients, comparé au groupe de patient ayant reçu un traitement standard (Chan et al. 2003). Bien connu pour leurs

propriétés anti-inflammatoires, les corticostéroïdes, notamment la méthylprednisolone, ont également été utilisés pour traiter les infections à SARS-CoV. De nombreux effets indésirables ont été recensés ainsi qu'une efficacité moindre, avec une charge virale plus forte, une virémie prolongée et un taux de mortalité plus important (Zumla et al. 2016). A l'inverse, la combinaison de corticostéroïdes avec des IFN α a montré une amélioration plus rapide de l'état de santé des patients, avec notamment une meilleure saturation en oxygène. De plus, le plasma de patients convalescents a lui aussi démontré son effet antiviral contre le SARS-CoV, avec une réduction de la mortalité. Cependant, ce type de traitement dépend de la disponibilité de celui-ci et de la rapidité de son administration.

La ribavirine a également fait l'objet de tests lors de l'émergence du MERS-CoV, mais aucun effet significatif a été relevé chez les patients infectés, y compris combinée avec les IFN α -2a, IFN α -2b et IFN β -1a. L'utilisation de corticostéroïdes s'est aussi montrée inefficace contre le MERS-CoV. En outre, aucune thérapie testée n'a montré quelque amélioration de l'état des patients.

Il est important de prendre en considération le faible nombre de patients admis dans les études citées ci-dessus. Afin de tirer des conclusions plus significatives, des cohortes de patients beaucoup plus larges doivent être réalisées.

La pandémie de la COVID-19 a nécessité d'accélérer les recherches de molécules antivirales réellement efficaces contre les coronavirus. Deux essais cliniques importants ont rapidement été mis en place : (i) l'essai Recovery (essai randomisé) et (ii) Solidarity, mis en place par l'OMS. Plusieurs molécules ont été testées lors de ces essais, comme la combinaison lopinavir/ritonavir, la dexaméthasone, le tocilizumab, le plasma de patients convalescents, l'hydroxychloroquine, l'azithromycine et l'IFN β -1a. Parmi toutes ces molécules, seule la dexaméthasone, un corticoïde, a permis de réduire la mortalité des patients souffrant d'une forme grave de la COVID-19, en inhibant notamment la tempête cytokinique grâce à son activité anti-inflammatoire (Ledford 2020). Le remdesivir, un analogue nucléosidique de l'adénosine et initialement testé pour traiter les cas d'infection par le virus Ébola, a prouvé son rôle antiviral contre le SARS-CoV-2 et d'autres coronavirus *in vitro* ($EC_{50} < 1 \mu\text{M}$) ainsi qu'en modèle animal (Sheahan et al. 2017; Brown et al. 2019; Williamson et al. 2020a). En revanche, certaines données issues d'essais

randomisés montrent que le remdesivir ne permet pas l'amélioration de l'état de santé des patients testés (Pardo et al. 2020). Cependant, une autre étude a tout de même observé une réduction du temps de rétablissement des patients lorsqu'ils ont reçu rapidement un traitement par le remdesivir, mais le faible nombre de patients inclus dans cette étude ne permet pas d'en tirer des conclusions significatives (Wang et al. 2020).

Outre le repositionnement moléculaire considéré dans les essais cliniques cités ci-dessus, les laboratoires Pfizer et BioNTech sont parvenus à réaliser le design moléculaire d'une molécule spécifique aux coronavirus, le nirmatrelvir, qui inhibe l'activité protéase de la 3CL^{pro} impliquée dans la maturation des polyprotéines pp1a et pp1ab formant la machinerie de réplication/transcription (voir chapitre 1.6.3). Le nirmatrelvir, associé au ritonavir, a obtenu son autorisation de mise sur le marché (AMM) en 2022 et est commercialisé sous le nom de Paxlovid®. Les données montrent une bonne tolérance des personnes traitées, avec de rares effets secondaires comme des troubles gastro-intestinaux, et une diminution des hospitalisations et des admissions en unité de soins intensifs comparé aux personnes non-traitées par le Paxlovid®.

1.4.2 Les vaccins

La vaccination reste le moyen le plus efficace pour lutter contre les maladies infectieuses. Malgré l'absence de disponibilité de vaccin contre les coronavirus avant l'émergence du SARS-CoV-2, plusieurs candidats vaccins ont été développés suite aux épidémies de SARS-CoV et de MERS-CoV, basés sur différentes technologies vaccinales. Le développement de ces candidats-vaccins anti-SARS-CoV et anti-MERS-CoV a, par ailleurs, largement bénéficié au développement de vaccins anti-SARS-CoV-2.

Des vaccins basés sur du virus inactivé ont été testés contre le SARS-CoV, MERS-CoV et le SARS-CoV-2. Ce type de vaccin est couramment utilisé, notamment contre la grippe et la polio. Des candidats vaccins développés contre le SARS-CoV et le MERS-CoV ont montré leur capacité à induire la production d'anticorps neutralisant en modèle animal, mais des cas d'éosinophilie chez des souris âgées ont été observés, ainsi qu'une faible protection après challenge (Bolles et al. 2011). Des tests ont permis de mettre en évidence que la protéine de nucléocapside (N) du virus serait à l'origine de l'éosinophilie. Dans

d'autres cas, une réponse immunitaire de type-TH2 (ex : induite dans les allergies) a été observée, indiquant que certains éléments du virus inactivé provoquent une hypersensibilité (Enjuanes et al. 2016). Le SARS-CoV-2 inactivé a été utilisé pour élaborer 3 vaccins contre la COVID-19 (BBIBP-CorV ; CoronaVac ; BBV152). BBIBP-CorV et BBV152 se sont montrés efficaces à 78 % contre les formes symptomatiques de la maladie, contre 50 % pour CoronaVac. Cependant, cette efficacité diminue face aux différents variants du SARS-CoV-2, s'expliquant par une capacité plus faible des anticorps à neutraliser la protéine spike (S) à cause d'introduction de mutations au niveau des épitopes.

Une seconde stratégie consiste à utiliser un vaccin à vecteur viral, basée sur l'insertion d'un gène d'intérêt codant une protéine virale dans le génome d'un virus peu pathogène (ex : adénovirus). Dans la très grande majorité des cas, le gène de la protéine S est utilisé. Le candidat vaccin le plus prometteur contre le MERS-CoV est un vaccin basé sur le virus de la vaccine Ankara modifié (MVA). Différentes formes de la protéine S (entière ou fragmentée) ont été testées, et une réponse cellulaire et humorale a été mesurée, quelle que soit la forme de S utilisée (Enjuanes et al. 2016). L'efficacité de ce vaccin a notamment été démontrée chez des dromadaires, dont l'excrétion de particules virales a été drastiquement réduite (1/1000) (Haagmans et al. 2016). Cette étude suggère alors qu'il serait possible de prévenir la propagation du virus au sein des réservoirs animaux de MERS-CoV. Par ailleurs, ce même vaccin a fait l'objet d'un essai clinique de phase I, dont les résultats montrent qu'une réponse cellulaire et humorale a été induite chez les personnes testées, mais cette réponse ne semble pas persister dans le temps (Koch et al. 2020). Des vaccins anti-SARS-CoV sont également en voie de développement, utilisant des adénovirus ou le virus de la rougeole comme vecteur viral, et la protéine S (entière ou fragmentée) comme protéine immunogène. Shim *et al.*, ont utilisé un adénovirus dépourvu de capacité répliquative et exprimant une forme tronquée de S (rADV-S). Après immunisation sublinguale de souris femelle BALB/c âgées de 6 semaines avec rADV-S, ils ont observé une production d'anticorps neutralisant IgG systémiques et d'IgA au niveau des muqueuses, neutralisant spécifiquement SARS-CoV-S. Ils ont également observé une réponse lymphocytaire T CD8⁺ (Shim et al. 2012). Les adénovirus ont aussi été utilisés comme vecteur pour créer une immunité contre le SARS-CoV-2. C'est le cas des vaccins AZD1222 (ChAdOx1) et Sputnik V, qui contiennent la séquence de la protéine S entière, et

Ad26.COVS2 utilisant une forme de conformation pré-fusion stabilisée de la protéine, grâce à l'insertion d'une mutation au niveau du site de clivage par la furine et de deux prolines stabilisatrices dans le domaine S2. Ces trois vaccins démontrent une efficacité correcte (81 % pour AZD1222, 66 % Ad26.COVS2 et 92 % pour Sputnik V) contre les formes symptomatiques de la COVID-19. Cette efficacité tend à diminuer selon les variants, et des cas de thromboses et de thrombocytopenies ont été recensés chez des personnes vaccinées, ayant notamment engendré la suspension de la vaccination par AZD1222 dans plusieurs pays.

Des sous-unités immunogènes d'une protéine peuvent être directement utilisées pour créer un vaccin. En effet, les vaccins à base de sous-unités protéiques ont montré leur capacité à conférer une immunité cellulaire et humorale sur des souris contre des infections à SARS-CoV et MERS-CoV. Ma C. et son équipe ont fusionné la région constante (Fc) d'IgG humain avec le RBD (résidus 377-662) de MERS-CoV-S. Après immunisation intranasale ou sous-cutanée de souris femelles BALB/c âgées de 4 à 6 semaines, ils ont observé une réponse humorale comparable entre les deux modes d'administration, mais une réponse cellulaire plus robuste après vaccination par voie intranasale (Ma et al. 2014). Une réduction de la charge virale et des cas de pneumonies ont aussi été constatées chez des macaques vaccinés contre le SARS-CoV et le MERS-CoV (Enjuanes et al. 2016). Ces types de vaccin se basent principalement sur des fragments de la protéine S, notamment sur le domaine S1 contenant le RBD, ou sur la protéine entière, qui est fortement immunogène, en faisant donc la cible majoritaire des anticorps neutralisants. Le vaccin anti-COVID-19 NVX-CoV2373 de Novavax utilise la protéine S dans son état préfusionnel et son efficacité de protection se situe vers 90 % selon les résultats de la phase III de l'essai clinique après deux injections (Heath et al. 2021). Comme pour les autres vaccins, son efficacité diminue en fonction des variants.

Expérimentés mais jamais utilisés auparavant, les vaccins à ADN/ARN s'affichent comme une nouvelle classe de vaccin. Cette méthode consiste à injecter la séquence ADN ou ARNm d'un gène codant une protéine virale immunogène. Un vaccin contenant la séquence codant la protéine S du MERS-CoV a induit chez des souris, des dromadaires et des macaques rhésus, la production d'anticorps neutralisant anti-MERS-CoV-S (Muthumani et al. 2015). Six macaques vaccinés sur 8 n'ont montré aucun signe

d'infiltration du MERS-CoV au niveau pulmonaire après challenge. Ce vaccin a fait l'objet d'un essai clinique de phase I, qui a montré sa capacité à stimuler la réponse humorale et cellulaire, avec 50 % des patients vaccinés possédant des anticorps neutralisants anti-MERS-CoV-S (Modjarrad et al. 2019). Deux vaccins, mRNA-1273 (Moderna) et BNT162b2 (Pfizer-BioNTech) basés sur la technologie ARNm sont largement utilisés contre le SARS-CoV-2. Le vaccin mRNA-1273 contient un ARNm codant la protéine SARS-CoV-2-S entière dans une conformation pré-fusion stabilisée grâce à l'ajout de deux prolines dans le domaine S2 de la protéine. Son efficacité contre les formes symptomatiques de la COVID-19 a été évaluée à 94 % et a légèrement diminué selon les variants. Le vaccin BNT162b2 utilise la protéine S entière, également stabilisée dans sa conformation pré-fusion dans le but d'augmenter son immunogénicité. L'efficacité de BNT162b2 est estimée à 95 % suite aux premiers essais cliniques (Polack et al. 2020). Dans les deux cas, des chocs anaphylactiques sans gravité ont été observés suite à la primo-injection, ainsi que des cas de myocardites chez des adolescents et jeunes adultes (Hadj Hassine 2022).

Afin de rebooster l'immunité face aux souches BA.4 et BA.5 du variant Omicron du SARS-CoV-2, des formes bivalentes des vaccins développés par Moderna et Pfizer-BioNTech sont utilisées. Ils contiennent la séquence ARNm de la protéine S des souches BA.4 et BA.5 en complément de la séquence ARNm de la protéine S de la souche originelle.

1.5 La biologie des coronavirus

1.5.1 La particule virale

Les coronavirus sont des particules pléomorphiques, bien qu'elles soient majoritairement sous forme sphérique, d'un diamètre compris entre 80 et 120 nm en moyenne. Ce sont des virus enveloppés d'une bicouche lipidique dérivée de la cellule hôte, dans laquelle trois protéines structurales sont enchâssées : la protéine spike (S), la protéine d'enveloppe (E) et la protéine de membrane (M) (**Figure 7**). Certains coronavirus, comme le coronavirus bovin (BCoV), arborent une quatrième protéine structurale : l'hémagglutinine estérase (HE). Sous l'enveloppe lipidique se trouve une nucléocapside hélicoïdale formée grâce à l'interaction entre l'ARN génomique viral et la

protéine de nucléocapside (N). Les coronavirus ont la particularité de présenter une « couronne » à leur surface, visible en coloration négative, causée par une projection en forme de massue (« club-like ») des trimères de protéine S à 17-20nm de la surface des virions (Masters 2006).

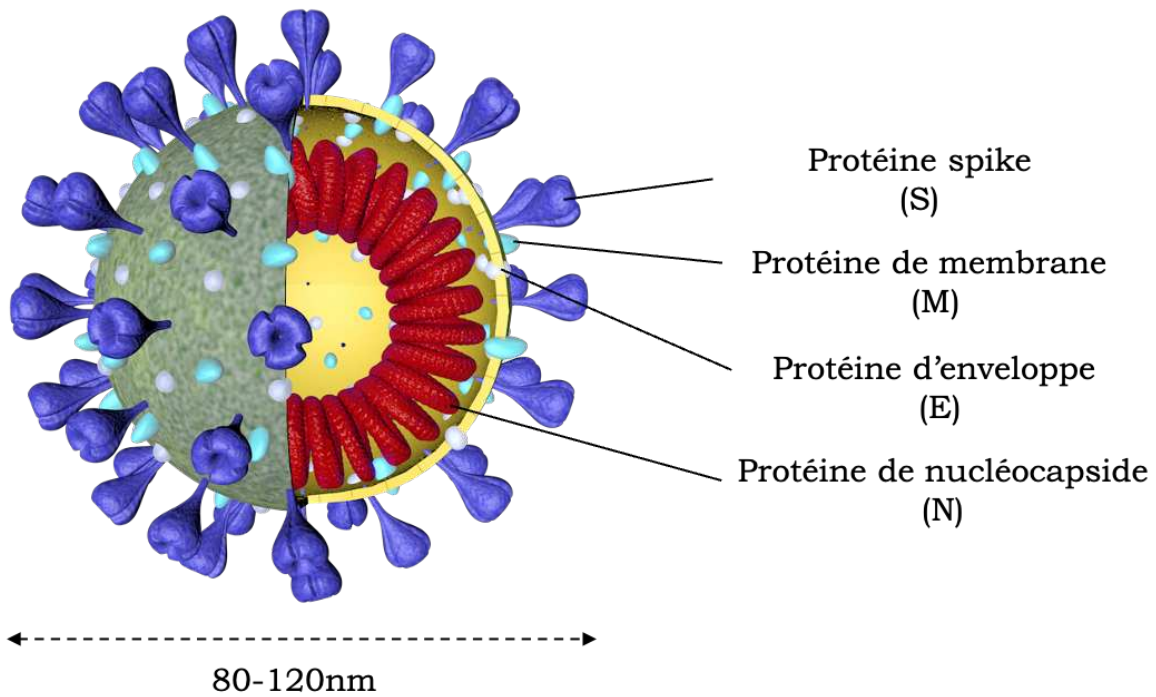


Figure 7. Représentation schématique de la particule des coronavirus.

Les coronavirus sont des virus enveloppés d'une bicouche lipidique issue des membranes de la cellule hôte dans laquelle sont enchâssées les protéines structurales spike (S), de membrane (M) et d'enveloppe (E). La protéine M est la plus protéine la plus abondante de la particule virale. La trimérisation de la protéine S induit des projections en forme de massue (« club-like ») donnant une structure en forme de couronne (« corona » en latin). La protéine de nucléocapside (N) est associée avec l'ARN génomique et forme une capsid virale hélicoïdale. Le diamètre des coronavirus est compris entre 80 et 120 nm. Figure adaptée de Juckel et al. 2020.

1.5.2 Le génome

Le génome des coronavirus est non-segmenté, simple brin et de polarité positive. Il est constitué de 27 000 à 32 000 bases en moyenne, ce qui en fait le plus long génome à ARN viral. Il présente les mêmes caractéristiques qu'un ARNm eucaryote, à savoir qu'il est coiffé à l'extrémité 5' et polyadénylé à son extrémité 3'. Comme pour tous ARN viraux de polarité positive, il a été montré que le génome des coronavirus TGEV, IBV et MHV est infectieux lorsqu'il est transfecté dans des cellules permissives (Masters 2006).

Deux cadres ouverts de lecture (ORF), ORF1a et ORF1b, représentent deux tiers du génome, et codent pour deux polyprotéines (pp1a et pp1ab) précurseurs de 16 protéines non-structurales, qui formeront la machinerie de réplication et de transcription virale après maturation (**Figure 8**). Le tiers restant porte des ORF codant les protéines structurales, séparés par un nombre variable d'ORF codant pour des protéines accessoires. L'architecture génomique est largement conservée sous la forme 5'-leader-UTR-ORF1a/b-S-E-M-N-3' UTR-poly(A). Les régions flanquantes 5' et 3' non-traduites (UTR) du génome contiennent des éléments *cis*-régulateurs nécessaires à la réplication et la transcription. Ces éléments de régulation se caractérisent par des structures secondaires de l'ARN de type tiges-boucles ainsi que par une séquence « leader » dans la région 5'-UTR. Des séquences de régulations transcriptionnelles (TRS) sont présentes en amont de chaque ORF codant pour les protéines structurales et accessoires ainsi que dans la région 5'-leader, permettant la synthèse d'ARNm sous-génomiques (ARNsg) qui codent pour ces ORF.

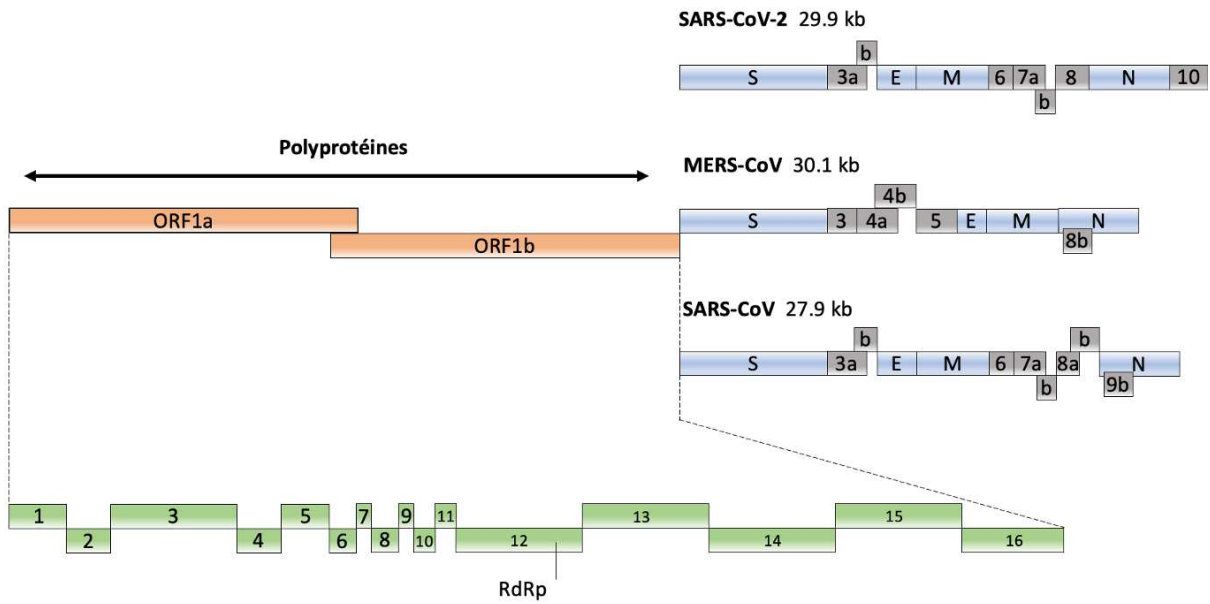


Figure 8. Architecture génomique des coronavirus.

La structure du génome est conservée chez tous les coronavirus. En orange sont représentés les gènes ORF1a et ORF1b codants pour les polyprotéines pp1a et pp1ab, précurseurs de l'ensemble des protéines non-structurales, indiquées en vert. RdRp : ARN polymérase ARN-dépendante. Ces deux cadres de lecture représentent les deux tiers du génome viral. En bleu figurent les gènes codant pour les protéines structurales S, E, M et N, espacés par un nombre variable de gènes codants pour des protéines accessoires, représentés en gris. Figure adaptée de Juckel et al. 2020.

1.5.3 Les protéines non-structurales

Les protéines non-structurales (nsp) sont codées par les gènes ORF1a et ORF1b, présents dans la région 5' du génome viral. Les coronavirus expriment 16 nsp sous la forme de deux précurseurs polyprotéiques, pp1a et pp1ab. La polyprotéine pp1a est le précurseur de nsp1-11 et pp1ab de nsp1-16. L'expression d'ORF1b, codant pour pp1ab, est permise grâce à un décalage du cadre de lecture, dont le mécanisme sera décrit plus tard. La maturation de ces deux polyprotéines est assurée par deux protéases virales, catalysant un clivage protéolytique à la jonction de chaque nsp.

La protéine **nsp1** semble favoriser la synthèse des protéines virales par la machinerie cellulaire au détriment de la synthèse des protéines de l'hôte (dégradation des ARNm), empêchant, de ce fait, l'expression des gènes impliqués dans la réponse immunitaire innée (Kamitani et al. 2006). Les fonctions de **nsp2** ne sont pas encore définies. Néanmoins, il a été démontré que son expression est dispensable pour la réplication de MHV et du SARS-CoV en culture cellulaire. Cependant, sa délétion dans le génome de ces deux virus engendre une baisse de la croissance virale (Graham et al. 2005). De plus, nsp2 interagit avec les prohibitines PHB1 et PHB2, qui s'associent sous la forme d'un complexe impliqué, entre autres, dans la biogénèse mitochondriale et la signalisation intracellulaire, laissant penser que nsp2 puisse également être impliquée dans ces mécanismes (Cornillez-Ty et al. 2009).

La protéine **nsp3** est une large protéine à multi-domaines transmembranaires. Elle est la première protéase virale (PLpro), de type protéase à sérine et semblable à la papaine. Elle est impliquée dans la maturation des nsp1-3. Elle possède également une activité déubiquitinase, un domaine Ubl1 lui permettant d'interagir avec la protéine N, et est impliquée dans la biogénèse des vésicules double-membranaires (DMV) dans lesquelles se déroulent les étapes de réplication et de transcription génomique (Angelini et al. 2013; Hurst et al. 2013; Wolff et al. 2020). Les protéines **nsp4** et **nsp6** sont également impliquées dans la production des DMV. Nsp4 semble être importante pour la stabilité et la structuration de ces vésicules, puisque sa délétion est corrélée à une morphologie aberrante des DMV en cellules infectées, ayant ainsi des conséquences sur la synthèse des ARN viraux et donc sur la croissance virale (Gadlage et al. 2010; Sakai et

al. 2017). En plus de sa fonction dans la formation des DMV, **nsp6** stimule la formation d'autophagosomes dérivés du RE. Cependant, ces autophagosomes induits par nsp6 présentent un diamètre inférieur à la normale (Cottam et al. 2014). Il est alors possible que nsp6 régule la voie d'autophagie au profit de la réplication virale. La protéine **nsp5** est la seconde protéase virale (3Cl^{pro} ou M^{pro}), faisant partie de la famille des protéases à cystéine. Elle est impliquée dans la maturation des nsp4-16.

Les protéines **nsp7** et **nsp8** forment un complexe hexadécamérique avec une activité ARN polymérase (te Velthuis et al. 2012). Ces deux protéines agissent en tant que cofacteur (1 copie de nsp7 et 2 copies de nsp8) de **nsp12** qui porte l'activité ARN polymérase ARN-dépendante (RdRp) principale. L'association nsp7-nsp8 semble augmenter la processivité de nsp12 (Subissi et al. 2014). La protéine **nsp14** est également associée au complexe nsp7-nsp8-nsp12, assurant la fonction de relecture du génome via son activité 3'-5' exoribonucléase (ExoN). Elle est aussi impliquée dans le coiffage m⁷G de l'extrémité 5' du génome par son activité N7-méthyltransférase (N7-MTase), permettant de protéger le génome des 5'-3' exoribonucléases cellulaires. La protection du génome est également assurée par **nsp16** par son activité 2'-O-méthyltransférase (2'-O-MTase). Cette modification de l'ARN permet de leurrer le système immunitaire en faisant passer le génome pour une molécule du « soi », empêchant sa reconnaissance par le senseur aux ARN double-brin (ARNdb) MDA5 (Züst et al. 2011). **Nsp15** est une endoribonucléase et agit également sur les senseurs à ARNdb en inhibant leur activation (Deng et al. 2017). Cette protéine co-localise avec les facteurs du complexe réplication/transcription (RTC). **Nsp10** est décrite comme un cofacteur de nsp14 et nsp16 puisqu'elle stimulerait les activités ExoN et 2'-O-MTase (Bouvet et al. 2010; Subissi et al. 2014). **Nsp13** est une hélicase à ARN, permettant de dérouler les intermédiaires ARNdb lors de la réplication/transcription du génome et porte également une activité 5'-triphosphatase. La protéine **nsp9** est une protéine de liaison à l'ADN et ARN simple brin (ADN/ARNss) et est présente uniquement chez les coronavirus (Egloff et al. 2004). **Nsp11** est un peptide de très petite taille (13 résidus), n'ayant cependant jamais été détecté en cellule infectée.

Pendant longtemps, la résolution structurale du RTC est restée un défi. C'est seulement qu'à partir de 2019, que la première structure a été publiée (Kirchdoerfer et Ward 2019). Depuis, plusieurs structures de RTC ont été résolues, notamment du SARS-

CoV-2 (Hillen 2021). Ces données structurales ont permis d'identifier les différentes sous-unités qui forment ce complexe. Ainsi, l'hétérodimère nsp7-8, et les protéines nsp9, 10, 12, 13, 14 et 16 sont des sous-unités du RTC.

1.5.4 Les protéines structurales

1.5.4.1 Protéine spike (S)

La protéine spike (S) est une glycoprotéine de type I dont la masse moléculaire varie entre 180 et 200 kDa. Elle se compose d'un long ectodomaine extracellulaire ou exposé à la surface des virions, d'un domaine transmembranaire et d'un court endodomaine intracellulaire ou orienté vers l'intérieur des virions, et s'associe sous forme d'homotrimère à la surface des virions (Delmas et Laude 1990). C'est une protéine largement glycosylée, présentant entre 21 et 35 sites potentiels de N-glycosylation (Belouzard et al. 2012). La protéine S est impliquée dans l'entrée virale en permettant : (i) la liaison du récepteur situé à la surface de la cellule hôte et (ii) la fusion de l'enveloppe virale avec une membrane cellulaire grâce à d'importants changements de sa conformation.

La protéine spike possède deux domaines fonctionnels (**Figure 9A**). Le premier domaine, nommé S1, se situe dans la région N-terminale et est important pour le processus d'entrée du virus puisqu'il contient le domaine de liaison au récepteur (RBD). Ce domaine S1 se décompose en un sous-domaine N-terminal (NTD) et C-terminal (CTD), et est la région la plus variable de la protéine. En effet, la très grande majorité des mutations de la protéine S des variants du SARS-CoV-2 sont retrouvées dans ce domaine, dont beaucoup sont identifiées dans le RBD (**Figure 3**). Cela a donc des conséquences directes sur l'affinité RBD/récepteur et donc sur l'infectiosité virale. Le second domaine est le domaine S2, qui est la région la plus conservée de la protéine. Il est important pour le processus de fusion virus-cellule et se décompose en multiple sous-domaines : un peptide de fusion (FP), deux « heptad repeat » (HR1 et HR2), une hélice centrale (CH), un domaine de connexion (CD) et un domaine transmembranaire (TM). Les sous-domaines HR1 et HR2, caractéristiques des protéines de fusion de classe I, correspondent à des

répétitions en tandem d'un heptapeptide contenant deux résidus hydrophobes, en position 1 et 4 de la séquence.

Les structures tri-dimensionnelles des protéines S des coronavirus MHV, IBV, HCoV-HKU1, HCoV-NL63, SARS-CoV, MERS-CoV et SARS-CoV-2 ont été résolues par cryo-EM (Kirchdoerfer et al. 2016; Walls et al. 2016b, 2016a; Yuan et al. 2017; Shang et al. 2018; Zhang et al. 2021). Celle de la protéine S du MERS-CoV est représentée en **Figure 9B**. La comparaison des structures 3D montre que l'architecture de la protéine S est conservée chez les différents genres de coronavirus.

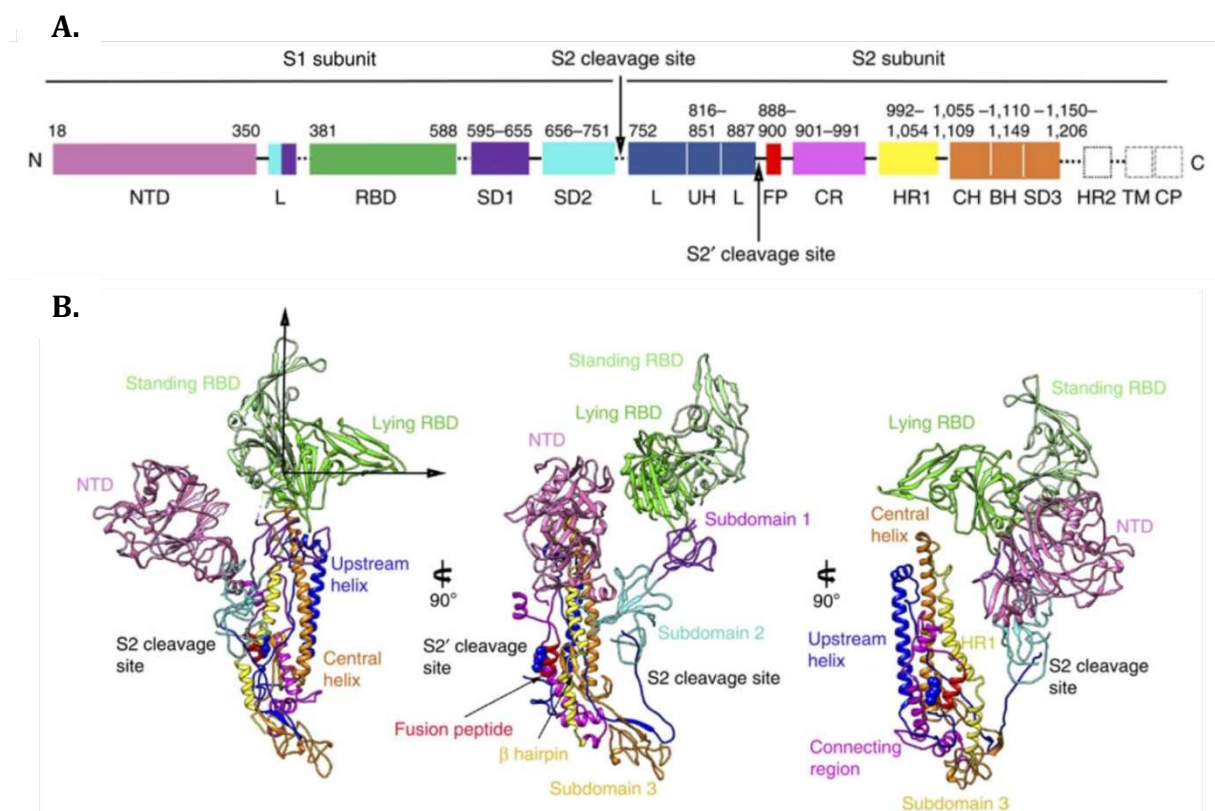


Figure 9. Structure de la protéine spike du coronavirus MERS-CoV.

(A). Représentation schématique des différents domaines de S. Les sites de clivages en S1/S2 et S2' sont représentés par des flèches. (B). Représentation en mode ruban de l'organisation structurale de la protéine S du MERS-CoV résolue par cryo-EM. Trois orientations de la protéine sont représentées. Les couleurs utilisées correspondent aux différents domaines décrits en (A). Figure adaptée de Yuan et al. 2017.

La protéine S peut se distinguer chez certains coronavirus par sa capacité à être clivée à la jonction entre les domaines S1 et S2 (S1/S2). En effet certains coronavirus, comme le MERS-CoV, IBV ou le SARS-CoV-2 contiennent un site de clivage par la furine au niveau de cette jonction.

1.5.4.2 Protéine d'enveloppe (E)

La protéine d'enveloppe (E) des coronavirus est une petite protéine de 75 acides-aminés environ, de poids moléculaire compris entre 8,4 et 12 kDa. Elle se compose d'un court domaine N-terminal hydrophile, d'un domaine transmembranaire et d'une longue queue C-terminale qui constitue la majorité de la taille de la protéine. C'est la protéine la moins abondante de la particule virale et présente de nombreuses variabilités au niveau de sa séquence en acides-aminés (Yadav et al. 2021). Néanmoins, seulement 4 positions varient entre la protéine E du SARS-CoV et celle du SARS-CoV-2, situées dans la région C-terminale.

La protéine E fait partie des protéines essentielles pour le processus d'assemblage des particules virales. En effet, la co-expression, à minima, des protéines E et M dans des cultures de cellules permet le bourgeonnement et la sécrétion de particules sub-virales non-infectieuses (ou VLP pour virus-like particles) (Vennema et al. 1996; Corse et Machamer 2000). Cependant, quelques études suggèrent que la protéine E peut être dispensable pour l'assemblage des particules virales. En effet, la délétion du gène codant la protéine E (ΔE) n'empêche pas la production de particules de MHV et de SARS-CoV, visibles en microscopie électronique. Cependant, la charge virale de MHV- ΔE et SARS-CoV- ΔE peut être jusqu'à 200 fois inférieure aux conditions sauvages, démontrant tout de même l'importance de cette protéine pour la croissance virale (Kuo et Masters 2003; DeDiego et al. 2007). Ces observations ne se sont pas confirmées avec la protéine TGEV-E, puisque sa délétion ne permet que la production de particules virales TGEV- ΔE immatures, non-infectieuses et non-sécrétées (Ortego et al. 2007). De plus, la délétion de la protéine E bloque le trafic intracellulaire des particules TGEV au niveau du compartiment ERGIC (ER-Golgi intermediate compartment).

Une étude a démontré qu'il existait deux populations de protéine E en cellules transfectées et infectées. Ces deux populations se caractérisent par des états d'oligomérisations différents de la protéine, soit sous forme monomérique ou homopentamérique/hexamérique (Westerbeck et Machamer 2015). De plus, la protéine E assurerait des fonctions différentes selon son état oligomérique. En effet, il semblerait que l'oligomérisation de la protéine soit indispensable pour assurer ses fonctions dans les étapes d'assemblages. De plus, la palmitoylation de trois résidus cystéines (C₄₀, C₄₄, C₄₇) situés dans la région C-terminale de la protéine est également nécessaire pour l'assemblage des particules (Boscarino et al. 2008). En effet, la substitution de ces résidus cystéines en alanine ou glycine abroge la production de VLP, bien que le trafic intracellulaire de E n'ait pas été impacté. L'oligomérisation de la protéine E lui permettrait également d'assurer une fonction de viroporine. Les viroporines sont des protéines transmembranaires qui forment un pore permettant le passage sélectif d'ions à travers les membranes de la cellule hôte. L'activité viroporine de la protéine E permet un passage sélectif des cations monovalents Na⁺ et K⁺. Il a aussi été observé que cette activité canal ionique permet de moduler le pH du compartiment ERGIC, lorsque la protéine E est retenue de façon intracellulaire (Cabrera-Garcia et al. 2021). En revanche, sous forme monomérique, la protéine E perturberait la voie de sécrétion, en inhibant le transport des protéines cargo vers la surface cellulaire ainsi qu'en déstructurant l'appareil de Golgi. De plus, cette activité permettrait à la protéine E de réguler le trafic intracellulaire de la protéine S, en la retenant au niveau du compartiment ERGIC (Ruch et Machamer 2011; Boson et al. 2021).

Enfin, la protéine E jouerait également un rôle dans la modulation de la réponse antivirale de l'hôte. En effet, l'infection de cellules Vero-E6 et MA-104 par du virus SARS-CoV recombinant déficient en protéine E (rSARS-CoV-ΔE) induit (i) la surexpression d'au moins 25 gènes du stress, (ii) réduit la voie IRE-1 mais pas les voies PERK et ATF-6 de la réponse UPR (*unfolded protein response*) et (iii) augmente la mort cellulaire par apoptose (DeDiego et al. 2011).

1.5.4.3 *Protéine de nucléocapside (N)*

La protéine de nucléocapside (N) est une phosphoprotéine de 50 kDa dont la fonction principale est d'interagir avec l'ARN génomique viral, formant la capsidale hélicoïdale des coronavirus. Elle est constituée de trois domaines conservés, espacés par des régions très variables (NTD et CTD). L'analyse du potentiel électrostatique de la surface de la protéine indique que les domaines NTD et CTD sont riches en acides-aminés basiques, laissant suggérer que l'interaction entre la protéine N et l'ARN viral s'établit au niveau de ces domaines (Bai et al. 2021). Au-delà de sa fonction de liaison à l'ARN, la protéine N semble jouer un rôle d'inhibiteur de la réponse antivirale déclenchée par la cellule hôte. En effet, une inactivation de la voie ARN interférence est observée en présence de la protéine N du SARS-CoV. Cependant, cette inactivation est moins importante avec la protéine N du MERS-CoV (Cui et al. 2015). Il semble également que cette protéine agisse comme un antagoniste de la voie IFN en perturbant la phosphorylation et la translocation nucléaire des facteurs de transcription STAT1 et STAT2 (Bai et al. 2021). Enfin, il a été montré que la protéine N interagit avec nsp3, et que cette interaction permet le recrutement de N au sein du complexe RTC (Hurst et al. 2013; Cong et al. 2020). De plus, cette association N-nsp3 permet de stimuler la synthèse des ARN viraux.

1.5.4.4 *Protéine hémagglutinine-estérase (HE)*

La protéine hémagglutinine-estérase (HE) est présente à la surface de certains coronavirus seulement. Elle a été identifiée pour la première fois chez BCoV et HCoV-OC43 et était semblable à celle du virus grippal de type C (Masters 2006). C'est une protéine de 48 kDa allant jusqu'à 65 kDa après l'ajout de N-glycosylations, et s'associe en homodimère à la surface des virions, formant des petites projections semblables à celles formées par les trimères de protéine S. Elle possède un long ectodomaine N-terminal avec un peptide signal, suivi d'un domaine transmembranaire et d'un court endodomaine C-terminal. Son activité hémagglutinine lui permet de s'attacher aux acides sialiques exposés par les glycoprotéines à la surface de la cellule hôte. Une activité acétylestérase a aussi été observée, permettant le clivage des fonctions acétyl des acides 9-O-neuraminiques (pour la protéine HE de BCoV, HCoV-OC43, ECoV et MHV souche DVIM)

ou 4-O-neuraminiques (pour la protéine HE de RCoV et MHV souches S et JHM) (Masters 2006).

1.5.4.5 Protéine de membrane (M)

La protéine de membrane (M) des coronavirus est une protéine transmembranaire de type III composée d'environ 220 acides-aminés, et dont le poids moléculaire varie entre 20 et 35 kDa en fonction de son état de glycosylation (voir chapitre 2.3.2). Elle est constituée d'un court ectodomaine, N- ou O-glycosylé, orienté vers la lumière des compartiments intracellulaires ou à la surface du virus, de 3 passages transmembranaires prolongés par un long endodomaine orienté côté cytoplasmique ou à l'intérieur du virus, et faisant plus de la moitié de la taille totale de la protéine. Cependant, il a été décrit pour la protéine M du coronavirus TGEV, que la formation d'un quatrième passage transmembranaire était possible, exposant la queue C-terminale côté luminal ou à la surface des virions (Risco et al. 1995). La présence d'une région amphipathique dans le domaine C-terminal de la protéine pourrait être à l'origine de la formation de ce quatrième passage transmembranaire.

La structure tri-dimensionnelle de la protéine M du SARS-CoV-2 a été en grande partie, récemment, résolue par cryo-EM (Dolan et al. 2022; Zhang et al. 2022) (**Figure 10A-C**). Ces données montrent que les 3 passages transmembranaires (TM1-3) sont organisés en hélices- α et l'endodomaine en 8 feuillets- β antiparallèles. Les extrémités N-terminales et C-terminales n'ont pas été résolues, ce qui suggère qu'elles ne seraient pas structurées. En revanche, la taille des différents domaines de M varie en fonction des structures publiées par les équipes de Dolan et de Zhang. Selon la structure résolue par Dolan et ses collègues, les TM1 à TM3 sont composés des résidus 17-36, 43-71 et 79-105, alors que le domaine N-terminal comprend les résidus 1-16, et le domaine C-terminal, les résidus 205-222. En revanche, selon Zhang *et al.*, les résidus 1-8 forment le domaine N-terminal de la protéine, suivis des résidus 9-105 constituant les hélices- α des TM1-3. Ensuite, une région charnière formée par une structure coude-hélice-coude, et composée des résidus 106-116, relie les segments transmembranaires aux feuillets- β du domaine C-terminal, constitués des résidus 117-201. Enfin, les résidus 207-222 forment la queue C-terminale de M. La formation d'homodimères rigides de 50 kDa de protéine M a

notamment pu être confirmée. Vue du dessus, les domaines TM1-3 s'arrangent sous la forme d'une ellipse aplatie. Les domaines TM1 et TM2 de chaque monomère sont distants l'un de l'autre, contrairement aux TM2 et TM3 qui se juxtaposent fortement. Dans le contexte des homodimères, le TM1 d'un monomère forme de larges interactions avec les TM2 et TM3 du second monomère. Les feuillets- β à la surface de la structure monomérique sont composés des brins $\beta 1$, $\beta 2$, $\beta 6$, la moitié N-terminale de $\beta 7$ et la moitié C-terminale de $\beta 8$. En revanche, les feuillets- β enfouilles dans la structure sont constitués des brins $\beta 3$, $\beta 4$, $\beta 5$, la moitié N-terminale de $\beta 8$ et la moitié C-terminale de $\beta 7$. Ils interagissent sur une large surface d'environ 690 \AA^2 avec les résidus L138, V139, V143, L145, F193 et A195 via des interactions hydrophobes et polaires.

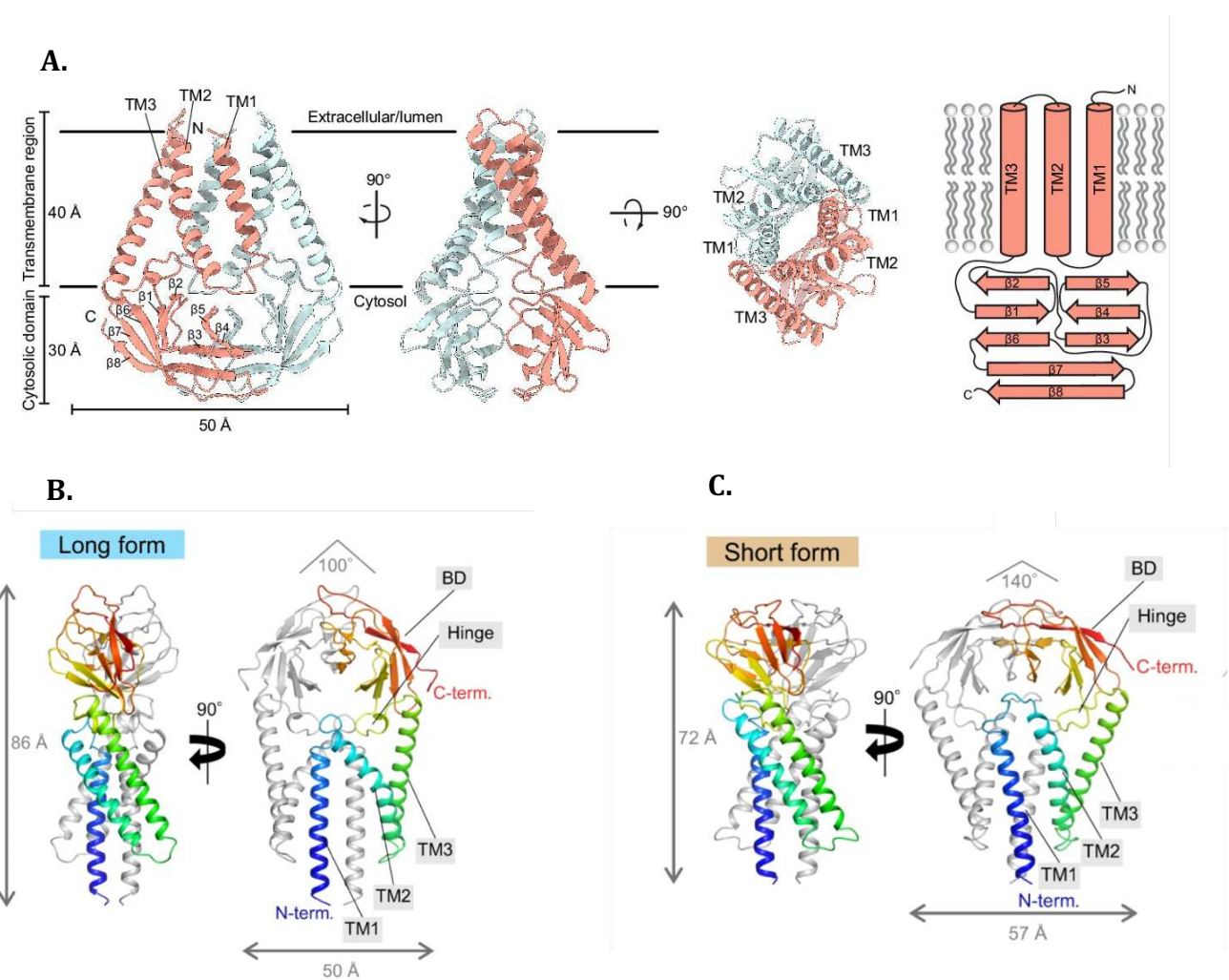


Figure 10. Structures de la protéine M du SARS-CoV-2.

Représentations en mode ruban des structures de la protéine M du SARS-CoV-2 obtenues par cryo-EM. (A). Structure publiée par Dolan et al. 2022. Trois angles de vue sont montrés. Une représentation en mode cartoon est insérée à droite du panel A. (B-C). Structures de SARS-CoV-2-M en conformation allongée (B) et compactée (C) publiées par Zhang et al. 2022.

Etonnamment, la structure tri-dimensionnelle de SARS-CoV-2-M présentent de très fortes homologies avec celle de la protéine accessoires ORF3A, décrite dans le chapitre 1.5.5, malgré seulement 20 % de similarité des séquences (Kern et al 2021). Il est possible que ces deux protéines aient une origine commune (**Figure 11**).

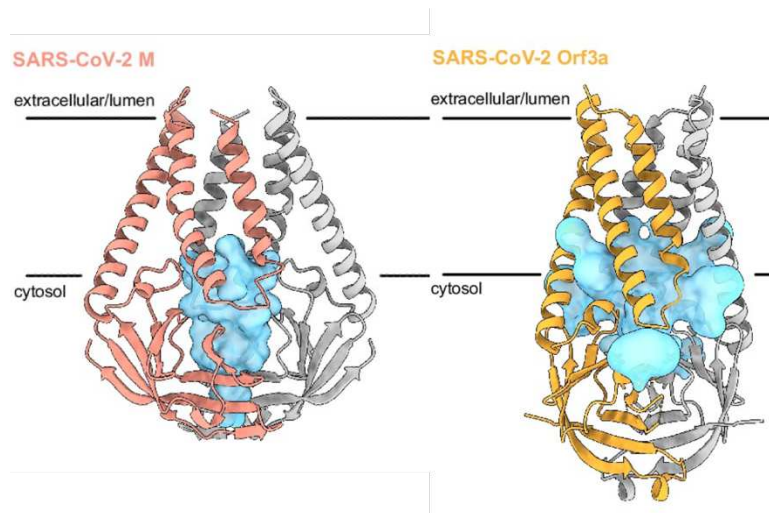


Figure 11. Comparaison structurale des protéines M et ORF3a du SARS-CoV-2.

Représentation en mode ruban des structures homodimériques des protéines M et ORF3a. Chaque monomère est indiqué par une couleur différente. En bleu figure la cavité polaire formée par l'oligomérisation des protéines. La structure de ORF3a provient de l'étude de Kern et al. 2021. Figure adaptée de Dolan et al. 2022.

Les données de cryo-EM montrent que les homodimères d'ORF3a forment une poche aqueuse, ouverte côté cytoplasmique et transmembranaire, favorable au passage des cations (Kern et al. 2021). En revanche, selon Miller et son équipe, ce pore serait trop étroit pour permettre le passage des ions. Également, le pore serait composé essentiellement de résidus chargés positivement, ce qui ne le rendrait pas perméable aux cations (Miller et al. 2023).

La protéine M a été caractérisée comme étant la protéine fondamentale pour le processus d'assemblage des coronavirus. En effet, elle est en mesure d'effectuer des interactions intermoléculaires avec toutes les protéines structurales, y compris avec elle-même (Masters 2006). De plus, cette protéine est indispensable pour synthétiser des VLP (Vennema et al. 1996; Siu et al. 2008). Les VLP sont des particules sub-virales non-infectieuses. En présence de la protéine spike, leur apparence en microscopie électronique est similaire à celle des coronavirus. Pour le MERS-CoV et le SARS-CoV-2, la

production de telles particules est possible en co-exprimant au minimum les protéines M et E dans des cultures de cellule. Dans certain cas, la protéine N peut être nécessaire, comme pour la production de VLP de type SARS-CoV (Siu et al. 2008).

Il a été observé que l'endodomaine de la protéine M du MERS-CoV contient des motifs importants, tel que la séquence $_{199}\text{KxGxYR}_{204}$ pour le trafic intracellulaire et l'assemblage des nouveaux virions (Perrier et al. 2019). La mutation de cette séquence $_{199}\text{KxGxYR}_{204}$ en alanines induit (i) une relocalisation de la protéine depuis le compartiment TGN (*Trans-golgi network*) vers la surface cellulaire, et (ii) une incapacité à synthétiser des VLP ainsi que des particules virales infectieuses (manuscrit en préparation). Un second motif a également été identifié dans la partie C-terminale de MERS-CoV-M. Il s'agit du motif di-acidique $_{211}\text{DxE}_{213}$, déjà identifié chez la glycoprotéine du virus VSV (Nishimura et Balch 1997). Cette séquence est impliquée dans l'export de la protéine M du réticulum endoplasmique (RE) vers l'appareil de Golgi (Perrier et al. 2019). En revanche, il a été observé que MERS-CoV-M pouvait être exportée du RE de façon $_{211}\text{DxE}_{213}$ -indépendante lorsque la protéine E est exprimée.

Une première étude a démontré que la protéine M du SARS-CoV et de MHV adoptait deux conformations : une conformation étendue (M_{LONG}) et une conformation compacte (M_{COMPACT}) (Neuman et al. 2011). Ces deux formes de M ont ensuite été confirmées grâce à la résolution de la structure de SARS-CoV-2-M (**Figure 10B et C**) (Zhang et al. 2022). Selon les données de Neuman *et al*, le conformère M_{LONG} correspondrait à une structure plus rigide que le conformère M_{COMPACT} . Également, ils ont observé que M_{LONG} était plus apte à induire des courbures membranaires, et que davantage de protéines spikes interagissait avec ce conformère. Ils ont ainsi suggéré que M_{LONG} était le conformère le plus favorable à l'assemblage des virions.

En plus de sa participation aux mécanismes d'assemblage, la protéine M joue un rôle dans la pathogénèse des coronavirus. Elle fait partie des quelques protéines virales qui inhibent la réponse immunitaire innée en bloquant la voie des IFN. Il a été observé que la protéine M du SARS-CoV, du MERS-CoV et du SARS-CoV-2 interfèrent avec la voie de signalisation RIG-I/MDA-5 qui est impliquée dans l'activation de la réponse IFN (Siu et al. 2009; Lui et al. 2016; Zheng et al. 2020; Fu et al. 2021). Pour la protéine M du SARS-

CoV, cette activité antagoniste est médiée par son premier passage transmembranaire (Siu et al. 2014).

1.5.5 Les protéines accessoires

Au-delà des protéines non-structurales et structurales décrites ci-dessus, les coronavirus produisent une troisième classe de protéines virales, dites « accessoires » (ORF). Les gènes codant ces protéines sont tous situés dans la région 3' du génome viral, insérés entre les gènes codant les protéines structurales. De plus, ces gènes sont présents en nombre variable : 8 gènes codant des protéines accessoires ont été identifiés chez le SARS-CoV, contre 11 chez le SARS-CoV-2 et 5 chez le MERS-CoV. Ces protéines ont la particularité de ne présenter aucune similarité avec d'autres protéines virales et cellulaires. Pendant longtemps, leurs fonctions étaient inconnues.

Il a été démontré *in vitro* que les protéines accessoires n'étaient pas essentielles pour la réplication du virus (de Haan et al. 2002; Yount et al. 2005). Il semblerait qu'elles soient davantage impliquées dans la pathogénèse des coronavirus, en modulant notamment la réponse immunitaire innée. Chez le MERS-CoV, la délétion de l'ensemble des gènes codant les protéines accessoires (MERS-dORF3-5) entraîne une atténuation du virus *in vitro* dans certains types cellulaires et *in vivo* (Menachery et al. 2017). Cela s'est traduit par l'induction d'une réponse inflammatoire plus robuste, avec une augmentation des IFN- β , IFN- γ , IL-1 β , RANTES et de la protéine MIP-1 α . Il s'est avéré que la réplication de MERS-dORF3-5 était moins efficace que le virus sauvage, expliquée par les auteurs par la forte réponse inflammatoire induite. Les protéines ORF4a, ORF4b et ORF5 ont montré, une à une, leur fonction antagoniste de la voie IFN (Siu et al. 2014). Comparé à ORF4b et ORF5, la protéine ORF4a présente la plus forte activité anti-inflammatoire (Yang et al. 2013). Le mode d'action d'ORF4a est d'inhiber la voie de signalisation RIG-I/MDA5 en liant, de manière ARN-dépendante, le facteur PACT.

Chez le SARS-CoV et SARS-CoV-2, les protéines ORF3b et ORF6 participent à l'évasion du virus à la réponse immunitaire, en inhibant l'expression des IFN- β . ORF3b agit sur le facteur IRF3, alors qu'ORF6, empêche la translocation nucléaire du facteur STAT1 phosphorylé (Kopecky-Bromberg et al. 2007; Miorin et al. 2020; Redondo et al.

2021). ORF9b participe également à l'inhibition de la réponse immunitaire, mais par des mécanismes différents. Transportée jusqu'aux mitochondries, ORF9b cible le signalosome MAVS impliqué dans la réponse antivirale par les RLR (*RIG-I-like receptor*) (Shi et al. 2014). De plus, ORF9b agit sur le morphodynamisme mitochondrial, en provoquant l'ubiquitination et la dégradation par le protéasome du facteur DRP1, responsable de la fission mitochondriale (Shi et al. 2014). Cette étude montre également qu'ORF9b module la voie de l'autophagie, caractérisée par une production importante d'autophagosome.

Les protéines ORF3a et ORF7a du SARS-CoV sont des protéines pro-inflammatoires, puisqu'elles stimulent les voies de signalisation NF- κ B et JNK des cellules A549 (Kanzawa et al. 2006). Le facteur NF- κ B est notamment connu pour activer la transcription de plusieurs gènes pro-inflammatoires. A contrario, SARS-CoV-2-ORF7a inhibe la voie IFN après avoir été polyubiquitinée sur la Lys119 par la machinerie cellulaire, ce qui bloquerait l'activité du facteur STAT2 (Cao et al. 2021).

Il a été observé que plusieurs protéines accessoires étaient responsables d'un arrêt du cycle cellulaire en phase G0/G1. C'est le cas des protéines ORF3a et ORF7a, qui module l'expression de la protéine cycline D3 et inhibe la phosphorylation de la protéine de rétinoblastome (Rb) sur les résidus Ser795 et Ser809-811 (Yuan et al. 2006, 2007). De plus, l'expression de ces protéines est responsable de l'activation de l'apoptose-dépendante des caspases dans différents types cellulaires (Tan et al. 2004; Law et al. 2005; Redondo et al. 2021). ORF3b semble également être une protéine pro-apoptotique.

De manière surprenante, certaines protéines accessoires peuvent être qualifiées de protéines structurales par leur capacité à être incorporées dans les particules lors de l'assemblage. Des observations en microscopie électronique ont permis de montrer qu'ORF3a était associée aux particules virales (Ito et al. 2005; Shen et al. 2005; Huang et al. 2006b). Cependant, cette protéine n'a pas de fonction dans les mécanismes d'assemblage, puisque son absence n'abroge pas la synthèse de VLP (Shen et al. 2005). De plus, il a été montré que des particules SARS-CoV- Δ 3a sont viables et se répliquent efficacement dans les cellules, confirmant la non-importance de cette protéine pour l'assemblage (Yount et al. 2005). Les protéines ORF6, ORF7a et ORF7b sont aussi incorporées dans les particules virales, comme le rapporte plusieurs études (Huang et al.

2006a, 2007; Schaecher et al. 2007). Notamment, il a été montré que le gène *ORF7b* pouvait être substituée par celui de la *GFP*, permettant la production de particules virales fluorescentes (Sims et al. 2005).

En plus de la protéine E, une activité viroporine a également été observée pour ORF3a, ainsi que pour ORF3 de PEDV et ORF4a de HCoV-229E (Wang et al. 2012; Zhang et al. 2014). Leur fonction viroporine a d'abord été un postulat, due à leur localisation membranaire, leur état oligomérique, et les similitudes de leur domaine C-terminal avec l'ATPase calcium de *Plasmodium falciparum* et la porine de *Shewanella oneidensis*. Cette activité a par la suite été démontrée expérimentalement sur des ovocytes de *Xenopus laevis*, où ORF3a permettrait le passage sélectif des ions K^+ (Lu et al. 2006). Cependant, d'autres études ont démontré qu'ORF3a ne modifiait pas le courant électrique au niveau des membranes de ces ovocytes (Grant et Lester 2021; Miller et al. 2023).

1.6 Le cycle viral

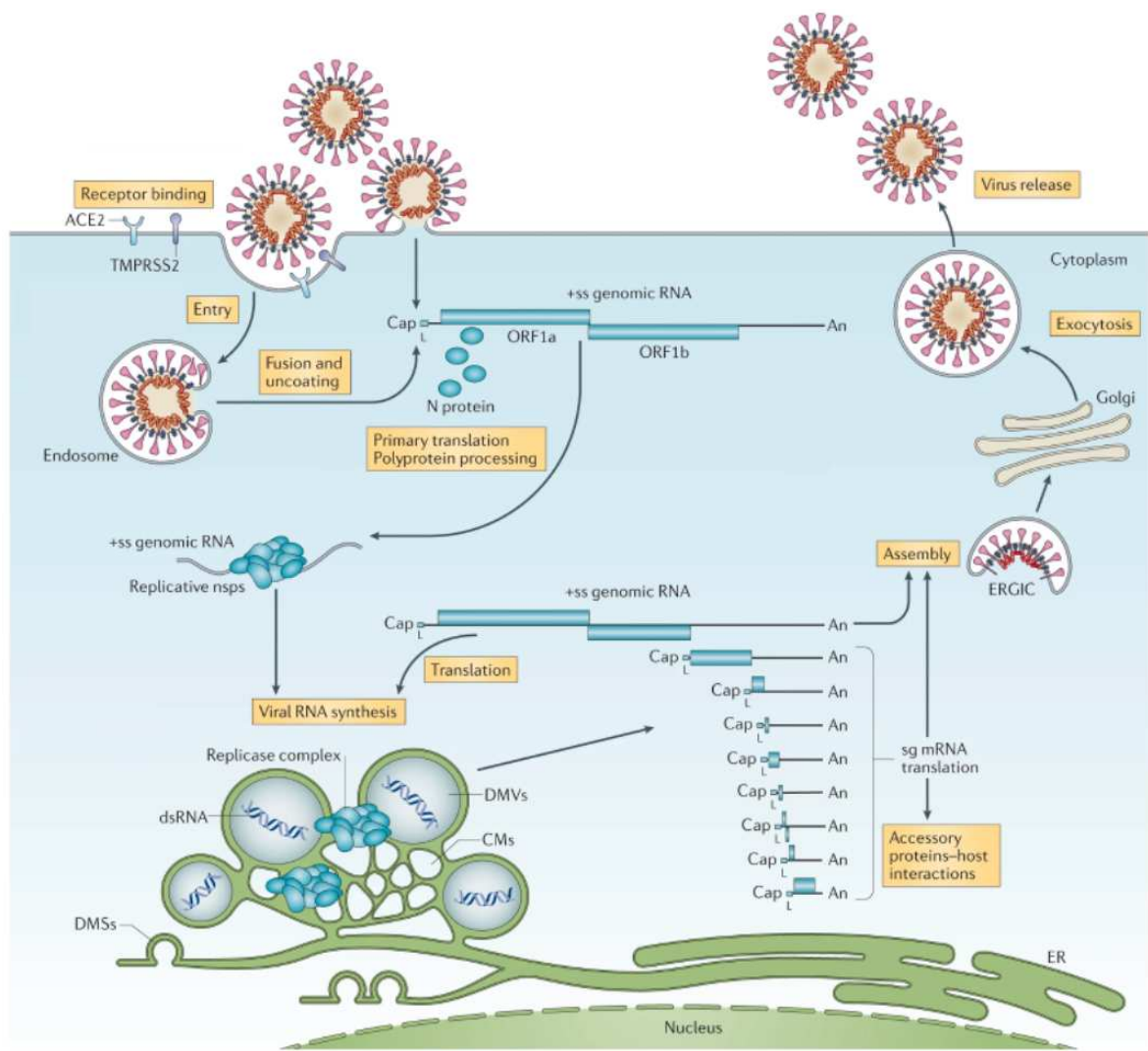


Figure 12. Les étapes du cycle viral des coronavirus.

Deux voies possible d'entrées des coronavirus dans leur cellule cible ont été décrites, la voie de fusion à la surface (ou voie précoce) et la voie endosomale. La voie précoce implique une fusion directe de la particule virale avec la membrane plasmique de la cellule. La voie endosomale nécessite une internalisation du virus par endocytose afin d'induire avant la fusion. Suite à cette étape, le génome est relargué dans le cytoplasme et sert de matrice ARNm pour les ribosomes cellulaires afin de synthétiser les protéines non-structurales (nsp), qui formeront le complexe de réplication/transcription viral. Le génome est ensuite répliqué (ARNg) et transcrit (ARNsg) à l'intérieur de vésicules double-membranaires (DMV) issues du RE. Les protéines structurales S, E et M sont synthétisées au niveau du RE et sont ensuite exportées vers le compartiment ERGIC, où tous les constituants nécessaires à la formation des nouveaux virions sont adressés. Les virions néosynthétisés sont ensuite transportés par voie vésiculaire jusqu'à la membrane plasmique, afin d'être libérés par exocytose. Figure adaptée de V'kovski et al 2021.

1.6.1 Attachement à la cellule cible

L'initiation du cycle infectieux des coronavirus requiert deux étapes fondamentales : (i) l'attachement du virus à sa cellule cible et (ii) à l'induction de la fusion de l'enveloppe virale avec les membranes cellulaires, aboutissant au relargage de la nucléocapside dans le cytoplasme de l'hôte.

Ces étapes sont dépendantes de la protéine S. Comme décrit dans la **section 1.5.4.1**, la sous-unité S1 de la protéine spike contient un domaine de liaison au récepteur (RBD). Ce domaine RBD contient un motif de liaison au récepteur (RBM) qui est directement impliqué dans les interactions avec les récepteurs cellulaires exposés à la surface. Ces récepteurs ne sont, pour la plupart, pas exprimés de façon ubiquitaire. En d'autres termes, cela signifie qu'ils sont exprimés de manière tissu-dépendant. Par conséquent, le tropisme cellulaire des coronavirus est dirigé par la protéine spike. Cela a notamment pu être montré grâce à la génération de virus MHV chimériques. Il existe une multitude de souches du virus MHV, avec des tropismes différents. Ces souches se distinguent essentiellement par la présence de variation au niveau de leur protéine S. La souche JHM est une souche majoritairement neurotrope, induisant des encéphalites sévères. En revanche, la souche MHV-2 se réplique davantage dans les cellules hépatiques. La capacité de MHV-A59 à induire des encéphalites et des hépatites montrent que cette souche est à la fois neurotrope et hépatotrope. Cependant, la recombinaison de la protéine S des souches JHM ou MHV-2 avec celle de la souche MHV-A59 a permis d'augmenter, respectivement, le neurotropisme et le tropisme hépatique de cette souche (Navas et al. 2001; Navas et Weiss 2003). Le rôle important de la protéine S dans le tropisme cellulaire des coronavirus a également pu être confirmé chez IBV (Casais et al. 2003).

Il existe une grande diversité de récepteurs cellulaires permettant au virus de s'attacher à la surface. A ce jour, 4 récepteurs ont été identifiés.

L'aminopeptidase N (APN), aussi connu sous le nom de CD13, est une protéine transmembranaire de type II. C'est une protéine de la famille des métalloprotéases et est exprimée dans de nombreux tissus cellulaires, comme les cellules respiratoires, intestinales et dans le SNC. Il a été montré qu'APN servait de récepteur à plusieurs

alphacoronavirus, tels que le sérotype 2 de FCoV et CCoV, TGEV, PEDV et le coronavirus humain HCoV-229E. L'interaction de ces coronavirus avec le récepteur APN s'effectue de manière espèce-dépendante.

La protéine S du MERS-CoV est spécifique au facteur dipeptidyl peptidase 4 (DPP4 ou CD26) pour médier l'attachement du virus à la surface de la cellule cible (Raj et al. 2013). DPP4 est une exopeptidase à sérine multifonctionnelle, notamment impliquée dans la régulation de la réponse immunitaire, de la transduction du signal et de l'apoptose. DPP4 est très majoritairement exprimé dans l'appareil digestif et rénal. Au niveau du système respiratoire, l'expression de DPP4 est majoritaire dans les bronches, faisant lien avec les pneumopathies sévères induites par le MERS-CoV. La structure cristallographique du complexe MERS-S-RBD/DPP4 a pu être obtenue, permettant d'identifier 2 régions importantes pour l'interaction, où s'effectuent de nombreuses liaisons énergétiquement faibles (Wang et al. 2013). La substitution Y499A du RBD de MERS-CoV-S empêche totalement l'interaction entre le RBD et DPP4. L'analyse de la structure du complexe suggère que le résidu Y499 effectue des liaisons hydrogènes essentielles avec R336 de DPP4, assurant la stabilité du complexe.

Plusieurs coronavirus utilisent l'enzyme de conversion de l'angiotensine 2 (ACE2) comme récepteur cellulaire. ACE2 est le récepteur clé pour l'infection par les *sarbecovirus* humains (SARS-CoV ; SARS-CoV-2), mais aussi par HCoV-NL63 (Li et al. 2003; Hofmann et al. 2005; Shang et al. 2020). ACE2 est une protéine transmembranaire de type I, exprimée par les cellules endothéliales pulmonaires, cardiaques, rénales et est également présente dans le tube digestif. Son rôle est de catalyser la conversion de l'angiotensine 1 et 2 en angiotensine (1-9) et (1-7). Par son activité, ACE2 est au cœur des mécanismes de régulation de la pression artérielle. L'utilisation du récepteur ACE2 par le SARS-CoV-2 peut être associée au développement de forme sévère de COVID-19 chez des patients souffrant de comorbidités, comme l'hypertension. En effet, il semblerait que certains médicaments hypertenseurs entraînent une surexpression du récepteur ACE2, favorisant l'infection (Cinaud et al. 2021).

Le quatrième récepteur identifié est la molécule CEACAM1 (molécules d'adhérence cellulaire liées aux antigènes carcino-embryonnaires). C'est une protéine

transmembranaire de type I appartenant à la grande famille des immunoglobulines. Elle assure de multiples fonctions, comme dans l'adhésion et la signalisation cellulaire. A ce jour, MHV est le seul coronavirus connu à utiliser CEACAM1 comme récepteur cellulaire. Il a été montré que le RBD de MHV-A59 interagissait avec le domaine N-terminal de CEACAM1 (Dveksler et al. 1993). Cependant, une étude a démontré que la neurovirulence de MHV-JHM était indépendante à CEACAM1. En effet, Miura *et al.*, ont observé une meilleure efficacité répliquative de MHV-JHM dans des souris déficientes *ceacam1*^{-/-} (Miura et al. 2008). Ces observations suggèrent que MHV-JHM puisse utiliser un récepteur alternatif à CEACAM1.

Certains coronavirus, comme les *betacoronavirus* BCoV, HCoV-OC43 et HCoV-HKU1, se lient à des acides sialiques terminaux exposés à la surface de la cellule cible par des glycoprotéines ou des glycolipides (Hulswit et al. 2019). Cette action de liaison aux acides sialiques pourrait s'expliquer par la présence de la protéine HE à la surface de ces virus (Desforges et al. 2013). Pour d'autres coronavirus, les acides sialiques peuvent servir de facteur d'attachement.

Dans le but de favoriser l'interaction entre leur domaine RBD et le récepteur cellulaire principal, certains coronavirus peuvent être amenés à utiliser des facteurs d'attachement. Notamment, le SARS-CoV, le SARS-CoV-2 et HCoV-NL63 sont connus pour interagir avec des héparanes sulfates (HSPG). Les HSPG sont des protéoglycannes complexes appartenant à la famille des glycosaminoglycannes (GAG). Concernant le SARS-CoV et le SARS-CoV-2, il a été montré que l'interaction avec les HSPG s'effectuait en complément de l'interaction avec le récepteur ACE2 (Lang et al. 2011; Clausen et al. 2020). Cette interaction semble favoriser les changements conformationnels du domaine RBD nécessaire à la liaison à ACE2. De plus, l'inhibition de l'interaction du RBD avec les HSPG par la lactoferrine, ne permet plus l'infection de cellules HEK293E/ACE2 par des particules pseudovirales de SARS-CoV. Ces données permettent de mettre en évidence l'importance des HSPG dans le cadre des infections à SARS-CoV et SARS-CoV-2. Concernant HCoV-NL63, il a été démontré que ses interactions avec les HSPG n'étaient pas dépendantes de la protéine S, mais plutôt de la protéine M (Milewska et al. 2014; Naskalska et al. 2019).

Certains coronavirus sont en mesure d'utiliser les lectines de type-C DC-/L-SIGN comme récepteur alternatif ou d'attachement. Ces lectines sont des récepteurs de la réponse immunité innée (PRR) et reconnaissent des glycanes riches en mannose portés par des agents pathogènes (PAMP) exposés à la surface des cellules présentatrices d'antigène. L'utilisation de ces lectines comme récepteur a été rapportée pour les coronavirus FCoV, HCoV-229E, HCoV-NL63, SARS-CoV et SARS-CoV-2 (Hofmann et al. 2006; Han et al. 2007a; Regan et Whittaker 2008; Amraei et al. 2021).

1.6.2 La fusion

Les virus sont des parasites obligatoires. Il leur est donc indispensable de délivrer leur génome dans le cytoplasme d'une cellule cible afin que celui-ci soit répliqué. C'est pourquoi, suite à l'attachement au récepteur, les virus enveloppés doivent fusionner leur enveloppe avec une membrane cellulaire, afin de créer un pore permettant le passage du génome viral vers le cytoplasme.

Les mécanismes de fusion des coronavirus sont contrôlés par la protéine S, grâce au peptide de fusion (FP) contenu dans la sous-unité S2. Il existe deux modes d'entrée des coronavirus dans leur cellule cible. Le premier mode d'entrée consiste en la fusion de l'enveloppe virale directement à la surface cellulaire (voie précoce), alors que le second mode s'effectue par l'intermédiaire du compartiment endolysosomal, nécessitant au préalable l'internalisation du virus par endocytose (voie endosomale). Lors de l'étape de fusion, la protéine S subit d'importants changements conformationnels, afin de passer d'un état « préfusion » vers un état « postfusion ». Ces changements conformationnels peuvent être induits (i) directement après liaison au récepteur cellulaire, (ii) par exposition à un pH-acide, (iii) ou par clivage protéolytique.

IBV est un rare coronavirus pour lequel il a été montré qu'un pH acide pouvait activer la fusion (Chu et al. 2006). Pour la plupart des coronavirus, il a été montré que la fusion était activée par un clivage protéolytique de la protéine spike. Différentes protéases cellulaires peuvent médier la fusion des coronavirus et leur localisation va conditionner le mode d'entrée (précoce ou tardif) du virus. Ainsi, la protéase TMPRSS2 exprimée à la surface des cellules au niveau des tissus respiratoires joue un rôle fondamental dans

l'entrée des coronavirus SARS-CoV, SARS-CoV-2 et MERS-CoV (Glowacka et al. 2011; Shirato et al. 2013; Bestle et al. 2020). En absence de cette protéase, le virus est internalisé et la fusion peut être activée par des protéases endosomales, telles que les cathepsines (Simmons et al. 2005; Zhao et al. 2021).

Le clivage des protéines spike au niveau de la région S2' (voir Figure 9A) est central dans l'activation de la fusion afin d'exposer le peptide de fusion qui est alors libre pour s'insérer dans la membrane de la cellule cible (Belouzard et al. 2009; Madu et al. 2009; Jackson et al. 2022) (**Figure 13**). La sous-unité S2 adopte alors une structure pré tige-boucle (« prehairpin »). A ce stade, les domaines HR1 et HR2 sont étendus et alignés. Une restructuration des domaines HR1 et HR2 s'ensuit, induisant la formation d'un faisceau de 6 hélices alphas. Cette restructuration provoque un rapprochement de la membrane virale et cellulaire avec une juxtaposition du peptide de fusion et du domaine transmembranaire de la protéine et permet de former le pore de fusion à travers lequel la nucléocapside transitera pour rejoindre le cytoplasme de la cellule cible.

Le rôle du clivage de la jonction S1/S2 dans l'entrée dépend probablement du virus. Ainsi pour le SARS-CoV dont la spike n'est pas clivée lors de sa synthèse, un clivage lors de l'entrée de cette jonction est importante pour faciliter le clivage en S2' (belouzard et al. 2009) contrairement au virus HCoV-229E (Bonnin et al. 2018). Le SARS-CoV-2 possède un site polybasique au niveau de cette jonction. Il a été démontré que cette séquence est clivée par la furine (Hoffmann et al. 2020). La furine est une protéase présente à la surface cellulaire, dans le TGN et dans le compartiment endolysosomal grâce à un motif de recyclage contenu dans son domaine cytoplasmique et permettant le transport rétrograde de la protéine (Bosshart et al. 1994; Molloy et al. 1994). Le remplacement du site polybasique par un site monobasique permet une meilleure propagation du virus en cellules Vero-E6, cependant, cet avantage est perdu en cellules Calu-3, cellules qui expriment TMPRSS2 (Johnson et al. 2021). Les modifications du site de clivage affectent aussi la pathogénicité du virus en modèle animal (Johnson et al. 2021).

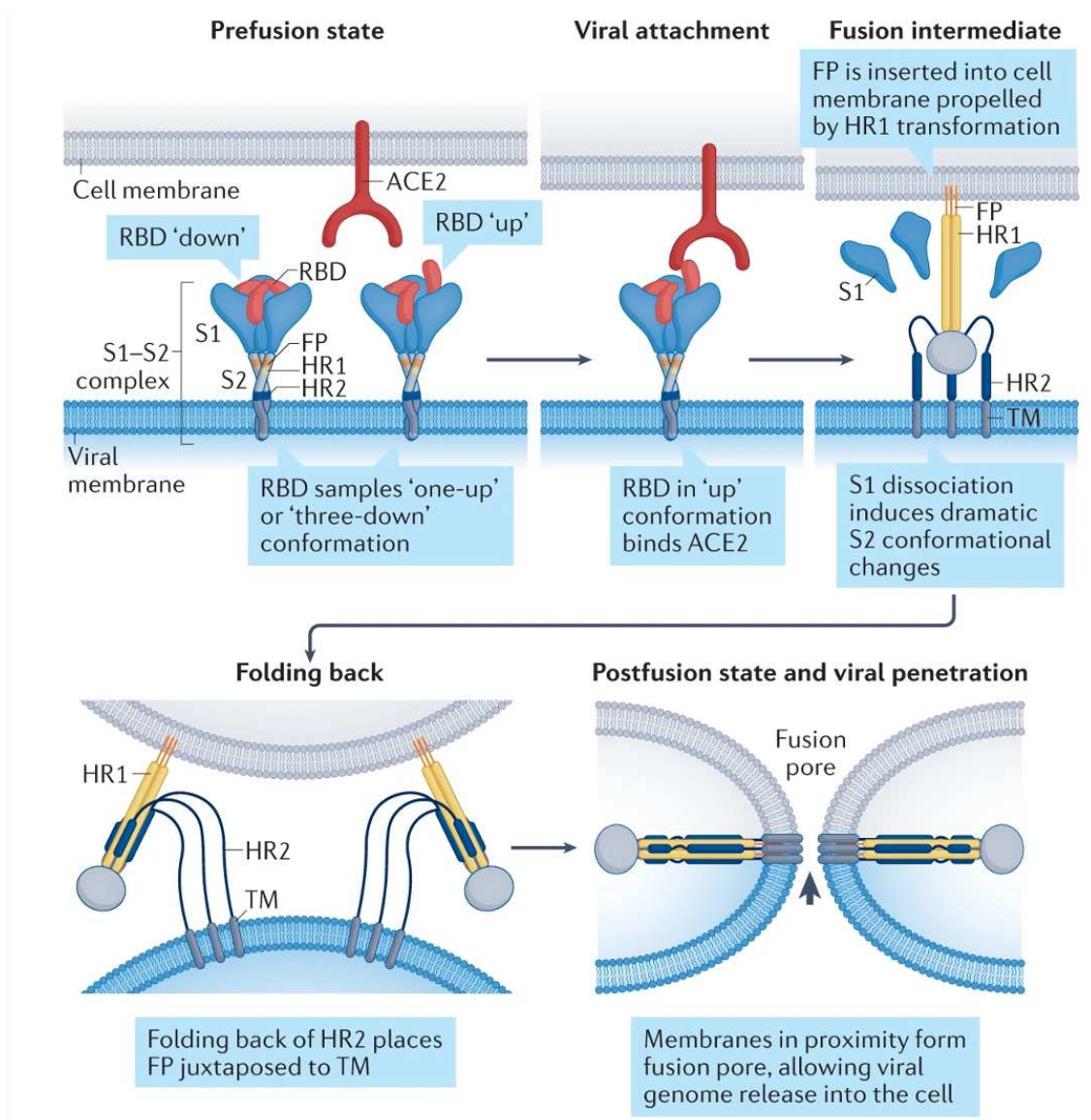


Figure 13. Mécanismes d'attachement et de fusion du SARS-CoV-2.

Le domaine RBD d'un monomère de protéine S oscille entre une conformation ouverte (« up ») et une conformation fermée (« down »). La conformation ouverte permet la liaison du récepteur ACE2. Après clivage de la protéine spike qui permet la dissociation du domaine S1 et la libération du peptide de fusion, ce dernier va s'ancrer dans une membrane de la cellule cible (stade pré tige-boucle). Les changements conformationnels successifs entraînent la formation d'un faisceau de 6 hélices, provoquant le rapprochement et la fusion des membranes virales et cellulaires aboutissant à la formation d'un pore à travers lequel sera relargué l'ARN génomique viral. Figure adaptée de Jackson et al. 2022.

1.6.3 Réplication et transcription du génome

Afin de produire des nouveaux virions, le génome viral doit être traduit, répliqué et transcrit après sa libération dans le cytoplasme de la cellule hôte. Dans un premier temps, la nucléocapside est dénudée afin de libérer le génome viral, qui servira de matrice ARNm pour les ribosomes. Ces-derniers vont traduire ORF1a et ORF1b directement sur l'ARN génomique (ARNg). Par ailleurs, cet ARNg se comporte comme un ARN messager eucaryote, puisque l'initiation de la traduction est dépendante de la coiffe m⁷G et un scan de la région 5'-UTR est réalisé. La traduction de ORF1a permet de synthétiser la polyprotéine pp1a, et celle de ORF1b permet la synthèse de pp1ab. Ces deux polyprotéines sont les précurseurs de l'ensemble des protéines non-structurales. ORF1a et ORF1b ont la particularité de partager un unique codon d'initiation pour la traduction. L'expression d'ORF1b est permise grâce à un décalage du cadre de lecture d'un nucléotide en arrière (-1). Ce décalage est permis par la présence d'une structure ARN de type pseudonœud ainsi que d'une séquence « glissante » de 7-nt (5'-UUUAAAC-3') commune à tous les coronavirus, et situés juste en amont du codon stop d'ORF1a. Chez MHV et SARS-CoV-2, il est estimé que le décalage -1 du cadre de lecture s'effectue dans 50 à 70 % des cas (Irigoyen et al. 2016; Finkel et al. 2021).

Après maturation des polyprotéines pp1a et pp1ab, les protéines non-structurales s'associent pour former le complexe de réplication/transcription viral (RTC). La formation du RTC et les étapes de réplication et de transcription de l'ARNg se déroulent au sein de vésicules double-membranaires (DMV) qui dérivent du réticulum endoplasmique (Knoops et al. 2008). Il a été montré que la formation de ces DMV était induite par les protéines nsp3, nsp4 et nsp6 (Angelini et al. 2013). Ainsi, la synthèse d'ARN viraux dans des compartiments physiquement séparés du cytoplasme, permettrait de masquer les ARN double-brins intermédiaires aux senseurs de l'immunité innée, tels que les récepteurs de type RLR (Rig-I-like receptor).

Une fois la formation du RTC achevée, celui-ci s'associe avec l'extrémité 3' de l'ARNg positif, afin de synthétiser une molécule intermédiaire complémentaire (-)ARNg par une synthèse ARN dite « continue ». Cet intermédiaire servira ensuite de matrice pour synthétiser de nouvelles copies (+)ARNg. Il existe également une synthèse ARN dite

« discontinue », permettant la transcription des gènes localisés dans la région 3' du génome et codant les différentes protéines structurales et accessoires. La transcription de ces gènes se fait sous la forme d'ARN sous-génomiques (ARNsg), qui ont la particularité de posséder les mêmes extrémités 5' et 3', ce qui en fait un caractère propre aux *Nidovirus*. La transcription du génome est possible grâce à la présence de « séquences de régulations transcriptionnelles » (TRS), situées en amont de chaque ORF de l'extrémité 3' (TRS_B) ainsi que dans la région 5'-leader (TRS_L). Chez le SARS-CoV et le SARS-CoV-2, cette séquence TRS a été définie comme 5'-ACGAAC-3'. Lors de la synthèse d'un (-)ARNsg, un modèle propose que le RTC fasse une pause au niveau d'un TRS_B. Ensuite, le complexe RTC ferait un « saut » jusqu'au TRS_L du (+)ARNg servant de matrice, dû à une complémentarité des séquences. Il est possible qu'une interaction ARN-ARN entre le (+)ARNg et le (-)ARNsg s'établisse entre le TRS_L et le TRS_B, permettant le saut du RTC, et résultant en une synthèse ARN discontinue. Les intermédiaires (-)ARNsg serviront ensuite de matrice pour la synthèse des (+)ARNsg, qui seront eux-mêmes traduits par la machinerie ribosomale de l'hôte pour produire les protéines structurales et accessoires.

1.6.4 Encapsidation, bourgeonnement et sécrétion

1.6.4.1 Encapsidation du génome

Après les étapes de répllication et de transcription réalisés par le RTC, le génome viral sort des DMV en traversant un pore qui semble être formé par un homohexamère de protéine nsp3 (~1,2 MDa) (Wolff et al. 2020). Une fois localisées dans le cytoplasme, les molécules d'ARNg interagissent avec les protéines N néosynthétisées afin de former la nucléocapsides des futurs virions. Il a été montré que la protéine N des coronavirus MHV et SARS-CoV-2 est localisée proche du pore des DMV à travers lequel transite le génome viral, facilitant ainsi l'interaction avec ce-dernier.

La protéine N est une protéine à multi-domaines (voir chapitre 1.5.4.3). Elle interagit avec l'ARNg via son domaine de liaison à l'ARN situé dans la partie N-terminale de la protéine (Bai et al. 2021). Au niveau de l'ARNg, l'interaction avec la protéine N s'effectue au niveau de signaux spécifiques (« packaging signal »), d'une taille variable

selon le virus. Chez le SARS-CoV, il a été montré *in silico* que ces signaux étaient longs de 70-nt et se localisaient proche de l'extrémité 3' d'ORF1b, région la plus variable de ce gène (Qin et al. 2003). Chez le SARS-CoV-2, ces signaux n'ont pas encore été déterminés. L'interaction entre la protéine N et l'ARNg est à l'origine de la formation de la capsidale hélicoïdale des coronavirus, dont la formation est une étape indispensable pour l'incorporation du génome au sein des nouveaux virions (Hsieh et al. 2005). En effet, Hsieh *et al.*, ont démontré un défaut d'incorporation d'un ARN témoin fluorescent dans des VLP de type SARS-CoV déficientes en protéine N. De plus, l'incorporation du complexe N-ARNg dans les nouveaux virions dépend de l'interaction directe entre les protéines N et M au niveau de leur domaine C-terminale (Luo et al. 2006).

1.6.4.2 Bourgeonnement

Rôle de la protéine M

La protéine M est la protéine la plus représentée à la surface des particules. De plus, cette protéine est considérée comme le moteur de l'assemblage des coronavirus. En effet, il est connu que la protéine M est engagée dans des interactions intermoléculaires avec l'ensemble des protéines structurales (Masters 2006). Elle est également impliquée dans des interactions homotypiques grâce à des contacts multiples au niveau des segments transmembranaires et de la queue C-terminale (De Haan et al. 2000; Neuman et al. 2011). Ces interactions homotypiques résultent, donc, en l'oligomérisation de M, récemment confirmée par la résolution de la structure 3D de la protéine (Dolan et al. 2022; Zhang et al. 2022). L'oligomérisation de M joue un rôle important dans le processus d'assemblage. En effet, M est en mesure de former des homodimères, qui s'associeront par la suite en tétramères et hexamères (Zhang et al. 2022). Ainsi, l'oligomérisation de M permet la formation d'une matrice protéique dans laquelle viendront s'insérer les protéines S et E.

Comme discuté en 1.5.4.5, la protéine M a été décrite sous deux conformères (Neuman et al. 2011 ; Zhang et al. 2022). Une forme allongée (M_{LONG}) et une forme compactée ($M_{COMPACT}$) ont été observées par cryo-EM. Il a été suggéré par Neuman *et al.*, que M_{LONG} était le conformère favorable à l'assemblage des particules. Dans cet état, l'endodomaine de la protéine est davantage exposé côté cytoplasmique, ce qui permet

l'interaction avec le complexe N-RNA et donc son incorporation dans les virions. Il est possible que la transition entre les deux conformations soit liée à la présence du complexe N-RNA dans l'environnement de M. De plus, M_{LONG} semble être davantage en mesure d'interagir avec la protéine S que le conformère $M_{COMPACT}$. Également, M_{LONG} serait à l'origine de courbure de la membrane, permettant d'initier le bourgeonnement des particules. En conclusion, la conformation adoptée par la protéine M joue probablement un rôle important dans les étapes d'assemblage.

Il est largement admis que la co-expression, à minima, des protéines M et E en culture cellulaire permet la production de VLP (Vennema et al. 1996; Baudoux et al. 1998; Corse et Machamer 2000; Liang et al. 2019) pour certains coronavirus. Dans le contexte du SARS-CoV, la protéine N semble être requise pour produire ces particules, bien que certaines études parviennent à produire des VLP en l'absence de SARS-CoV-N (Siu et al. 2008). Par conséquent, cela démontre l'importance des interactions M-E pour l'assemblage des particules (Corse et Machamer 2003). L'expression de la protéine E du coronavirus MHV-A59 induit d'importants réarrangements membranaires des membranes pré-golgiennes (Raamsman et al. 2000). De plus, des aberrations morphologiques ont été observées après l'insertion de mutations dans le domaine C-terminal de MHV-E, avec des virions à la structure allongée (Fischer et al. 1998). Il est donc fort possible qu'un défaut dans le mécanisme de scission soit à l'origine de l'allongement des virions.

Le domaine C-terminal de la protéine M joue également un rôle important pour l'assemblage. Chez MHV-M, l'introduction de mutation dans la séquence conservée SWWSFNPETNNL abroge la formation de VLP (Arndt et al. 2010). Cette séquence se situant du côté C-terminal, il se pourrait qu'elle soit importante pour la dimérisation de la protéine.

Rôle des modifications post-traductionnelles

L'assemblage des coronavirus semble également dépendre des modifications post-traductionnelles de certaines protéines structurales. En effet, il a récemment été montré que le résidu K15 de la protéine M du SARS-CoV-2 était ubiquitiné par l'ubiquitine ligase

E3 RNF5 (Yuan et al. 2022). La stabilité du complexe M-E semble dépendre de l'ubiquitination de K15. Les auteurs ont observé par microscopie électronique une diminution de particules de SARS-CoV-2 sécrétées après avoir diminué l'expression de RNF5. Chez le coronavirus MHV, la protéine E est palmitoylée sur trois résidus de cystéine (C40, C44 et C47). La mutation de ces trois résidus en alanine ou en glycine abroge la synthèse de VLP, bien que, ni le trafic intracellulaire de MHV-E, ni l'interaction entre M et E n'aient été impactés (Boscarino et al. 2008). La protéine MHV-S porte aussi des palmitoylations importantes sur plusieurs résidus de cystéine. En effet, l'utilisation d'un inhibiteur de palmitoyl-acyltransférases impacte la formation du complexe M-S, et donc l'incorporation de spike dans les particules (Thorp et al. 2006).

Trafic intracellulaire des protéines structurales

Alors que certains virus enveloppés à ARN s'assemblent et bourgeonnent à la membrane plasmique, les coronavirus effectuent ces étapes au niveau du compartiment intermédiaire entre le RE et l'appareil de Golgi (ERGIC) (Klumperman et al. 1994; Stertz et al. 2007). Il est donc primordial que l'ensemble des facteurs impliqués dans l'assemblage soient adressés et retenus au niveau du ERGIC. Cependant, les protéines virales ont tendance à se localiser au-delà du site ERGIC. C'est le cas de la protéine M du MERS-CoV, qui est retenue au niveau du réseau *trans*-golgien (TGN) en transfection transitoire (Perrier et al. 2019). De plus, l'ajout de O- ou N-glycosylations de type complexes confirme le trafic de la protéine jusqu'au TGN (Voß et al. 2006). En revanche, IBV-M est retenue plutôt au niveau du compartiment *cis*-golgien (Machamer et al. 1990). En cellules infectées, des données préliminaires obtenues au laboratoire montrent que la protéine M est présente dans toute la voie de sécrétion, suggérant un rôle des interactions protéine-protéine dans le trafic intracellulaire de M. Concernant la protéine S, il a été observé qu'elle migre jusqu'à la surface cellulaire, en cohérence avec la formation de syncytia. De plus, la protéine spike arbore de nombreuses glycosylations spécifiques du Golgi. Quant à la protéine E, elle est davantage retenue au niveau des compartiments ERGIC/Golgi (Venkatagopalan et al. 2015).

La protéine S possède un motif de recyclage KxHxx lui permettant d'interagir avec la protéine de manteau COPI (coatomer protein I) impliquée dans le transport rétrograde.

Principalement localisée à la surface cellulaire, la protéine S peut ainsi être rappatriée vers le site d'assemblage, où les interactions avec la protéine M permettraient de la retenir au niveau du ERGIC (McBride et al. 2007; Boson et al. 2021). L'étude de Boson *et al.*, montre que la protéine E est aussi impliquée dans la rétention de spike au niveau du ERGIC. Son activité viroporine induirait un ralentissement de la voie de sécrétion, favorisant la rétention de spike au niveau des compartiments intracellulaires. Pour la protéine M du MERS-CoV, des données obtenues au laboratoire montrent que la mutation des motifs $_{199}KxGxYR_{204}$ ou $_{211}DxE_{213}$, importants pour le trafic intracellulaire de la protéine, empêche la formation et la sécrétion de particules virales infectieuses (Desmarets et al. 2023, manuscrit soumis).

1.6.4.3 Sécrétion

L'ultime étape dans le cycle de vie des virus est la sécrétion des nouvelles particules après leur assemblage dans le ERGIC. Il est admis qu'après l'assemblage et le bourgeonnement des nouveaux virions, ces-derniers transitent à travers l'appareil de Golgi, puis par voie vésiculaire, rejoignent la membrane plasmique afin qu'ils soient libérés par exocytose. Ce mécanisme de sécrétion est notamment le cas du virus de l'hépatite C, de la dengue ou encore du Nil Occidental.

Les mécanismes de sécrétion des coronavirus sont encore mal compris et plusieurs hypothèses sont envisagées. Tout d'abord, il a longtemps été considéré que les coronavirus utilisaient la voie de sécrétion constitutive. Chez les coronavirus, la protéine E semble également avoir une fonction dans le transit des particules vers la surface cellulaire. En effet, des mutations dans son domaine transmembranaire, nécessaire pour l'oligomérisation et l'activité canal ionique de la protéine, induisent une diminution de la quantité de particules virales sécrétées (Ye et Hogue 2007). De plus, des particules TGEV recombinantes déficientes en protéine E (rTGEV Δ E) sont retenues dans la voie sécrétion, entre le RE et le *cis*-golgi (Ortego et al. 2007). En revanche, lorsque les cellules infectées par rTGEV Δ E expriment TGEV-E, les particules virales sont visibles dans des vésicules d'exocytose. Une étude rapporte aussi que les particules de SARS-CoV-2 sont observées dans de petites vésicules de sécrétion (Eymieux et al. 2021).

Une étude récente a montré que les *betacoronavirus* utiliseraient une voie de sécrétion non-canonique pour sortir de la cellule hôte, puisque des particules de MHV et de SARS-CoV-2 ont été détectées, par microscopie électronique à transmission, dans des vésicules endolysosomales (Ghosh et al. 2020). Par ailleurs, le coronavirus PEDV avait également été observé dans des lysosomes (Ducatelle et Hoorens 1984). Si la voie lysosomale est utilisée par les coronavirus pour sortir de la cellule hôte, il est nécessaire qu'une désacidification des vésicules ait lieu, afin d'empêcher l'action des protéases. En cellules infectées, cette étape de désacidification a été observée, et pourrait être induite par l'activité canal ionique de la protéine ORF3a (Ghosh et al. 2020). Egalement, cette protéine réside dans les lysosomes, comme l'indique la co-localisation avec le marqueur LAMP1. Chez le SARS-CoV, il a également été suggéré que l'activité canal ionique de ORF3a était requise pour la libération des nouveaux virions (Lu et al. 2006).

Il a aussi été montré qu'ORF3a du SARS-CoV-2 modulait la voie d'autophagie afin de promouvoir la sécrétion des particules par les lysosomes (Chen et al. 2021). En effet, ORF3a perturbe la fusion des autophagosomes avec les lysosomes, ce qui aurait induit la dégradation du contenu intravésiculaire (Miao et al. 2021). Pour cela, ORF3a, retenu au niveau des endosomes tardifs/lysosomes, séquestre le facteur VPS39. Ce facteur est une sous-unité du complexe HOPS médiant la fusion autophagosomes-lysosomes. Cependant, une telle activité de SARS-CoV-2-ORF3a n'a pas été confirmée avec son homologue chez le SARS-CoV. Par conséquent, d'autres mécanismes permettant aux particules de transiter jusqu'à la membrane plasmique par l'intermédiaire des lysosomes pourraient exister.

1.7 La N-glycosylation

1.7.1 Généralités

La glycosylation est une réaction biochimique commune à tous les êtres vivants. Elle consiste en la liaison covalente de glucides, souvent aux structures complexes, sur des molécules. Ces molécules peuvent être de nature protéique ou lipidique. L'ajout de ces glucides s'effectue sur des sites accepteurs. Ainsi, on distingue différents types de glycosylation, dont deux principales : la N- et O-glycosylation. On parle de N-glycosylation

quand une chaîne glucidique est ajoutée sur la fonction NH d'une asparagine appartenant à une séquence consensus Asn-X-Ser/Thr, où X correspond à n'importe quel acide-aminé sauf une proline. La O-glycosylation correspond à l'addition de glucide sur la fonction OH des acides-aminés sérines ou thréonines. Cependant, aucune séquence consensus n'a été identifiée.

Il existe 3 types de N-glycanes : (1) riches en mannose ou oligomannose, (2) hybrides et (3) complexes (**Figure 14**). Chaque N-glycane partage une structure commune de type (Man₃GlcNAc₂)-Asn. Cette structure commune est étendue exclusivement par des résidus mannoses pour les glycanes oligomannoses, alors que les glycanes complexes peuvent posséder jusqu'à 5 branches, toutes initiées par un résidu GlcNAc et allongées par des unités N-acétyllactosamines ([Galβ1-4GlcNAc]). Des résidus fucosylés peuvent être ajoutés, ainsi que des acides sialiques aux extrémités. Enfin, les glycanes hybrides correspondent à des structures intermédiaires entre oligomannose et glycane complexe, présentant les caractéristique de chacun d'entre eux.

La N-glycosylation assure plusieurs fonctions sur les glycoprotéines : (i) de contrôle qualité du repliement des protéines, (ii) elle augmente la solubilité, (iii) est importante pour l'activité des protéines et (iiii) agit sur l'antigénicité.

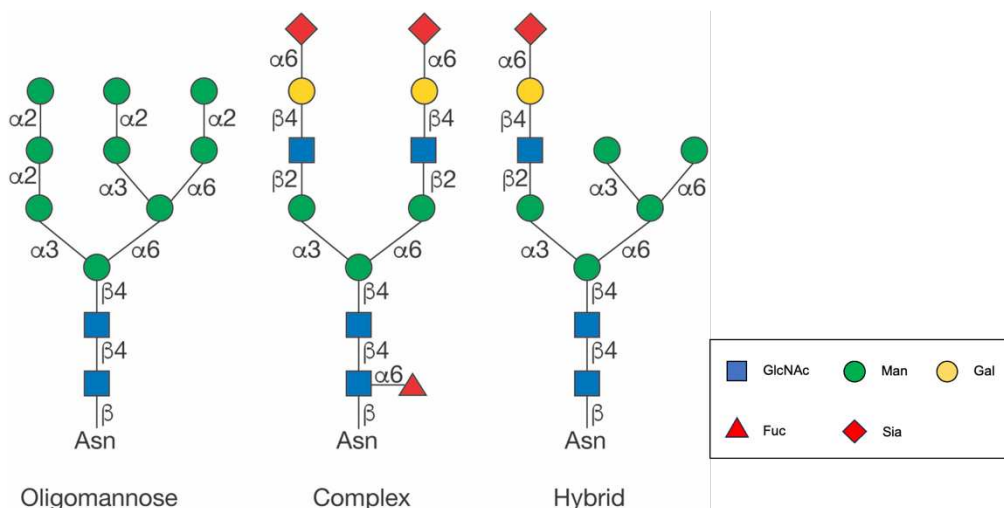


Figure 14. Représentations schématiques des N-glycanes riches en mannose, complexes et hybrides. Figure adaptée de Varki et al. 2022.

1.7.2 La voie de biosynthèse des N-glycanes

La N-glycosylation des protéines est un processus qui est organisé au sein des cellules eucaryotes. En effet, la synthèse s'initie au niveau du RE puis s'achève à travers les diverses citernes de l'appareil de Golgi (**Figure 15**) (Varki et al. 2022). Ce mécanisme débute par la synthèse d'un précurseur oligosaccharidique conjugué à un lipide polyisoprénique, le dolichol-pyrophosphate (Dol-P) ancré dans la membrane du RE. L'initiation de la synthèse de ce précurseur s'opère côté cytoplasmique du RE, où 7 monosaccharides ($\text{Man}_5\text{GlcNAc}_2$) vont être liés au Dol-P par l'action de multiples glycosyltransférases, utilisant comme substrat des nucléotides diphosphates UDP-GlcNAc et GDP-Man. Le précurseur $\text{Man}_5\text{GlcNAc}_2\text{-P-P-Dol}$ est ensuite transloqué vers la lumière du RE par l'action d'une flippase, où 4 unités mannoses et 3 unités glucoses vont lui être ajoutées. A la fin de cette étape, un précurseur composé de 14 monosaccharides ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$) est synthétisé, et est transféré en bloc sur un résidu Asn d'une protéine en cours de synthèse par l'activité du complexe oligosaccharyltransférase (OST).

Les oligosaccharyltransférases sont des complexes moléculaires multimériques ancrés dans la membrane du RE, dont deux ont été identifiés chez les eucaryotes (**Figure 16**). Ils sont nommés OST-A et OST-B, et partagent 6 sous-unités communes (Ribophorine I, Ribophorine II, OST4, OST48, DAD1 et TMEM258) (Ramírez et al. 2019). L'activité catalytique est portée par les sous-unités STT3A (OST-A) ou STT3B (OST-B). D'autres sous-unités sont spécifiques au complexe OST. Notamment, le facteur DC2 appartient à OST-A, et permet l'interaction entre la sous-unité STT3A et la protéine Sec61 du canal de translocation (Shrimal et al. 2017). Ainsi, OST-A est finement associé au translocon et permet le transfert co-translationnel du précurseur $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ sur les polypeptides en cours de synthèse. Concernant OST-B, TUSC3 ou MAGT1, des oxydoréductases, sont intégrés dans le complexe. L'absence de la sous-unité DC2 ne permet pas à OST-B d'être associé au translocon. Ainsi, il catalyse davantage le transfert du précurseur oligosaccharidique sur des polypeptides insérés dans la membrane du RE.

Après le transfert en bloc, la chaîne $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ est davantage modifiée dans la lumière du RE. L'action séquentielle des α -glucosidases I et II permet de retirer le résidu α 1-2Glc terminal et α 1-3Glc interne, formant ainsi une structure monoglucosylée. Cette

structure est un substrat pour les lectines chaperonnes calnexine (CNX) et calréticuline (CRT) intégrées dans la membrane du RE. Ces lectines ont pour fonction d'assister les glycoprotéines dans leur repliement. Elles sont associées au facteur ERp57 catalysant la formation des ponts-disulfures des protéines. Suite à cette étape, le résidu α 1-3Glc le plus interne peut être retiré par l' α -glucosidase II, ne permettant plus la reconnaissance de la protéine par CNX et CRT. Si la protéine est mal repliée à ce stade, alors l'UDP-glucose glucosyltransférase (UGGT) catalysera le transfert d'un résidu glucose, permettant à nouveau la reconnaissance par les lectines chaperonnes. Dans le cas où la protéine est toujours mal repliée après l'action de UGGT, alors elle sera dirigée vers la voie de dégradation ERAD (ER-associated degradation). En revanche, si la protéine a correctement adopté sa structure tri-dimensionnelle, elle sera exportée vers le Golgi après retrait d'un résidu mannose par une mannosidase. Une fois dans l'appareil de Golgi, la structure $\text{Man}_8\text{GlcNAc}_2$ sera davantage modifiée par les glycosyltransférases et glycosidases résidants dans les différentes citernes, permettant la formation de structures complexes et hybrides.

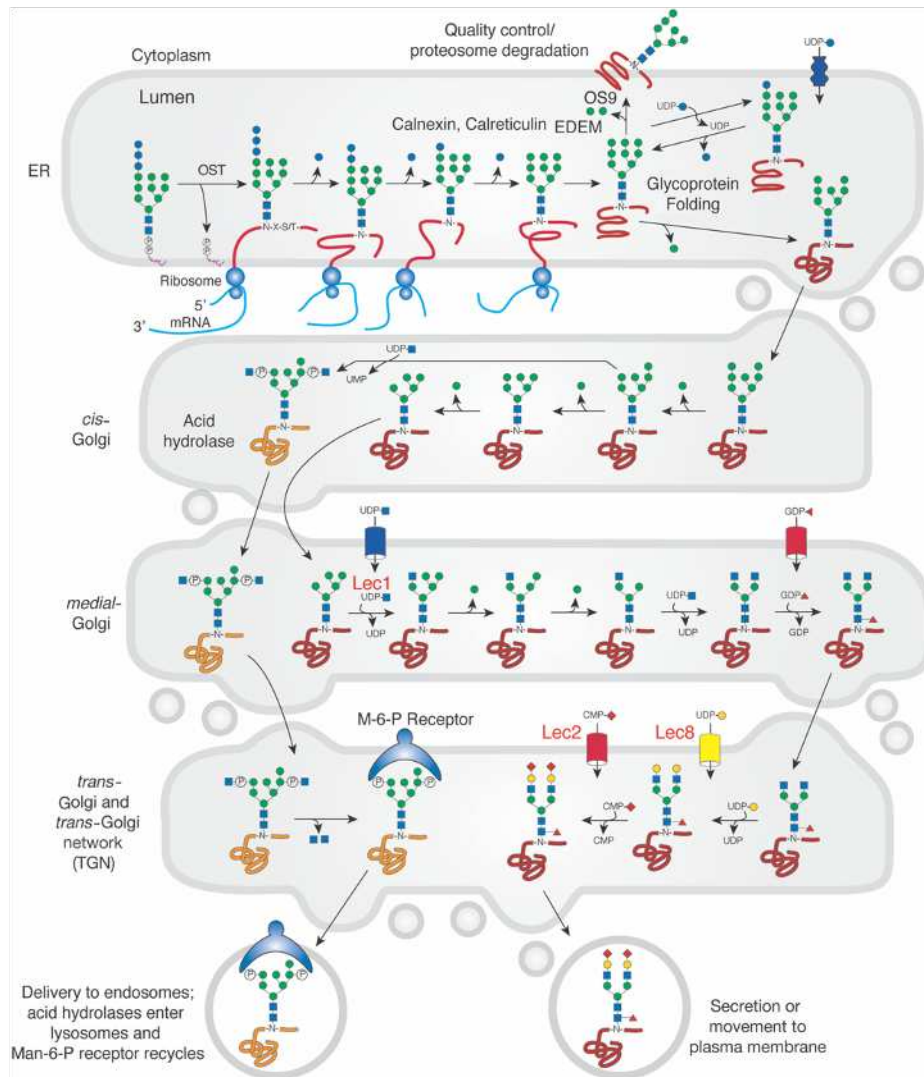


Figure 15. Voie de glycosylation des protéines en cellules eucaryotes.

La glycosylation des protéines débute dans la lumière du RE, où un précurseur oligosaccharidique composé de 14 sucres est transféré sur une asparagine acceptrice par le complexe OST. Ce précurseur est ensuite modifié par l'intervention de diverses glycosidases et glycosyltransférases du RE et de l'appareil de Golgi. Figure adaptée de Varki et al. 2022.

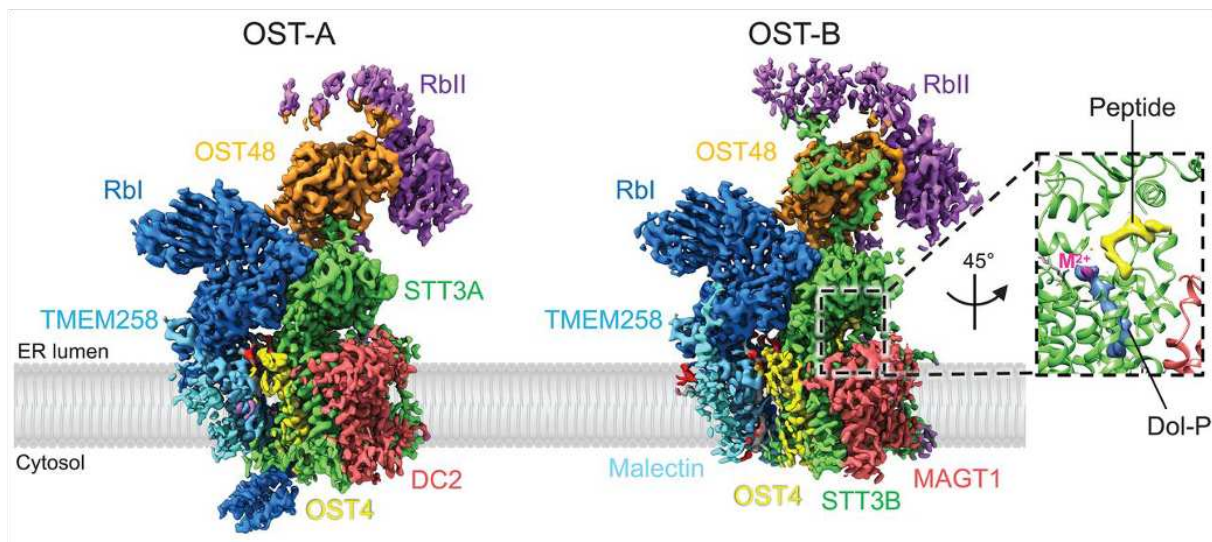


Figure 16. Structures cryo-EM des complexes oligosaccharyltransférases (OST) A et B.

Les différentes sous-unités des complexes OST-A et OST-B sont représentées par des couleurs différentes. Rb : Ribophorine. Figure adaptée de Ramírez et al. 2019.

1.7.3 Les polylactosaminoglycanes

Les chaînes de polylactosamines, ou polylactosaminoglycanes, sont de très grandes structures glucidiques, linéaires ou ramifiées, et qui peuvent être ajoutées sur des glycoprotéines et glycolipides. Ces chaînes sont composées de répétition d'unité *N*-acétyllactosamine (LacNAc) composée de galactose et de *N*-acétylglucosamine reliés par des liaisons osidiques de type β 1-4 (**Figure 17**). Chaque unité de LacNAc est reliée à une autre par des liaisons osidiques de type β 1-3. La synthèse des polylactosaminoglycanes se déroulent dans les citernes *médial* et *trans* de l'appareil de Golgi, par l'action alternée de β 1,4-galactosyltransférases (β 4GalT) et de β 3GnT. D'autres glycosyltransférases peuvent intervenir, comme des β 1,6-*N*-acétylglucosaminyltransférases. Ces enzymes transfèrent un résidu GlcNAc sur un résidu Gal interne par une liaison type β 1-6. Cela permet de générer des branchements des chaînes de polylactosamines.

Des études ont permis de mettre en évidence que les chaînes de polylactosamines étaient des structures biofonctionnelles. Notamment, les polylactosaminoglycanes forment les structures antigéniques des groupes sanguins « i » et « I ». La conversion des antigènes « i » vers « I » se déroule lors des 18 premiers mois de la vie des nouveau-nés, où les chaînes de polylactosamines vont passer d'une structure linéaire vers une structure

ramifiée grâce à l'addition de β 1-6GlcNAc sur des résidus internes de galactoses (Varki et al. 2022). Les chaînes de polylectosamines pourrait également servir de modulateur de la réponse immunitaire, puisqu'une hyperactivation des lymphocytes T est observée chez des souris déficiente en β 1,3-N-acétylglucosaminyltransférase 2 (β 3GnT2), une enzyme décrite comme la principale polylectosamine synthase cellulaire (Togayachi et al. 2007, 2010). Egalement, les polylectosaminoglycanes arborés par l'adénylate cyclase 3 (AC3) sont importants pour l'activité d'AC3 et sa localisation intracellulaire dans les neurones olfactifs sensoriels des souris (Henion et al. 2011). AC3 est responsable de la production de l'AMP cyclique, un second messager qui permet le bon fonctionnement des voies de signalisation dépendantes des récepteurs couplés aux protéines G (GPCR).

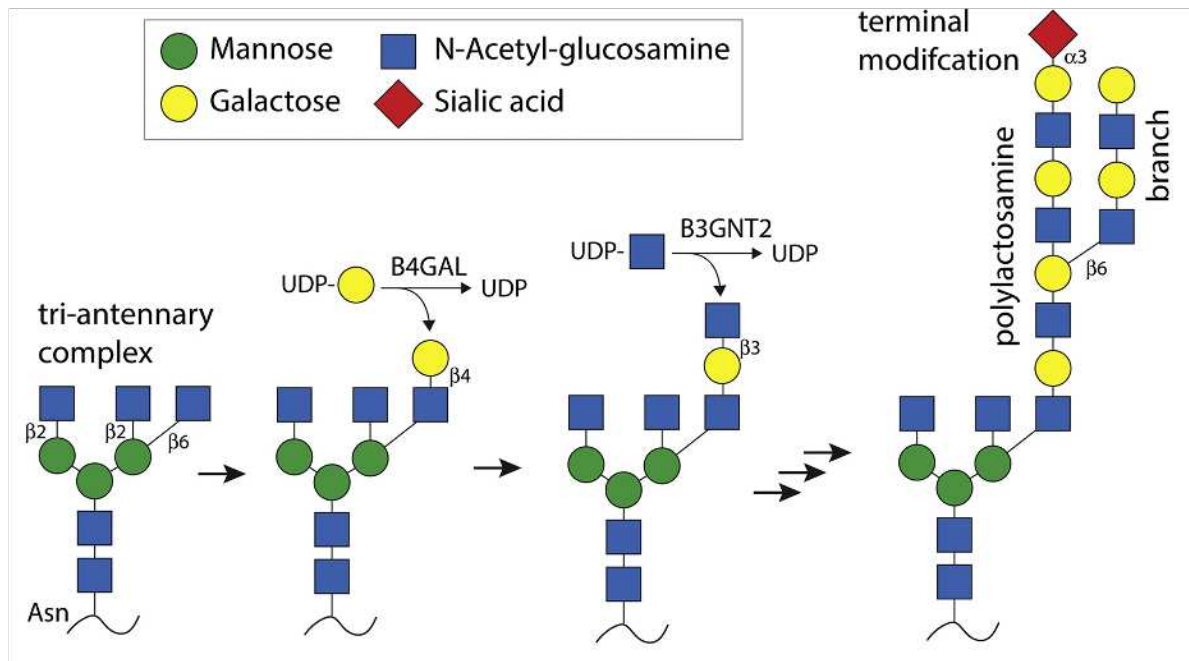


Figure 17. Mécanismes de synthèse des chaînes de polylectosamines.

Les chaînes de polylectosamines sont préférentiellement ajoutées sur des N-glycanes bi, tri ou tétra-antennaires. La synthèse de ces chaînes est possible par l'action alternée de β 1,4-galactosyltransférases (β 4GalT) et de β 1,3-N-acétylglucosaminyltransférase (β 3GnT). L'enzyme β 3GnT2 est décrite comme la polylectosamine synthase principale. Des ramifications peuvent être réalisées par des β 1,6-N-acétylglucosaminyltransférases (β 6GnT) transférant un résidu N-acétylglucosamine sur un galactose par une liaison osidique β 1-6. Figure adaptée de Kimber, 2021.

Les chaînes de polylectosamines sont très fortement étudiés dans le domaine de la cancérologie. En effet, il semblerait qu'un lien existe entre la structure de ces chaînes et la croissance des cellules tumorales et des métastases (Zhou 2003). Par exemple, une

réduction significative des métastases et de la croissances des tumeurs a été constatée chez des souris déficiente en β 6GnT-5 (MGAT5), responsable du branchement des chaines de polylactosamines (Granovsky et al. 2000). Il a été supposé que les polylactosaminoglycanes permettraient l'interaction des glycoprotéines avec des molécules d'adhésion, qui induirait l'activation de la voie de signalisation de la croissance tumorale et de l'invasion (Zhou 2003).

Les protéines membranaires associées aux lysosomes (LAMP) 1 et 2 sont les protéines cellulaires les plus modifiées par des polylactosaminoglycanes. Cependant, seulement certains sites de glycosylation arborent ces structures, signifiant qu'il existe des mécanismes de régulation (Carlsson et Fukuda 1990). Il semblerait que les polylactosaminoglycanes ajoutés sur LAMP1 et LAMP2 permettraient de maintenir l'intégrité des membranes lysosomales. La protéine Band3, exposée à la surface des erythrocytes, est aussi modifiée par des chaines de polylactosamines. Par ailleurs, de par leur structure et leurs modifications (ex : sialylation), ces chaines portées par les erythrocytes forment les erythroglycanes à l'origine des groupes sanguins ABO.

1.7.4 Fonctions des N-glycanes chez les coronavirus

1.7.4.1 Glycanes de S

La protéine spike des coronavirus est connue pour être très largement glycosylée. De plus, ces sites de glycosylation sont très conservés, suggérant un rôle important de ces modifications pour l'infection virale. Majoritairement, ces glycosylations ont été décrites comme étant des N-glycosylations. Cependant, la protéine S du SARS-CoV-2 a la particularité de posséder quelques O-glycanes de types core-1 et core-2 (Sanda et al. 2021).

Décrit précédemment (voir chapitre 1.7.1), l'ajout des N-glycosylations s'effectue sur des résidus Asn appartenant à une séquence conservée. Par conséquent, il est possible de prédire *in silico* les sites potentiels de glycosylation. Ainsi, l'analyse de la séquence protéique a permis d'identifier entre 21 et 35 sites potentiels de N-glycosylation par

monomère de protéine S en fonction des différents virus. Plus précisément, 23 sites potentiels ont été identifiés pour la protéine S du SARS-CoV et du MERS-CoV, contre 22 pour la protéine S du SARS-CoV-2. Par ailleurs, 12/22 sites potentiels de SARS-CoV-2-S sont conservés chez tous les *sarbecovirus*, accentuant l'importance de ces modifications pour le cycle viral (Allen et al. 2023). Selon les études, les sites occupés par des N-glycanes ne sont pas identiques. En effet, l'étude des glycanes de la protéine S du SARS-CoV-2 menée par Sanda et son équipe, a révélé que 17/22 sites étaient occupés par des N-glycanes lorsque la protéine spike est exprimée en cellules HEK293T (Sanda et al. 2021). En revanche, Watanabe *et al.*, ont observé des N-glycanes sur les 22 sites de spike (Watanabe et al. 2020a). Concernant le SARS-CoV, seulement 12 sites sur les 23 identifiés sont modifiés par des N-glycanes (Krokhin et al. 2003). Il est important de préciser que les systèmes d'expression utilisés pour produire la protéine S peuvent influencer les glycosylations portées. Par conséquent, ces données pourraient ne pas réellement refléter les N-glycoformes présentes dans les virions.

Une très forte micro- et macro-hétérogénéité des N-glycanes portés par la protéine S est observée selon le système d'expression utilisé. En effet, chaque site fonctionnel de SARS-CoV-2-S peut être modifié par des N-glycanes riches en mannose et complexes, avec une fréquence beaucoup plus importante pour les structures complexes lorsque spike est exprimée par des cellules humaines. En revanche, la protéine S porte davantage de structures riches en mannose lorsqu'elle est exprimée par des cellules d'insectes (Gong et al. 2021). Aussi, des structures complexes composées de répétition d'unité *N*-acétyllactosamine ont été observées sur 5 glycopeptides de la protéine S du SARS-CoV-2 (Sanda et al. 2021). Également, une étude estime que 52 % des N-glycanes de SARS-CoV-2-S sont modifiés par des fucosylations, et 15 % portent au moins un acide sialique, ajoutant davantage d'hétérogénéité et de complexité (Watanabe et al. 2020a).

Selon une étude structurale, il est estimé qu'en moyenne 40 % de la surface protéique de spike est recouverte par des glycanes, ce qui permet de générer un bouclier contre les anticorps neutralisants (Grant et al. 2020). Par conséquent, ces modifications participent à l'échappement des coronavirus à la réponse humorale. Ces boucliers glycaniques existent également chez d'autres virus, comme le VIH avec sa protéine Env, le virus HCV avec son hétérodimère E1E2, ou encore avec la protéine HA des virus

Influenza (Kobayashi et Suzuki 2012; Lavie et al. 2018; Berndsen et al. 2020). Chez le coronavirus HCoV-NL63, les N-glycanes portés par le site N358 permettent de masquer une structure du RBD en forme de boucle et importante pour l'attachement au récepteur cellulaire (Walls et al. 2016b). Une précédente étude avait mis en évidence la capacité de cette boucle à être reconnue par des anticorps neutralisants (Prabakaran et al. 2006).

Cependant, l'efficacité des boucliers glycaniques peut varier entre les virus. Par exemple, peu de glycanes sont présents au niveau du domaine RBD de la protéine S du MERS-CoV, qui est pourtant la région principalement ciblée par les anticorps neutralisants, contrairement à la spike du SARS-CoV (Watanabe et al. 2020b). De plus, des clusters d'oligomannose ont été identifiés à la surface de MERS-CoV-S, pouvant être reconnues par le système immunitaire par l'intermédiaire des lectines MBL (*Mannose-binding lectin*). De tels clusters n'ont pas été retrouvés à la surface de la protéine S du SARS-CoV et du HCoV-HKU1. Ainsi, le bouclier glycanique de MERS-CoV-S semblerait être moins efficace que celui d'autres coronavirus. A propos des *sarbecovirus* humains, 18/22 sites de N-glycosylation sont conservés entre SARS-CoV-S et SARS-CoV-2-S. Il est alors très probable que l'efficacité de leur bouclier soit semblable. Watanabe *et al.*, ont également comparés les boucliers glycaniques des coronavirus avec d'autres virus à ARN. Notamment, ils affirment que les glycanes représentent 25 % du poids moléculaire de la protéine S des coronavirus (CoV-S) et de la glycoprotéine complexe du virus Lassa (LASV GPC). Cependant, CoV-S est une protéine beaucoup plus large que LASV GPC. Par conséquent, moins de surface de CoV-S est recouverte par des glycanes, ce qui peut réduire l'efficacité du bouclier glycanique des coronavirus par rapport à celui du virus Lassa. Aussi, la fréquence de mutation du virus VIH-1 étant plus importante que celle des coronavirus, une évolution constante du bouclier glycanique a lieu chez ce virus, ce qui pourrait le rendre plus efficace que celui des coronavirus.

Au-delà de l'activité de bouclier, les glycanes qui arborent la protéine S peuvent assurer plusieurs autres fonctions. Notamment, ces modifications seraient importantes dans le maintien de la structure de spike. Par exemple, la digestion de la protéine par la PNGase F abolit la reconnaissance par des anticorps neutralisants (Koch et Kant 1990; Song et al. 2004). Dans une autre étude, il a été montré que les sites N165 et N234, juxtaposés au site RBD de SARS-CoV-2-S, permettent de stabiliser ce-dernier sous une

conformation « ouverte », nécessaire pour l'interaction avec le récepteur ACE2 (Casalino et al. 2020). Aussi chez le SARS-CoV-2, les glycosylations portés par les sites N331 et N343 (situés dans le RBD) seraient impliquées dans le maintien structural du RBD ou dans la liaison au récepteur, selon les auteurs (Li et al. 2020). En effet, l'étude montre que les mutations N331Q et N343Q engendrent une très forte baisse de l'infection des cellules 293T-hACE2 en comparaison aux conditions sauvages.

Il y a également des études qui mettent en évidence la participation des glycanes de spike dans l'attachement des particules virales à la surface des cellules hôtes. Bien que cette étape d'attachement soit principalement dépendante de l'interaction entre S-RBD et un récepteur cellulaire (voir chapitre 1.6.1), il a été observé que des cellules HeLa, n'exprimant pas constitutivement le récepteur ACE2, étaient permissives à des particules pseudovirales de types SARS-CoV lorsqu'elles surexpriment les lectines DC/L-SIGN (Han et al. 2007a). Cette étude a mis en évidence que 7 sites de N-glycosylations médient l'interaction avec ces lectines, à savoir les positions N109, N118, N119, N158, N227, N589 et N699. Également, il semblerait que la position N330 soit critique pour l'interaction avec DC/L-SIGN (Zhou et al. 2010). Chez IBV-M41, 10 sites potentiels de glycosylation sont présents dans le RBD, dont 9 portent des N-glycanes. Parmi ces sites fonctionnels, les positions N33, N59, N85, N126, N160 et N194 seraient impliquées dans l'interaction du RBD avec des résidus d'acides sialiques (Neu5Ac(α 2-3)) ajoutés aux extrémités des glycoprotéines de surface (Parsons et al. 2019). De plus, l'attachement des particules virales recombinantes à la surface des tissus de trachées de poulet via les résidus Neu5Ac(α 2-3) est moins efficace après mutations de ces sites en alanine.

Enfin, la maturation de la protéine S à la jonction S1/S2 pourrait également dépendre de certains de ces N-glycanes. Par exemple, Zheng *et al.*, ont identifié les positions N212 et N276 de la protéine S d'IBV-M41, comme étant des sites cruciaux pour la maturation de la protéine. En effet, les mutations N \rightarrow D affectent la maturation protéolytique de spike (Zheng et al. 2018). De plus, ce défaut de maturation a notamment pu être confirmé par l'incapacité des protéines S mutées à induire la fusion cellule-cellule *in vitro*, comparé à la protéine S sauvage.

Ainsi, l'ensemble de ces études a permis de montrer la complexité des glycanes qui arborent la protéine S des coronavirus, ainsi que la diversité de leurs fonctions dans l'infection virale.

1.7.4.2 Glycanes de M

La glycosylation de la protéine M est conservée chez tous les coronavirus, soit sous forme de N- ou O-glycosylation selon le genre du virus. En effet, la protéine M de certains *beta-coronavirus* arbore des O-glycanes (MHV, BCoV et HCoV-OC43), alors que celle des autres coronavirus est modifiée par des N-glycanes exclusivement.

Il a été montré que la glycosylation de M n'est nécessaire ni pour l'assemblage ni pour la réplication virale *in vitro* (De Haan et al. 1998; de Haan et al. 2003; Voß et al. 2009; Liang et al. 2019). Également, la mutation de l'asparagine acceptrice n'induit pas de changement dans le transport de la protéine M (Voß et al. 2009; Liang et al. 2019; Perrier et al. 2019).

La glycosylation de la protéine M est très conservée et n'est pas perdue au cours de l'évolution des coronavirus, ce qui suggère qu'elle pourrait avoir un rôle pour le virus. Elle pourrait moduler les interactions avec l'hôte. Il a été observé que les N-glycanes de la protéine IBV-M, portées par les sites accepteurs N3 et N6, induisent un stress du RE en cellules infectées. Le niveau d'expression de différents gènes marqueurs du stress du RE a été monitoré par RT-qPCR. Ainsi, il s'avère les gènes GRP78, CHOP et XBP1 sont surexprimés après infection de cellules Vero par du virus IBV sauvage recombinant (WTrIBV), comparé à l'infection de ces-mêmes cellules par un virus IBV recombinant déficient en N-glycosylation (rN3D/N6D) (Liang et al. 2019). Aussi, il a été montré que davantage de cellules Vero entrent en apoptose après infection par WTrIBV. La pathogénicité du virus IBV pourrait également dépendre des N-glycanes portés par la protéine M, puisqu'une léthalité légèrement plus importante, bien que non significative, a été observée sur des œufs embryonnés de poules SPF (Specific pathogen free) infectés par WTrIBV comparé à l'infection par rN3D/N6D.

Chez le coronavirus TGEV, il a été démontré que les N-glycanes de la protéine M était à l'origine d'une activité interférogénique *in vitro* sur des cellules PBMC (*Peripheral Blood Mononuclear Cells*) co-cultivées avec des cellules infectées et fixées à la glutaraldéhyde (Charley et Laude 1988; Charley et al. 1991; Laude et al. 1992). En effet, la mutation du site accepteur de M en glutamine provoque une diminution de la sécrétion des IFN α par les PBMC. Un tel effet a aussi été observé après traitement par la tunicamycine, un inhibiteur de la N-glycosylation, et également suite à un pré-traitement des particules virales TGEV par la PNGase F. Il est intéressant de préciser qu'un pré-traitement des particules virales par l'Endo-H et l'Endo-F, digérant spécifiquement les N-glycanes riches en mannose, n'a pas provoqué de diminution des titres IFN α . Par conséquent, les N-glycanes complexes de TGEV-M sont à l'origine de cette activité. Dans une autre étude, les capacités interférogéniques *in vitro* de la protéine M du coronavirus MHV-A59 ont également été mises en évidence (de Haan et al. 2003). Cependant, cette activité est beaucoup plus importante après avoir substitué le site de O-glycosylation de MHV-M en un site de N-glycosylation, par génétique inverse. Ainsi, les N-glycanes semblent être de meilleurs inducteurs que les O-glycanes. En revanche, cette activité interférogénique de MHV-M n'a pas été retrouvée *in vivo*, comme le montre un essai sur des souris C57BL/6.

Enfin, il se pourrait que les glycanes de la protéine M jouent un rôle dans le tropisme cellulaire des coronavirus. La souche A59 du coronavirus MHV est capable de se répliquer à la fois dans les cellules hépatiques et nerveuses (De Haan et al. 2003). Cependant, un virus MHV recombinant déficient en O-glycosylation se réplique moins efficacement dans le foie comparé au virus recombinant sauvage. En revanche, aucune différence significative n'a été observé dans le cerveau. Il est important de prendre en considération que les infections des souris n'ont pas été réalisées dans les mêmes conditions pour mesurer les titres viraux dans le foie et le cerveau, pouvant éventuellement expliquer les différences observées. En effet, le virus a été injecté par voie intranasale pour mesurer la réplication dans le cerveau, alors qu'il l'a été dans la cavité intrapéritonéale pour évaluer la réplication dans le foie. De manière surprenante, les titres infectieux d'un virus MHV recombinant portant des N-glycanes sur la protéine M, au lieu des O-glycanes originels, sont beaucoup plus importants dans le foie que le virus recombinant sauvage. Cependant, toujours aucune différence n'a été constatée dans le

cerveau. Par conséquent, il se pourrait que les glycanes portés par la protéine M puissent intervenir dans le tropisme cellulaire des coronavirus, et plus particulièrement les N-glycanes.

1.7.4.3 *N-glycanes de nsp3/nsp4*

Parmi les 16 protéines nsp encodées par le génome des coronavirus, seules nsp3 et nsp4 présentent des sites potentiels de N-glycosylation. Plusieurs études se sont concentrées sur les protéines nsp3/nsp4 du coronavirus MHV. Des expériences de digestion par l'Endo-H et de substitutions N → A ont permis d'identifier que le site N2357 de MHV-nsp3 était modifié par des N-glycanes riches en mannose (Kanjanaethai et al. 2007). Il existe un second site potentiel sur cette protéine, N2281. En revanche, le poids moléculaire du mutant nsp3-N2281A est identique à celui de la protéine sauvage, confirmant que seul le site N2357 est fonctionnel. Cependant, la fonction de ce site est à étudier.

Également, MHV-nsp4 possède deux sites accepteurs potentiels aux positions N176 et N237. La sensibilité de ces deux sites au traitement par l'Endo-H confirme la fonctionnalité de N176 et N237 (Clementz et al. 2008). Dans cette étude, les auteurs ont testé les capacités répliquatives de virus MHV recombinants dont la protéine nsp4 a été mutée par génétique inverse, au niveau des sites de glycosylation identifiés. Ils ont observé que le virus recombinant MHV-nsp4-N176A se répliquait aussi efficacement que le virus sauvage, contrairement au mutant nsp4-N237A et double mutant nsp4-N176A/N237A, où les titres viraux étaient plus faibles. Ces données suggèrent alors que les glycosylations portées par nsp4-N237 sont importantes pour la réplication de MHV. Une seconde équipe a également observé une diminution de la croissance virale en présence des mutations nsp4-237A ou nsp4-176A/237A (Gadlage et al. 2010). Ce changement dans la cinétique de croissance a été expliqué par une réduction de la synthèse d'ARN viraux en présence de ces mutants. Par conséquent, N237 semble être important pour la réplication et la transcription du génome de MHV. Curieusement, Gadlage *et al.*, ont observé que l'insertion de mutations dans nsp4, y compris au niveau des sites N176 et N237, génère des vésicules DMV (voir chapitre 1.5.3) aux morphologies aberrantes.

2 Objectifs de thèse

Les coronavirus sont une grande famille de virus émergents. Sur la période des vingt dernières années, trois coronavirus humains hautement pathogènes ont émergé. Le SARS-CoV en 2002, le MERS-CoV en 2012, et le SARS-CoV-2 à la fin de l'année 2019. Ces émergences virales ont été à l'origine d'épidémies à travers le monde, notamment le SARS-CoV-2 responsable de la maladie du COVID-19. Ces virus constituent une menace pour la santé publique. L'approfondissement des connaissances scientifiques concernant la biologie de ces virus est donc nécessaire.

Ce projet de thèse se concentre sur la protéine de membrane (M) des coronavirus MERS-CoV et SARS-CoV-2, qui fait figure de constituant protéique majeur de la particule virale et dont l'importance dans les mécanismes d'assemblage a été démontrée. La protéine M est une protéine transmembranaire de type III, dont le domaine N-terminal, ou ectodomaine, est glycosylé par des N- ou O-glycanes selon le genre viral. Ces glycosylations sont conservées au cours de l'évolution des coronavirus, indiquant un rôle important pour le cycle infectieux. L'étude de la séquence en acide-aminé du domaine N-terminal de la protéine M du MERS-CoV et du SARS-CoV-2, indique qu'un site potentiel de N-glycosylation (Asn-X-Ser/Thr, où X correspond à n'importe quel acide-aminé, sauf une proline) est présent en position N3 pour MERS-CoV-M et N5 pour SARS-CoV-2-M.

Dans une précédente étude, il a été observé que la protéine M du MERS-CoV présentait un profil de migration particulier en western-blot, avec la détection de plusieurs espèces (**Figure 18**) (Perrier et al. 2019). Cependant, une espèce unique a été détectée après mutation du résidu N3 en glutamine, indiquant que les autres espèces détectées correspondaient à des formes conjuguées à des N-glycanes (Perrier et al. 2019).

L'objectif de cette thèse était de mieux comprendre les mécanismes de glycosylation des protéines M du MERS-CoV et du SARS-CoV-2 et le rôle de ces glycanes.

Dans un premier temps, nous avons cloné la séquence de la protéine M du SARS-CoV-2 dans un vecteur d'expression et montré qu'elle était glycosylée de la même façon

que celle du MERS-CoV grâce à des digestions par des N-glycosidases et par la mutation du site N5 en glutamine.

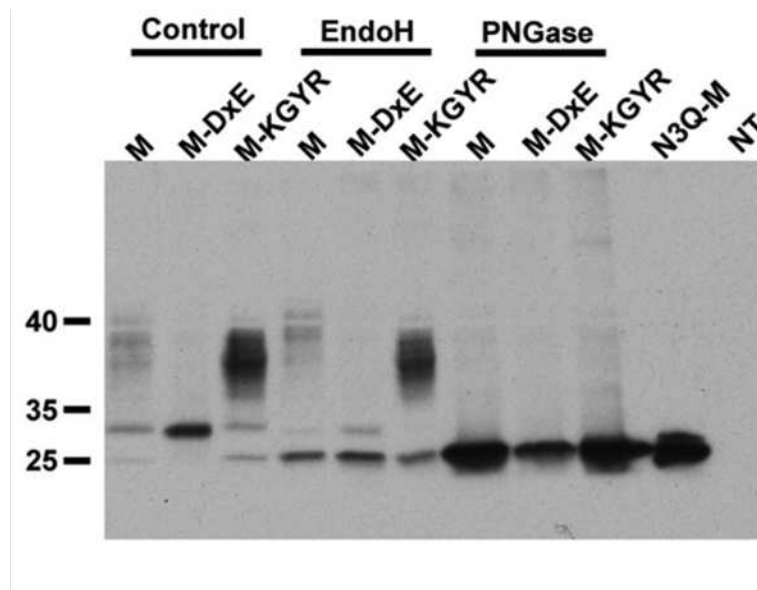


Figure 18. Profil de migration de la protéine M du MERS-CoV.

MERS-CoV-M présente un profil de migration particulier en western-blot, avec la détection d'une espèce non-glycosylée migrant vers 25 kDa, suivi d'une espèce intermédiaire d'une taille de 30 kDa et conjuguée à des N-glycanes sensibles à l'Endo-H (oligomannose), et une dernière espèce migrant sous la forme d'une bande diffuse et associée à des N-glycanes complexes (PNGase F sensible / Endo-H résistant). Figure adaptée de Perrier et al. 2019.

La protéine M du coronavirus aviaire IBV possède deux sites fonctionnels de N-glycosylation dans son domaine N-terminal. Une étude a affirmé avoir précipité la protéine IBV-M à l'aide des lectines *Lycopersicon esculentum* (LEL) et *Datura stramonium* (DSL) (Machamer, 1990). Ces lectines reconnaissent spécifiquement les chaînes de polylactosamines, de larges structures glucidiques constituées de plusieurs répétitions d'unité N-acétyllactosamine [Gal β 1-4GlcNAc] β 1-3 (Yamashita et al. 1987; Oguri 2005). Dans ce projet, nous avons montré la présence de chaînes de polylactosamines sur la protéine M du MERS-CoV et du SARS-CoV-2 en condition d'expression transitoire et infectieuse. Nous avons eu recours à l'utilisation des lectines mentionnées ci-dessus, ainsi que d'une lignée cellulaire déficiente pour une glycosyltransférase impliquée dans la voie de biosynthèse des chaînes de polylactosamines. D'autre part nous avons aussi essayé d'identifier des séquences importantes pour l'addition de chaînes de polylactosamines par mutagenèse dirigée. Seuls des résidus importants pour la glycosylation de la protéine ont pu être mis en évidence. La réalisation de chimères avec la protéine E1 du virus de

l'hépatite C ou avec ORF3a suggère que la protéine M ne contient probablement aucun signal spécifique pour cette modification.

La N-glycosylation des protéines est un processus qui est organisé au sein des cellules eucaryotes. En effet, la synthèse s'initie au sein de la lumière du RE puis s'achève à travers les diverses citernes de l'appareil de Golgi. Celle-ci débute par le transfert en bloc, côté luminal, d'un précurseur oligosaccharidique composé de 14 sucres ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$), préalablement lié à un lipide polyisoprénique ancré dans la membrane du RE. Ce transfert en bloc est permis grâce à des oligosaccharyltransférases (OST). Deux complexes OST ont été décrits chez les eucaryotes, OST-A et OST-B. Le complexe OST-A est directement associé à la protéine Sec61 du canal de translocation, permettant de transférer le précurseur lors de la synthèse du polypeptide (Shrimal et al. 2015, 2017). En revanche, OST-B n'est pas associé au canal de translocation, transférant, alors, principalement le précurseur de manière post-traductionnelle sur les polypeptides. La machinerie OST responsable de la glycosylation de la protéine M des coronavirus n'est pas connue. Nous l'avons identifiés à l'aide de lignées cellulaires déficientes.

3 Matériels et méthodes

3.1 Culture cellulaire

Des cellules Huh-7 ont été cultivées à 37 °C avec 5 % de CO₂ dans du milieu DMEM (Dulbecco's Modified Eagle Medium) supplémenté avec 10 % de sérum de veau fœtal et 2mM de Glutamax (Life technologies).

3.2 Plasmides

L'ensemble des réactions PCR suivantes ont été réalisées avec le Q5® High-Fidelity 2X Master Mix (New England Biolabs). Tous les clonages ont été réalisés en insérant les produits PCR entre les sites de restriction EcoRI et BamHI des vecteurs d'expression pcDNA3.1 contenant la séquence codant le tag V5.

Construction des vecteurs pcDNA3.1-V5-SARS-CoV-2-M et pcDNA3.1-V5-SARS-CoV-2-(N5Q)-M

Des extraits d'ARN totaux ont été collectés à partir de cellules infectées par le SARS-CoV-2. Une étape de RT-PCR a permis d'obtenir des ADN complémentaires (ADNc). Les ADNc correspondants à la séquence codante de la protéine M ont été spécifiquement amplifiés avec le primer sens Fw-1 5'-cgggatccgcagattccaacggtactatta-3' et antisens Rv-1 5'-cagaattcttactgtacaagcaaagcaata-3'.

La mutation N5Q a été insérée dans pcDNA3.1-V5-SARS-CoV-2-M par mutagenèse dirigée par PCR à l'aide du primer sens Fw-2 5'-cgggatccgcagattccaaggtactattaccggtt-3' et antisens Rv-1.

Génération des mutants du domaine N-terminal de M

Pour substituer les résidus du domaine N-terminal en alanine, la séquence codant la protéine M sauvage a été amplifiée en deux fragments PCR. Pour la protéine M du SARS-CoV-2, les deux fragments PCR ont ensuite été fusionnés avec les primers suivants : Fw-3 (5'-attgacgtcaatgggagttt-3') et Rv-2 (5'-cagaattcttactgtacaagcaaagcaata-3'). Pour le premier cycle de PCR pour SARS-CoV-2-M, les primers utilisés ont été listés dans le tableau 2.

Pour MERS-CoV-M, deux fragments PCR contenant les mutations ont également été générés, les primers utilisés pour le premier cycle de PCR sont aussi listés dans le tableau 1. Les deux fragments ont ensuite été fusionnés à l'aide des primers suivants : Fw-4 (5'-cgggatccaatatgacgcaactcact-3') et Rv-3 (5'-cagaattcctaagctcgaagcaatgcaa-3').

Génération des constructions M-linker

Pour générer la construction pcDNA3.1-V5-MERS-M-linker, deux fragments PCR ont été produits à l'aide des primers Fw-4 avec Rv-4 (5'-cttaatagttgtatttaatttatagtttccaatacggtagcggttgatgcgtatagatcaattggataaacg-3') et Fw-5 (5'-taaattaaatacaactattaaggacctggactacaaggacgacgatgacaagccaaaaggaggcttcccagataatctc-3') avec Rv-3. Les deux fragments PCR ont ensuite été fusionnés avec les primers Fw-4 et Rv-3. La construction pcDNA3.1-V5-MERS-M-linker a ensuite servi de matrice pour insérer les mutations N3Q et N88Q.

Pour le mutant pcDNA3.1-V5-N3Q-MERS-M-linker, un cycle PCR suffit avec les primers suivants Fw-6 (5'-cgggatcccaaatgacgcaactcactga-3') et Rv-3.

Pour générer pcDNA3.1-V5-N88Q-MERS-M-linker, deux cycles PCR ont été réalisés avec les primers suivants : Fw-4 avec Rv-5 (5'-ccttaatagttgtttgtaatttatagtttcc-3') et Fw-7 (5'-ggaaactataaattacaacaactattaagg-3') avec Rv-3. Les deux fragments PCR ont été fusionnés avec les primers Fw-4 et Rv-3.

Pour introduire la double mutation N3Q/N88Q dans la construction pcDNA3.1-V5-MERS-M-linker, deux fragments PCR ont été amplifiés avec les primers suivants : Fw-6

avec Rv-5 et Fw-7 avec Rv-3. Les fragments PCR ont ensuite été fusionnés avec les primers Fw-6 et Rv-3.

L'insertion de la mutation E9D18A dans la construction pcDNA3.1-V5-MERS-M-linker et pcDNA3.1-V5-N3Q-MERS-M-linker a été réalisée comme décrite précédemment pour la construction sans la séquence linker. Le plasmide pcDNA3.1-V5-E9D18A-MERS-M-linker a ensuite servi de matrice pour introduire la mutation N88Q comme décrit pour la construction pcDNA3.1-V5-N88Q-MERS-M-linker.

Génération des chimères E1-M

Pour la chimère E1-M, la séquence codant E1 a été insérée dans le primers sens utilisé pour la réaction PCR. La séquence codant la protéine M du SARS-CoV-2 a été amplifiée avec les primers Fw-8 (5'-acggatccgccaccatggctcaggtgaagaatagcagtggtcctctacgttgaagagcttaaaaagct-3') et Rv-6 (5'-ccgaattcctgtacaagcaagcaatattg-3'). La séquence correspond à E1 est indiquée en vert et celle de SARS-CoV-2-M l'est en bleue.

Génération des chimères M-ORF3A

La séquence codant ORF3a a été amplifiée à partir des ADNc de cellules infectées en utilisant les primers Fw-9 (5'-actctgagccagtgctcaaag-3') et Rv-7 (5'-gaaacgaatgagtacataagt-3'). Un second cycle PCR a permis d'insérer les sites de restriction *EcoRI* et *BamHI* introduits dans les primers suivants : Fw-10 (5'-tcggatccgccaccatggattgtttatgagaa-3') et Rv-8 (5'-ccgaattccaaggcagcgtagtagtcg-3').

Pour construire la chimère Nt-M/ORF3a, la séquence codant la protéine M a été amplifiée avec Fw-3 et Rv-9 (5'-gccaacaataagccatccggtaatagtagcttggaaatc-3'). La séquence codant ORF3a a été amplifiée avec Fw-11 (5'-gattccaacggctactattaccggatggcttattgttggc-3') et Rv-8. Les deux fragments PCR ont ensuite été fusionnés avec les primers Fw-3 et Rv-8.

Concernant la chimère TM1-M/ORF3a, la séquence codant la protéine M a été amplifiée avec les primers Fw-3 et Rv-10 (5'-ctcttttgagggtatgataaattgtagaagacaaatcc-3') et la séquence codant ORF3a avec les primers Fw-12 (5'-ggatttgcttctacaatttATCATAACCCTCAAAAAGAG-3') et Rv-8. Les primers Fw-3 et Rv-8 ont permis la fusion des deux fragments PCR.

Construction des plasmides pX330

La séquence ADN cible a été introduite dans un vecteur pX330 comme décrit par Ran et al 2013. Brièvement, 10 μ M d'oligonucléotides ont été hybridés et phosphorylés avec la T4 polynucléotide kinase (NEB). Ensuite, les oligonucléotides et le vecteur ont été digérés par l'enzyme de restriction BbsI et ligués en une seule étape. Les ADN linéaires résiduels ont été éliminés par traitement avec une exonucléase, et des cellules compétentes Top10 ont été transformées avec les produits ligués. Les oligonucléotides suivants ont été utilisés pour :

β 3GNT2 : 5'-caccgactactgccaacactgacctg-3' et 5'-aaaccaggtcaggttcgcagtagtc-3'

STT3A : 5'-caccgtactcacggatcaaactca-3' et 5'-aaactgagtttgatccgtgagtagac-3'

STT3B : 5'-caccgtacagcaaaagagtctacat-3' et 5'-aaacatgtagactcttttgctgta-3'.

3.3 Génération des lignées KO

Des cellules Huh-7 ont étéensemencées en plaque 6 puits et ont été transfectées par 1 μ g de vecteur pX330 codant pour un petit ARN guide ciblant *β 3GNT2*, *STT3A* ou *STT3B*, ou par un vecteur vide pX330 et 50 ng de plasmide portant un gène de résistance à la puromycine. Les cellules ont été transfectées à l'aide de l'agent de transfection TransIT®-LT1 (Mirus Bio) selon le protocole du fournisseur. 24 h post-transfection, 5 μ g/ml de puromycine ont été ajoutés dans le milieu de culture afin d'induire une pression de sélection. La puromycine a été renouvelée tous les jours pendant 4 jours. Pour les KO *STT3A* et *STT3B* les populations de cellules ont été utilisées pour exprimer les protéines M. Le niveau d'expression des protéines STT3a et STT3b a été contrôlé par western-blot à l'aide d'anticorps polyclonaux anti-STT3a (Fisher Scientific) et anti-STT3b

(Sigma). Pour les cellules $\beta 3GNT2$ -KO, des clones ont été isolés par dilutions limites des populations de cellules après sélection à la puromycine. Faute d'anticorps anti- $\beta 3GnT2$ fonctionnel, le niveau d'expression de la protéine n'a pu être contrôlé par western-blot. L'ADN génomique des clones isolés a été extrait avec le kit QIAamp® DNA mini kit (QIAGEN) et a été séquencé au niveau de la zone éditée. Un clone produisant une protéine $\beta 3GnT2$ tronquée à la position 109 a été sélectionné pour les expériences.

3.4 Traitements aux glycosidases

Des cellules Huh-7 ont étéensemencées en plaque 6 puits, et ont été transfectées par 1 μ g de plasmides pcDNA3.1 exprimant les constructions taguées V5-MERS-M, V5-N3Q-MERS-M, V5-SARS2-M ou V5-N5Q-SARS2-M avec l'agent de transfection TransIT®-LT1 (Mirus Bio) selon le protocole du fournisseur. 24 h post-transfection, les cellules ont été lysées sur glace dans du tampon de lyse B1 (50 mM Tris-HCl pH 7,5 ; 100 mM NaCl₂ ; 2 mM EDTA ; 1 % Triton X-100 ; 0,1 % SDS) contenant un cocktail d'inhibiteurs de protéase (Roche) et 1mM de PMSF. Les lysats cellulaires ont ensuite été centrifugés à 14 000 rpm pendant 5 min à 4 °C afin de les nettoyer des débris cellulaires. Ensuite, les lysats ont été traités avec de l'eau (mock), à l'Endoglycosidase H (Endo-H) (NEB) ou à la PNGase F (NEB) selon les instructions du fournisseur. Les protéines ont été séparées par SDS-PAGE et transférées sur une membrane de nitrocellulose. Les protéines M ont été détectées avec un anticorps monoclonal anti-V5 (Invitrogen, R96025) et un anticorps secondaire anti-souris couplé à la peroxydase de raifort (JacksonImmunoResearch 115-035-146). La détection protéique a été réalisée par chimioluminescence (Pierce).

Pour les quantifications des formes non-glycosylées de la protéine M, les lysats des cellules ayant exprimées les différents mutants ont été traités à l'eau (mock) ou à la PNGase F, puis analysés par western-blot. La bande correspondant à la forme non-glycosylée a été quantifié dans les deux conditions via la fonction « quantification des bandes » du logiciel ImageJ. Les conditions traitées par la PNGase F ont permis de déterminer la quantité totale de protéine M dans les échantillons.

3.5 Précipitations par les lectines

Des cellules Huh-7 ont été ensemencées en plaque 6 puits et ont été transfectées avec 1 µg de vecteur pcDNA3.1 exprimant les constructions taguées V5-MERS-M, V5-N3Q-MERS-M, V5-SARS2-M ou V5-N5Q-SARS2-M avec l'agent de transfection TransIT®-LT1 (Mirus Bio) selon les instructions du fournisseur. 24 h post-transfection, les cellules ont été lysées sur glace avec du tampon de lyse B1 supplémenté par un cocktail d'inhibiteurs de protéases (Roche) et 1mM de PMSF. Ensuite, les lysats ont été centrifugés à 14 000 rpm pendant 5 min à 4 °C afin d'éliminer les débris cellulaires. 50 µl d'échantillon ont été conservés pour une analyse par western-blot (input). Le reste a été incubé avec des billes d'agarose conjuguées avec la lectine DSL (*Datura stramonium* lectin) ou LEL (*Lycopersicon esculentum* lectin) (Eurobio) sur une roue à 4°C pendant la nuit. Le lendemain, les billes ont été nettoyées 3 fois avec du tampon B1 et resuspendues dans du tampon de charge Laemmli 2 X contenant 50 mM de DTT. Les billes ont ensuite été chauffées à 70 °C pendant 10 min avant d'être chargées sur SDS-PAGE pour une analyse par western-blot. Les protéines ont été révélées à l'aide d'un anticorps monoclonal anti-V5 et d'un anticorps secondaire anti-souris couplé à la peroxydase de raifort. Les protéines ont été détectées par chimioluminescence (Pierce).

3.6 Infection

Des cellules β 3GNT2-KO ont été ensemencées en plaque 6 puits et ont été infectées par le MERS-CoV ou le SARS-CoV-2 à une MOI de 0,1 pendant 16 h. 48 h avant l'infection par le SARS-CoV-2, les cellules ont été transduites avec des lentivirus exprimant ACE2. 16h post-infection, les cellules ont été lysées dans du tampon B1 et les lysats collectés ont été centrifugés à 14 000 rpm pour éliminer les débris cellulaires. Les lysats cellulaires ont ensuite été chauffés à 95 °C pendant 20 min avant de sortir du laboratoire P3. Les protéines ont été séparées par SDS-PAGE et analysées par western-blot. Les protéines M du MERS-CoV et du SARS-CoV-2 ont été détectées à l'aide d'anticorps polyclonaux anti-M spécifiques et d'un anticorps secondaire anti-lapin couplé à la peroxydase de raifort (JacksonImmunoResearch 111-035-144). Les protéines ont été détectées par chimioluminescence (Pierce).

3.7 Immunofluorescence

Les anticorps primaires et secondaires utilisés sont résumés dans le tableau 2.

Des cellules Huh-7 ont étéensemencées sur des lamelles en plaque 24 puits et ont été transfectées avec 250 ng de plasmide pcDNA3.1 exprimant les différentes constructions mutées taguées V5 ou exprimant HCV-E1 (souche JFH1). 24 h post-transfection, les cellules ont été fixées avec 3 % de paraformaldéhyde pendant 15 min, puis les membranes ont été perméabilisées avec 0,1 % de Triton X-100 pendant 5 min. Ensuite, les cellules ont été bloquées pendant 30 min avec 5 % de sérum de cheval dans du PBS, suivi de 30 min d'incubation avec une solution d'anticorps primaires préparée dans le tampon de blocage. Enfin, les cellules ont été rincées 3 fois avec du PBS et incubées pendant 30 min à l'obscurité avec une solution d'anticorps secondaires couplés à un fluorophore. Les noyaux ont été marqués avec 1µg/ml de 4,6-diamidino-2-phenylindole (DAPI).

Les images ont été acquises à l'aide d'un microscope confocal à balayage laser LSM 880 (Zeiss) avec un objectif à immersion au grossissement 63x.

3.8 Perméabilisations sélectives

Des cellules Huh-7 ont étéensemencées en plaque 24 puits, puis ont été transfectées avec 500 ng de plasmide pcDNA3.1 contenant la séquence codante de V5-M ou des mutants. 24 h post-transfection, ces cellules ont été fixées avec 3 % de paraformaldéhyde, précédant une perméabilisation totale des membranes cellulaires par 0,1 % de Triton X-100 ou partielle par un cycle de congélation/décongélation à -80 °C dans du PBS. Enfin, les cellules ont été bloquées par 5 % de sérum de chèvre/PBS pendant 30 min puis ont été traitées comme décrit ci-dessus.

Tableau 2. Oligonucléotides utilisés pour le premier cycle PCR.

	Mutants	Sens (5'-3')	Antisens (5'-3')
SARS-CoV-2-M	8ITV₁₀-M	attgacgtcaatgggagttt	agcggcagcagtagccttggaaatctgcggatcccg
		ctgctgccgctgaagagcttaaaaagctcc	cagaattcttactgtacaagcaagcaata
	11EEL₁₃-M	attgacgtcaatgggagttt	agccgctgcaacggtaaatagtagcgttgg
		tcagcgggctaaaagctccttgaacaatgg	cagaattcttactgtacaagcaagcaata
	14KKL₁₆-M	attgacgtcaatgggagttt	agccgctgcaagctcttcaacggtaaatagtagc
		tcagcgggctcttgaacaatggaacctagtaatagg	cagaattcttactgtacaagcaagcaata
	17LEQ₁₉-M	attgacgtcaatgggagttt	tgctgcagcagagcttttaagctcttcaac
		cgctgcagcatggaacctagtaaatagtttcc	cagaattcttactgtacaagcaagcaata
	E_{11,12,18A}	attgacgtcaatgggagttt	attgccaagtagcttttaagcgtgcaac
		aaaaagctccttgcgcaatgg	cagaattcttactgtacaagcaagcaata
MERS-CoV-M	6QLT₈-M	attgacgtcaatgggagttt	tcagcggctgccgcatattggatcccgtagaatc
		cagccgctgaggcgcagattattgccattattaaa	cagaattcctaagctcgaagcaatgcaa
	9EAQ₁₁-M	attgacgtcaatgggagttt	atcggcggcagtgagttgcgtcatattggatcc
		gcgggggcgattattgccattattaaagactggaa	cagaattcctaagctcgaagcaatgcaa
	12IIA₁₄-M	attgacgtcaatgggagttt	aatggctgcagcctgcgcctcagtgagttgcgtcatatt
		gcaggctgcagcattattaaagactggaactttgcatgg	cagaattcctaagctcgaagcaatgcaa
	15IHKD₁₈-M	attgacgtcaatgggagttt	ggctgcagcagcggcaataatctgcgcctcagtg
		gctgctgcagcctggaactttgcatggtcctgat	cagaattcctaagctcgaagcaatgcaa
	E9A-M	attgacgtcaatgggagttt	tctgcggcagtgagttgcgtcatattgg
		tcggcgagattattgccattattaaaga	cagaattcctaagctcgaagcaatgcaa
	E9D-M	attgacgtcaatgggagttt	cgcatcagtgagttgcgtcata
		cactgatgagcagattattgcc	cagaattcctaagctcgaagcaatgcaa
	E9D18A-M	attgacgtcaatgggagttt	caaagttccagcttaataatggcaataatctgcggcagtg
		cactgaggcagattattgccattattaaagcgtggaactttg	cagaattcctaagctcgaagcaatgcaa

Tableau 3. Anticorps primaires et secondaires utilisés pour les expériences d'immunofluorescence.

Nom	Espèce	Référence
IgG2a anti-V5	Souris	Invitrogen 37-7500 Clone 2F11F7
IgG anti-TGN46	Mouton	Biorad AHP500GT
Anti-calnexine		Abcam Ab22595
Anti-HCV-E1 (A6)	Humain	Mesalam <i>et al.</i> , 2018
Anti-MERS-M	Lapin	Proteogenix
IgG anti-SARS-CoV-2-M	Lapin	Abnova PAB31758
IgG anti-souris couplé cyanine-3	Âne	JacksonImmunoResearch 715-165-151
IgG anti-mouton couplé Alexa488	Âne	JacksonImmunoResearch 713-545-147
IgG anti-mouton couplé cyanine-3	Âne	JacksonImmunoResearch 713-165-147
IgG anti-lapin couplé Alexa488	Âne	JacksonImmunoResearch 711-545-152
IgG anti-lapin couplé cyanine-3	Âne	JacksonImmunoResearch 711-165-152
IgG anti-humain couplé Alexa488	Âne	JacksonImmunoResearch 709-545-149

4 Résultats

4.1 Les protéines M du MERS-CoV et du SARS-CoV-2 sont N-glycosylées

Dans une étude précédente réalisée dans le laboratoire, la protéine M du MERS-CoV présentait un profil de migration particulier en western-blot (Perrier et al. 2019). Il a été montré que ce profil de migration particulier est causé par le statut de glycosylation de la protéine qui présente un seul site de N-glycosylation. En effet, la substitution du résidu asparagine (N3) en glutamine (N3Q-M), empêche ainsi la glycosylation de la protéine et engendre la migration de MERS-CoV-M sous forme d'une bande unique autour de 25 kDa. La protéine M du SARS-CoV-2 possède également un site de glycosylation unique (N5). Nous nous sommes donc intéressés, dans un premier temps, au profil de migration de la protéine M du SARS-CoV-2 en western-blot.

Pour cela, des cellules Huh-7 ont été transfectées par un vecteur pcDNA3.1 contenant la séquence codant la protéine SARS-CoV-2-M avec une étiquette V5. Comme contrôle, nous avons muté le seul site potentiel de N-glycosylation en substituant l'asparagine N5 en glutamine (N5Q). En parallèle, des cellules Huh-7 ont également été transfectées afin d'exprimer la protéine M sauvage et mutée au niveau du site de glycosylation (N3Q-M) du MERS-CoV. 24h post-transfection, les lysats cellulaires ont été collectés, puis les protéines ont été séparées sur SDS-PAGE 12 % et analysées par western-blot avec un anticorps anti-V5 (**Figure 19A**). Il s'avère que SARS-CoV-2-M (panel gauche) présente un profil de migration identique à celui observé avec MERS-CoV-M (panel droit). En effet, nous pouvons constater que la protéine M-WT du SARS-CoV-2 migre sous la forme de 3 espèces. En revanche, la protéine mutée N5Q-M migre comme une seule espèce vers 25 kDa. Par conséquent, ces résultats démontrent que la protéine M du SARS-CoV-2 est N-glycosylée sur le site N5.

Afin de caractériser les formes glycoconjuguées de SARS-CoV-2-M, les lysats cellulaires ont été digérés par l'Endo-H et la PNGase F. L'Endo-H digère spécifiquement les N-glycanes qui n'ont pas été modifiés par l' α -mannosidase II, c'est-à-dire, les oligomannoses et certaines structures hybrides. En revanche, la PNGase F est capable de digérer tous les types de N-glycanes. Nous pouvons observer que la forme de SARS-CoV-2-M migrant vers 30 kDa est sensible à l'Endo-H, alors que les formes migrant plus lentement sous la forme d'une bande diffuse y sont résistantes. En revanche, le traitement par la PNGase F a permis la digestion de toutes les formes de SARS-CoV-2-M, sauf l'espèce migrant autour de 25 kDa, correspondant à une forme non-glycosylée de la protéine. Ainsi, ces données démontrent qu'en cellules transfectées transitoirement, les protéines M du MERS-CoV et du SARS-CoV-2 (25 kDa) sont conjuguées à des N-glycanes riches en mannose (30 kDa) et complexes (>35 kDa).

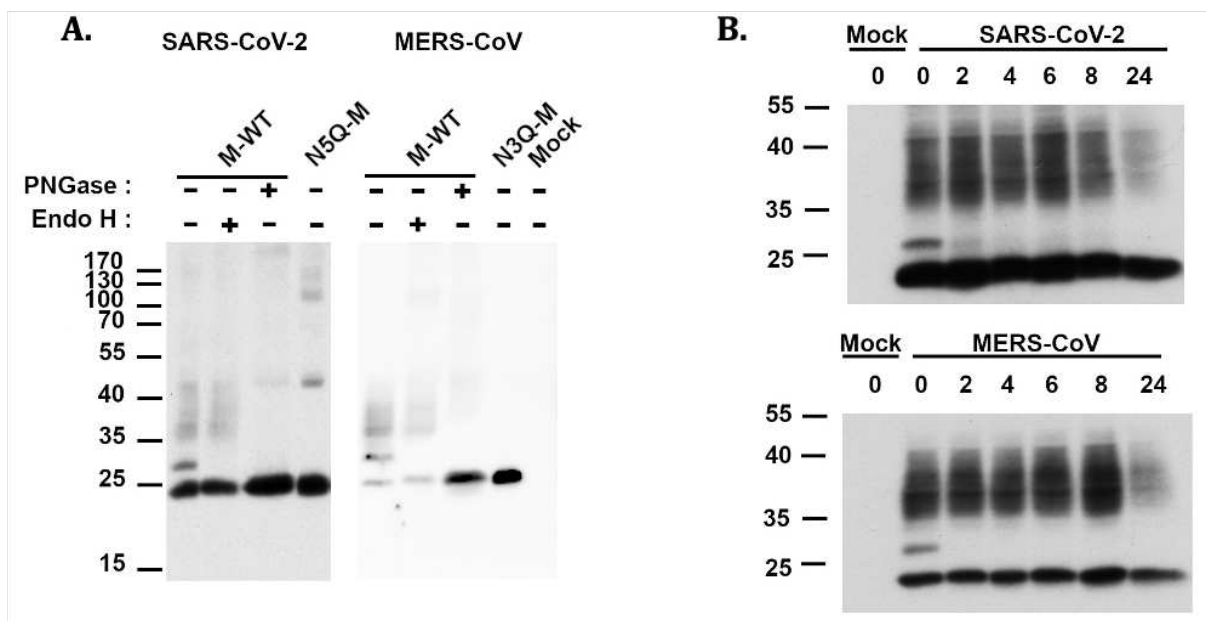


Figure 19. La protéine M du MERS-CoV et du SARS-CoV-2 est N-glycosylée.

(A). Des lysats de cellules Huh-7 transfectées pendant 24 h par un vecteur d'expression pcDNA3.1 contenant la séquence codant les protéines taguées V5 SARS-CoV-2-M et MERS-CoV-M ont été traités par l'Endo-H ou la PNGase F. Un anticorps anti-V5 a permis de révéler les protéines. Des lysats de cellules ayant exprimé les protéines déficientes en site de glycosylation (N3Q-M pour MERS-CoV ; N5Q pour SARS-CoV-2) ont été chargés comme contrôle. (B). Des cellules Huh-7 ont exprimé pendant 24 h les protéines MERS-CoV-M ou SARS-CoV-2-M avec une étiquette V5. Ces cellules ont ensuite été traitées par 100 μ M de cycloheximide afin de bloquer la synthèse protéique. Les lysats ont été récoltés aux temps indiqués (0, 2, 4, 6, 8 et 24 h). Un anticorps anti-V5 a permis de révéler les protéines. Les poids moléculaires sont exprimés en kDa. Mock : non-traité.

A l'équilibre, nous observons donc trois formes de la protéine. Nous avons voulu savoir si les formes non-glycosylées et sensibles à l'Endo-H étaient des intermédiaires de glycosylation. Pour cela, des cellules Huh-7 exprimant V5-MERS-CoV-M ou V5-SARS-CoV-2-M ont été traitées par la cycloheximide afin de bloquer la synthèse protéique, 24h-post-transfection. Les lysats cellulaires ont ensuite été collectés à différents temps de 2h à 24h après le début du traitement, puis les protéines ont été analysées par western-blot (**Figure 19B**). Nous pouvons constater qu'à partir de 2h post-traitement, l'espèce conjuguée à des N-glycanes riches en mannose et migrant vers 30 kDa n'est plus détectée, signifiant que cette forme sert de précurseur pour la synthèse des espèces conjuguées à des N-glycanes complexes. En revanche, l'espèce non-glycosylée est toujours présente jusqu'à 24h post-traitement et est relativement stable au cours du temps, signifiant que certaines protéines ne sont jamais modifiées.

Les formes conjuguées à des glycanes complexes sont parfois difficiles à observer en western-blot à cause de leur hétérogénéité, ce qui peut suggérer qu'une proportion importante de la protéine n'est pas glycosylée. Nous avons quantifié la bande correspondant à la forme non glycosylée en western-blot (**Figure 21C**). Afin de déterminer la quantité totale de protéine, une fraction de l'échantillon a été traité par la PNGase F. Après quantification, 3,32 % (+/- 1,13 %) de la protéine M du MERS-CoV et 5,73 % (+/- 2,18 %) de celle du SARS-CoV-2 ne sont pas glycosylés.

4.2 SARS-CoV-2-M et MERS-CoV-M sont modifiées par des chaînes de polylactosamines

Les chaînes de polylactosamines sont de très grandes structures linéaires ou ramifiées de glycanes qui peuvent être ajoutées sur des glycoprotéines et glycolipides. Ces chaînes sont composées de répétition d'unité *N*-acétyllactosamine (LacNAc) composée de galactose (Gal) et de *N*-acétylglucosamine (GlcNAc) reliés par des liaisons osidiques β 1-4. Chaque unité de LacNAc est reliée à une autre par des liaisons osidiques β 1-3. Des études ont permis de mettre en évidence que les chaînes de polylactosamines étaient des structures importantes, puisqu'elles sont directement ou indirectement impliquées dans de nombreuses fonctions biologiques. Notamment, les polylactosaminoglycanes forment

les structures antigéniques des groupes sanguins « i » et « I ». La conversion des antigènes « i » vers « I » se déroule lors des 18 premiers mois de la vie des nouveau-nés, où les chaînes de polylactosamines vont passer d'une structure linéaire vers une structure ramifiée grâce à l'addition de β 1-6GlcNAc sur des résidus internes de galactoses (Varki et al 2023 – essentials of glycobiology). La synthèse des polylactosaminoglycanes se déroulent dans les citernes *médial* et *trans* de l'appareil de Golgi, par l'action alternée de β 1,4-galactosyltransférases (β 4GalT) et de β 1,3-N-acétylglucosaminyltransférase (β 3GnT) (Nilsson et al. 1991; Lee et al. 2009; Kimber 2021).

Chez le coronavirus IBV, il a été montré que la protéine M est retenue au niveau du compartiment *cis*-golgien, en cohérence avec les N-glycanes ajoutés sur les sites N3 et N6 de la protéine qui restent majoritairement sensibles à l'Endo-H (Machamer et al. 1990). En revanche, la délétion du premier et second segment transmembranaire induit une fuite d'IBV-M jusqu'à la surface cellulaire. Dans ces conditions, il a été observé qu'une plus grande fraction d'IBV-M devenait résistante à la digestion par l'Endo-H. En western-blot, l'augmentation de cette fraction s'est caractérisée par la migration de la protéine sous la forme d'une bande diffuse, comme nous avons pu l'observer précédemment avec la protéine M du MERS-CoV et du SARS-CoV-2. De plus, dans leur manuscrit, Machamer *et al.*, mentionne sans le montrer que la protéine IBV-M tronquée est reconnue par les lectines *Datura stramonium* (DSL) et *Lycopersicon esculentum* (LEL), dont l'affinité pour les chaînes polylactosamines a été démontrée (Yamashita et al 1987 ; Oguri et al 2005). Par conséquent, nous nous sommes intéressés à la possible présence de polylactosaminoglycanes sur les sites N3 de MERS-CoV-M et N5 et SARS-CoV-2-M.

Dans un premier temps, des cellules Huh-7 ont été transfectées par des vecteurs pcDNA3.1 exprimant les protéines V5-MERS-CoV-M et V5-SARS-CoV-2-M sauvages et mutées au niveau de leurs sites de glycosylation respectifs (N3Q-M et N5Q-M). Les lysats cellulaires ont été collectés 24h post-transfection, puis ont été incubés avec des billes d'agarose conjuguées aux lectines LEL ou DSL. Les protéines précipitées ont ensuite été chargées sur SDS-PAGE 12 % et analysées par western-blot à l'aide d'un anticorps anti-V5. Afin de s'assurer du niveau d'expression des protéines M, un aliquot de lysat cellulaire a été déposé sur gel (**Figure 20A**). Les produits de précipitation sont visibles en **Figure 20B**. Nous pouvons constater que les formes de M glycoconjuguées à des N-glycanes

complexes sont précipitées par la lectine LEL, suggérant la présence de chaînes de polylectosamines sur la protéine M du MERS-CoV et du SARS-CoV-2 en cellules transfectées. Cependant, la lectine DSL ne semble pas reconnaître les glycanes associés à M. Afin de contrôler la qualité de cette lectine, nous avons vérifié que cette lectine était bien capable de précipiter la protéine M de IBV (données non-montrées). Par conséquent, l'absence de précipitation de MERS-CoV-M et SARS-CoV-2-M par la lectine DSL a une raison biologique. Il a été montré que DSL reconnaît davantage les chaînes de polylectosamines ramifiées, contrairement à LEL qui possède plus d'affinité pour les chaînes linéaires (Varki et al. 2022). Il a également été montré que l'affinité de la lectine DSL envers les polylectosaminoglycanes décroît en présence de certaines modifications, telles que des fucosylations et sialylations (α 2,6 acide sialique) ajoutées aux extrémités (Yamashita et al. 1987; Bojar et al. 2022).

Pour confirmer la présence de chaînes de polylectosamines sur la protéine M, nous avons généré une lignée cellulaire Huh-7 déficiente pour la β 1,3-N-acétylglucosaminyltransférase 2 (β 3GnT2) par CRISPR/Cas9 (β 3GNT2-KO). L'enzyme β 3GnT2 a été caractérisée comme la polylectosamine synthase principale (Togayachi et al. 2010). De plus, la génération d'une lignée de souris déficiente en β 3GnT2 montre que l'édition du gène β 3GNT2 n'est pas létale.

L'absence d'anticorps anti- β 3GnT2 fonctionnel ne nous a pas permis de vérifier le niveau d'expression de l'enzyme dans les populations Huh-7 β 3GNT2-KO par western-blot. Par conséquent, les populations de cellules obtenues après sélection ont été sous-clonées par dilution limite, et la région éditée du gène β 3GNT2 de chaque clone isolé a été séquencée. Un clone, dont un codon stop prématuré a été introduit en position 109, a été utilisé pour réaliser nos expériences. Ce clone et un clone contrôle (pX330) ont été transfectés afin d'exprimer la protéine M-WT taguée V5 du MERS-CoV et du SARS-CoV-2, ainsi que leurs mutants non-glycosylés N3Q-M et N5Q-M respectifs. Les lysats cellulaires ont été analysés par western-blot, 24h post-transfection (**Figure 20C**). Les différentes espèces de M identifiées précédemment sont retrouvées dans la lignée contrôle pX330. En revanche, la bande diffuse n'est plus visible lorsque la protéine M est exprimée dans la lignée β 3GNT2-KO. A la place, une espèce migrant plus lentement que l'espèce glycoconjuguée à des N-glycanes riches en mannose (30 kDa) est détectée. La différence

de poids moléculaire et la sensibilité à l'Endo-H a permis de discriminer ces deux espèces. En effet, contrairement à l'espèce de 30 kDa, celle à plus haut poids moléculaire est résistante au traitement par l'Endo-H (données non-montrées). Ainsi, cette espèce correspond bien à une forme de M associée à des N-glycanes complexes. Également, la migration de M sous la forme d'une bande à haut poids moléculaire au lieu d'une bande diffuse montre une réduction de l'hétérogénéité des N-glycanes complexes ajoutés dans la lignée $\beta 3GNT2$ -KO. Par conséquent, ces résultats indiquent que l'enzyme $\beta 3GnT2$ participe à la maturation des N-glycanes de MERS-CoV-M et SARS-CoV-2-M, ce qui confirme l'ajout des chaînes de polylactosamines.

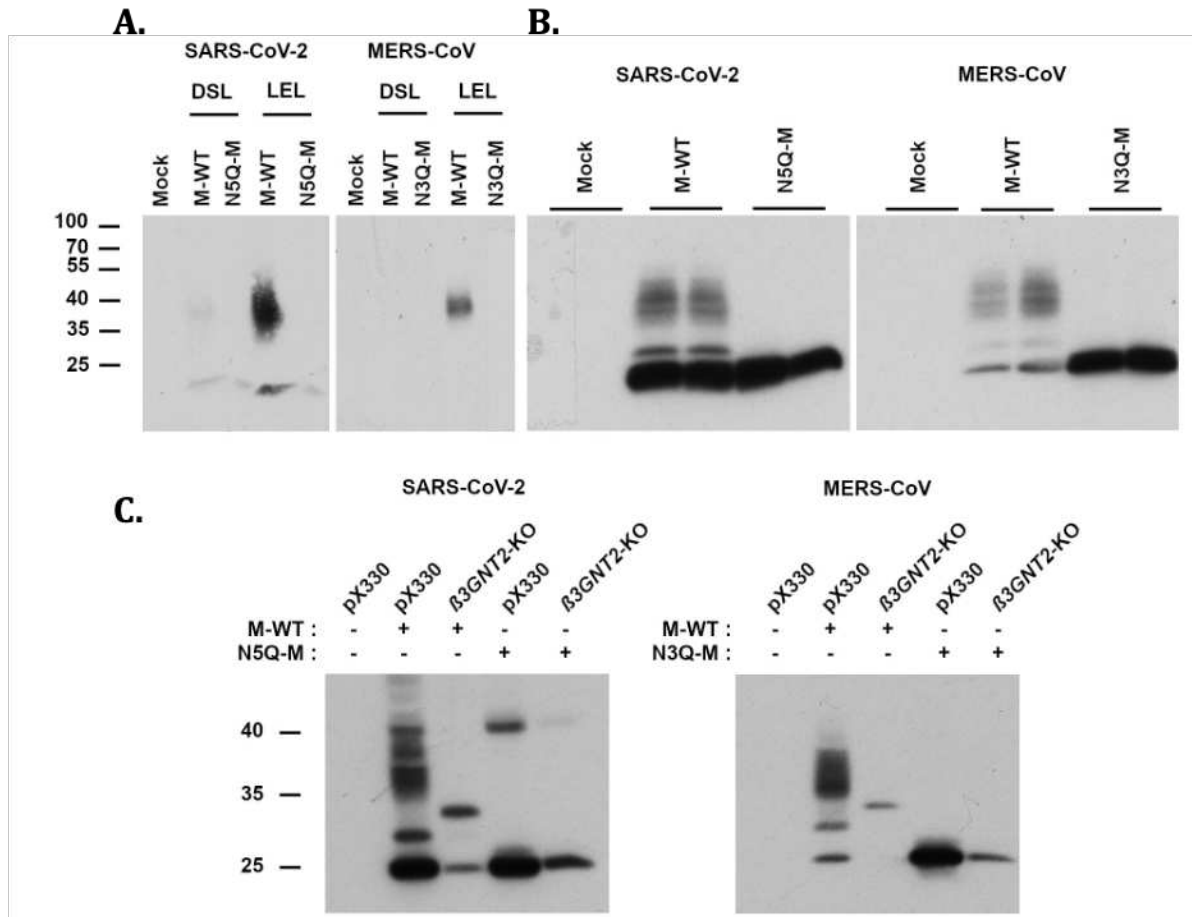


Figure 20. MERS-CoV-M et SARS-CoV-2-M sont modifiées par des chaînes de polylactosamines.

(A). Des tests de précipitation par les lectines *Datura stramonium* (DSL) et *Lycopersicon esculentum* (LEL) conjuguées à des billes d'agarose ont été effectués sur des lysats de cellules ayant exprimé pendant 24 h les protéines M taguées V5 du MERS-CoV ou du SARS-CoV-2. Les protéines ont été révélées à l'aide d'un anticorps anti-V5. (B). Un aliquot des échantillons utilisés en (A) a été déposé afin de contrôler les niveaux d'expression de M (INPUT). (C). Des cellules Huh-7 contrôles (pX330) ou déficientes pour la β 3GnT2 (β 3GnT2-KO) ont été transfectées pendant 24 h afin d'exprimer les protéines M taguées V5 mentionnées. Les lysats cellulaires ont été chargés sur gel SDS-PAGE et les protéines ont été révélées par un anticorps anti-V5. Les poids moléculaires sont exprimés en kDa.

4.3 Les résidus acides proches du site de glycosylation sont importants pour la N-glycosylation de la protéine M

Les diverses glycosyltransférases nécessaires à la synthèse des chaînes de polylactosamines sont localisées dans les citernes *médial* et *trans* de l'appareil de Golgi. Il est alors admis que l'ajout de ces chaînes sur les glycoprotéines s'effectue dans ces compartiments cellulaires. En revanche, les glycoprotéines transitant à travers l'appareil de Golgi ne sont pas systématiquement toutes modifiées par des polylactosaminoglycanes, indiquant qu'il existe des mécanismes de régulation. Cependant, ces mécanismes restent mal compris.

C'est pourquoi, nous avons cherché à identifier une séquence en acide-aminé pouvant guider l'ajout des chaînes de polylactosamines sur la protéine M. Nous nous sommes focalisés sur les résidus du domaine N-terminal dû à leur proximité avec le site de glycosylation. Notre étude ayant été initiée avant la résolution de la structure 3D de la protéine M, nous avons prédit *in silico* que le domaine N-terminal de la protéine se composait des 20 premiers résidus. Ainsi, nous avons substitué 3 par 3 ces résidus situés en aval du site de glycosylation de MERS-CoV-M et SARS-CoV-2-M, en alanine (**Figure 21A**). Des cellules Huh-7 ont ensuite été transfectées pendant 24h afin d'exprimer les différents mutants. Leur statut de glycosylation a ensuite été analysé par western-blot avec un anticorps anti-V5 (**Figure 21B**). Nous pouvons remarquer que la majorité des mutants construits sont correctement glycosylés et portent des chaînes de polylactosamines, comme l'indique la présence d'une bande diffuse. Par conséquent, aucunes de nos mutations introduites n'a eu un effet spécifique sur l'ajout des chaînes de polylactosamines. En revanche, certaines mutations impactent la glycosylation de M, comme la substitution en alanine du motif ⁹EAQ₁₁ de MERS-CoV-M. En effet, nous observons uniquement la présence d'une espèce non-glycosylée de ce mutant, puisque son poids moléculaire est identique à celui de N3Q-M. Nous avons donc poursuivi nos investigations à la recherche de séquences ou de résidus nécessaires pour la glycosylation de M.

Ainsi, nous avons quantifié l'intensité de la bande correspondant à la forme non-glycosylée de chacun des mutants, et l'avons normalisée par un traitement des échantillons à la PNGase F reflétant la quantité totale de protéine M. La quantité de M non-glycosylée a pu ensuite être exprimée en pourcentage de M total (**Figure 21C**). Selon les quantifications réalisées, comme mentionné précédemment, il s'avère que la forme non-glycosylée de M est minoritaire dans les cellules, puisque nous estimons qu'elle représente 3,32 % +/- 1,13 % de la quantité totale de MERS-CoV-M et 5,73 % +/- 2,18 % de la quantité totale de SARS-CoV-2-M. En revanche, ces proportions sont portées à 91,00

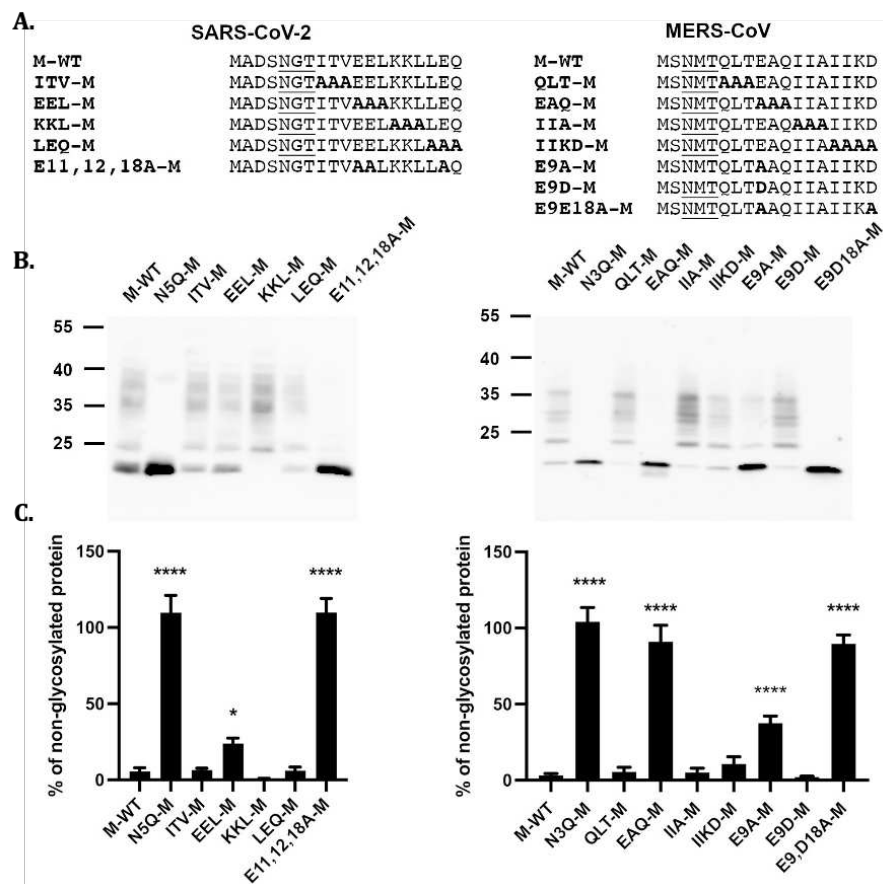


Figure 21. La mutation des résidus acides de MERS-CoV-M et SARS-CoV-2-M perturbe la glycosylation.

(A). Séquences en acide-aminé des différents mutants construits. (B). Analyse du profil de migration des différents mutants ayant été exprimés pendant 24 h en cellules Huh-7. Un anticorps anti-V5 a été utilisé pour la révélation des protéines. (C). Quantification de la proportion de protéine M non-glycosylée pour chacun des mutants. Des cellules Huh-7 ont été transfectées afin d'exprimer les constructions mentionnées, puis les lysats cellulaires récoltés ont été traités par la PNGase F ou non-traités avant d'être analysés par western-blot. L'intensité de la bande correspondant à la forme non-glycosylée de M a été quantifiée et normalisée avec celle des échantillons traités par la PNGase F reflétant la quantité totale de protéine M. Les résultats présentés correspondent à la quantité de protéine non-glycosylée et correspondent à la moyenne de quatre expériences indépendantes. Les barres d'erreur correspondent à l'erreur type sur la moyenne (SEM). Les résultats ont été analysés à l'aide d'un test ANOVA (* P<0,1 ; **** P<0,0001).

% +/- 10,82 % pour le mutant EAQ-M de MERS-CoV-M et 23,95 % +/- 3,43 % pour le mutant EEL-M de SARS-CoV-2-M.

Ces deux constructions ont la particularité d'avoir au moins un résidu d'acide glutamique substitué en alanine. Afin d'étudier si les acides glutamiques du domaine N-terminal de M jouent un rôle dans le processus de glycosylation, nous avons muté ponctuellement le résidu E9 du motif EAQ en alanine (E9A-M) dans la protéine M du MERS-CoV. En western-blot, la bande associée à l'espèce non-glycosylée présente une plus forte intensité que celle de la protéine sauvage. Cependant, des formes glycoconjuguées sont visibles, portant ainsi la quantité de E9A-M non-glycosylée à 37,51 % (+/- 4,73 %) de la quantité totale de M. Ces formes glycoconjuguées n'ayant pas été observées avec le mutant EAQ-M, nous avons muté ponctuellement le résidu glutamine Q11 en alanine. Cependant, cette mutation n'a eu aucun impact sur le niveau de glycosylation de la protéine, ce qui ne nous permet pas d'expliquer pourquoi la mutation du motif EAQ en alanine impacte davantage la glycosylation de M que la mutation E9A (données non-montrées).

L'acide glutamique est classé parmi les acide-aminés acides, puisqu'il porte une charge négative au niveau de sa fonction -COOH présente dans sa chaîne latérale. L'acide aspartique est le deuxième acide-aminé acide. Afin de caractériser davantage le rôle des acides-aminés acides pour la glycosylation de M, nous avons substitué le résidu E9 en acide aspartique (E9D-M). Le profil de migration de ce mutant est comparable à celui de la protéine sauvage, indiquant que la glycosylation de M n'a pas été impactée. Par conséquent, le mutant E9D-M permet de montrer que la glycosylation de M dépend de la présence d'une charge négative portée par un résidu acide dans le domaine N-terminal de la protéine.

Suite à ces résultats, nous avons souhaité substituer l'ensemble des résidus acides du domaine N-terminal de MERS-CoV-M et SARS-CoV-2-M afin de confirmer leur implication pour la glycosylation. Ainsi, nous avons généré les mutants E11,12,18A-M pour SARS-CoV-2-M et E9D18A-M pour MERS-CoV-M. En western-blot, seulement une espèce non-glycosylée est détectée pour ces deux mutants. De plus, les proportions de E11,12,18A-M et E9D18-M non-glycosylées sont similaires à celles des mutants N5Q-M et N3Q-M. En conclusion, ces résultats permettent d'affirmer que les résidus acides situés

proche des sites accepteurs N3 (MERS-CoV-M) et N5 (SARS-CoV-2-M) sont nécessaires pour la glycosylation de M.

Comme décrit dans le chapitre 1.7.1, la N-glycosylation peut jouer un rôle dans le contrôle qualité du repliement des glycoprotéines au niveau du RE. Seules les protéines correctement repliées pourront être exportées vers l'appareil de Golgi. Dans le cas contraire, les protéines seront retenues au niveau du RE, avant d'être exportées vers le cytoplasme pour être dégradées par le protéasome. Par conséquent, nous avons voulu vérifier que les mutations introduites ne sont pas à l'origine d'un mauvais repliement de la protéine qui pourrait expliquer l'absence de glycosylation. Pour cela, nous avons vérifié que les mutants ne s'accumulaient pas dans le RE. En effet, dans une étude précédemment menée au laboratoire, il a été montré que la protéine M du MERS-CoV est localisée dans le réseau *trans*-golgien (TGN) lorsqu'elle est exprimée seule dans la cellule (Perrier et al. 2019). Ainsi, 24h post-transfection en cellules Huh-7, nous avons réalisé des expériences de co-marquages des mutants V5-E9D18A-M et V5-E11,12,18A-M avec la protéine TGN46, marqueur du compartiment TGN (**Figure 22**). Nous pouvons observer que E9D18A-M et E11,12,18A-M co-localisent avec le marqueur TGN46, ce qui indique que les mutations introduites n'ont pas eu d'impact sur le repliement protéique.

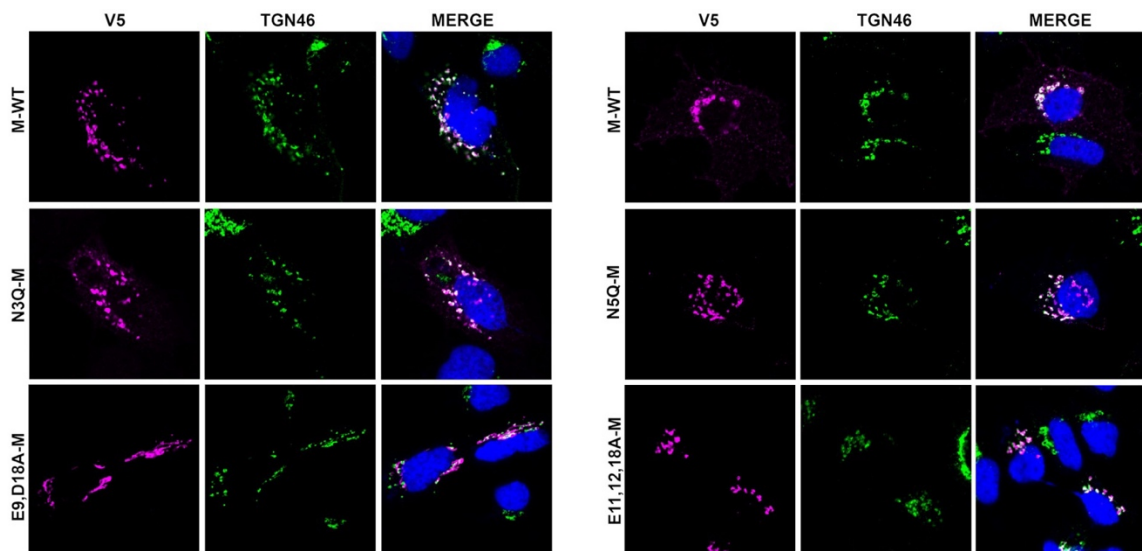


Figure 22. Les mutants sont localisés dans le compartiment TGN.

Des cellules Huh-7 ont été transfectées pendant 24 h par un vecteur contenant la séquence codant les protéines taguées V5 mentionnées. Ces cellules ont ensuite été fixées et perméabilisées. Des co-marquages ont été réalisés avec un anticorps anti-V5 (magenta) et anti-TGN46 (vert). Les noyaux ont été marqués au DAPI (bleu).

4.4 Les mutants E9D18A-M et E11,12,18A-M conservent la topologie N-exo/C-endo

La protéine M des coronavirus est une protéine qui possède 3 segments transmembranaires, et dont le domaine N-terminal est exposé vers le lumen du RE ou à l'extérieur du virus, et le domaine C-terminal vers le cytoplasme ou à l'intérieur du virus (Armstrong et al. 1984; Voß et al. 2009). Elle présente ainsi une topologie de type N-exo/C-endo. Cette topologie est obligatoire pour que le domaine N-terminal de la protéine soit N-glycosylé. Par conséquent, nous avons souhaité vérifier la topologie membranaire de nos mutants par perméabilisation sélective des membranes cellulaires.

Pour cela, des cellules Huh-7 ont été transfectées pendant 24h afin d'exprimer les protéines M sauvages et mutées taguées V5 en N-terminal, puis un double marquage des cellules avec un anticorps anti-V5 et un anticorps dirigé contre le domaine C-terminal de la protéine a été réalisé dans différentes conditions de perméabilisation. Les cellules ont été perméabilisées soit avec 0,1 % Triton X-100 (Triton(+)) ce qui permet de perméabiliser les membranes de tous les organelles intracellulaires, soit partiellement (Triton(-)) par un cycle de congélation à -80°C suivi d'une décongélation à température ambiante (**Figure 23**). Dans la condition de perméabilisation partielle par congélation/décongélation, seule la membrane plasmique est perméabilisée (Mardones et González 2003). Ainsi, nous pouvons vérifier l'accessibilité de l'étiquette V5 ajoutée côté N-terminal par un anticorps anti-V5 selon la méthode de perméabilisation utilisée. Le côté C-terminal de la protéine est marqué par un anticorps anti-M spécifique à MERS-CoV-M et SARS-CoV-2-M et est accessible quelle que soit la méthode de perméabilisation. Pour les protéines sauvages, l'absence de marquage avec l'anticorps anti-V5 en condition Triton(-) confirme l'orientation du domaine N-terminal vers le lumen du RE. Comme attendu, le domaine C-terminal est marqué quelle que soit la méthode de perméabilisation des membranes cellulaires, confirmant son orientation cytoplasmique. Un marquage similaire aux protéines sauvages est observé pour les deux protéines mutées. Les mutants E9D18A-M et E11,12,18A-M semblent donc avoir une orientation de type N-exo/C-endo, comme décrit pour la protéine M sauvage. Ainsi, une inversion de la topologie membranaire ne semble pas être à l'origine du défaut de glycosylation des mutants E9D18A-M et E11,12,18A-M.

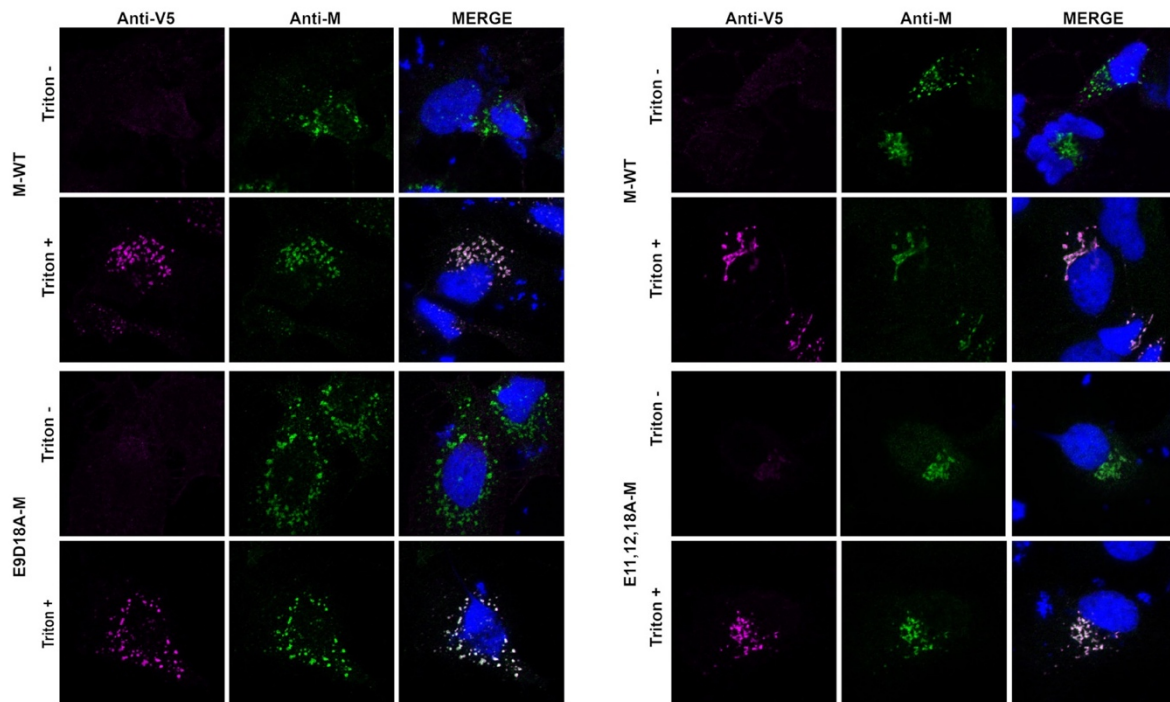


Figure 23. L'extrémité N-terminale des mutants est orientée côté luminal des compartiments intracellulaires.

Des cellules Huh-7 ont été transfectées pendant 24 h afin d'exprimer les différentes constructions indiquées. Elles ont ensuite été fixées et les membranes cellulaires ont été perméabilisées totalement (Triton +) par 0,1 % de Triton X-100 ou partiellement (Triton -) par un cycle de congélation/décongélation à -80 °C. L'extrémité N-terminale a été marquée par un anticorps anti-V5 (magenta) et l'extrémité C-terminale par un anticorps anti-M spécifique (vert). Les noyaux ont été marqués au DAPI (bleu).

Afin de confirmer la topologie N-exo/C-endo de ces mutants, nous avons eu recours à une seconde stratégie expérimentale, qui a consisté en l'ajout dans la boucle reliant le deuxième (TM2) et le troisième (TM3) segment transmembranaire de MERS-CoV-M et du mutant E9D18A-M, d'une séquence linker constituée de 31 acides-aminés hydrophiles (YAYNRYRIGNYKL**NTTI**KDLDYKDDDDKPKKE) et contenant un site de N-glycosylation en son centre (₈₈NTT₉₀) (**Figure 24A**). Cette boucle a été choisie car elle est orientée du côté luminal du RE. Cette séquence linker a été précédemment utilisée pour vérifier la topologie membranaire de la protéine M du SARS-CoV (Voß et al. 2009). Des contrôles ont également été construits où les sites N3 et N88 ont été mutés en glutamines (N3Q-M-linker ; N88Q-M-linker ; N3Q,N88Q-M-linker). Les constructions ont été exprimées en cellules Huh-7 pendant 24h avant de récolter les lysats cellulaires. Les protéines ont ensuite été séparées sur SDS-PAGE 12 % et analysées par western-blot avec un anticorps anti-V5 (**Figure 24B**).

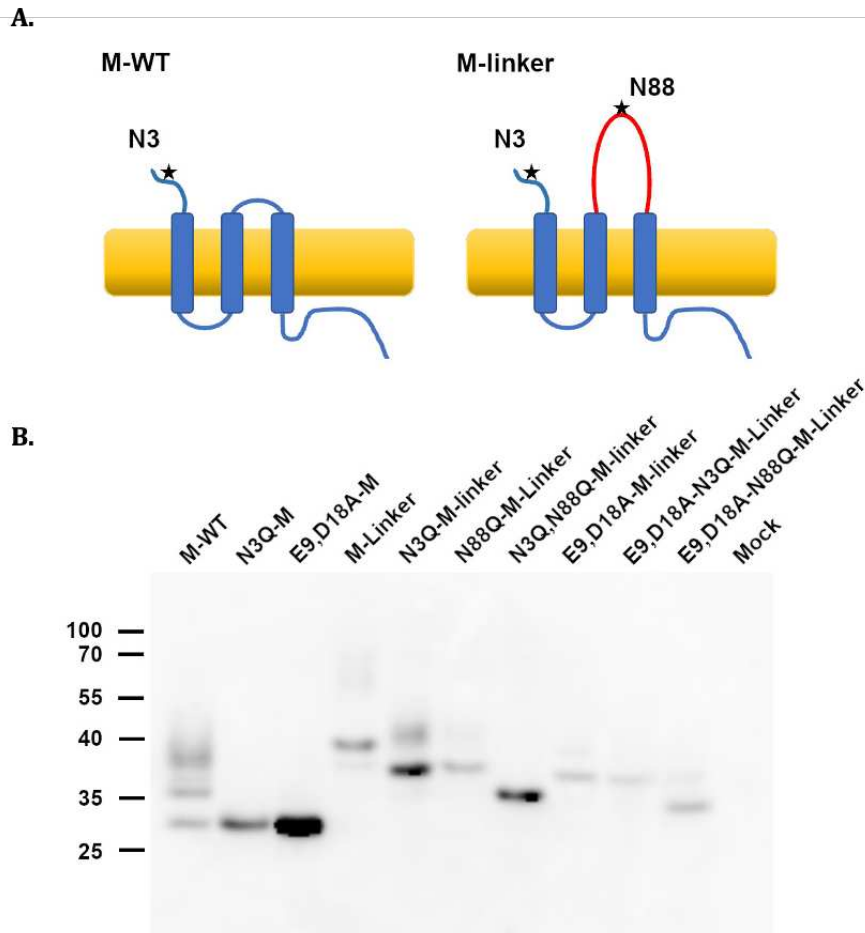


Figure 24. La N-glycosylation d'une séquence linker insérée entre le second et le troisième segment transmembranaire de MERS-CoV-M n'est pas dépendante des résidus acides du domaine N-terminal.

(A). Représentation schématique des constructions. La séquence linker est indiquée en rouge. (B). Des cellules Huh-7 ont exprimé pendant 24 h les différentes protéines mentionnées. Les lysats cellulaires ont été récoltés et les protéines ont été analysées par western-blot à l'aide d'un anticorps anti-V5. Les poids moléculaires sont exprimés en kDa. Mock : vecteur pcDNA3.1 vide.

Dans un premier temps, nous pouvons confirmer la fonctionnalité du système puisque les constructions N3Q-M-linker et N88Q-M-linker sont glycosylées et portent des chaînes de polylactosamines. De plus, le décalage de poids moléculaire avec la construction N3Q,N88Q-M-linker confirme la glycosylation des sites N3 et N88. Également, M-linker migre plus lentement que N3Q-M-linker et N88Q-M-linker, ce qui indique que les sites N3 et N88 portent des N-glycanes. A propos de la mutation E9D18A, nous pouvons observer que la construction E9D18A-M-linker est glycosylée puisqu'elle migre au même niveau que N3Q-M-linker et N88Q-M-linker. De plus, la mutation E9D18A impacte uniquement la glycosylation du site N3 puisque E9D18A-M-linker présente le

même profil de migration que E9D18A-N3Q-M-linker. En revanche, E9D18A-N88Q-M-linker migre essentiellement de la même façon que N3Q,N88Q-M-linker, confirmant l'absence de glycosylation sur le site N3 en présence de la mutation E9D18A. Par conséquent, l'ensemble de ces analyses montrent que la construction E9D18A-M-linker est correctement glycosylée sur le site N88. La séquence linker insérée entre le TM2 et TM3 est donc bien orientée vers la lumière du RE. Ainsi, nous confirmons une nouvelle fois que la mutation des résidus acides du domaine N-terminal de la protéine M ne provoque pas une inversion de la topologie membranaire. D'autre part, ces résultats suggèrent que les résidus E9 et D18 dans la protéine M du MERS-CoV agissent localement, puisque leur mutation n'affecte pas la glycosylation du site présent dans le linker.

4.5 Rôle du trafic intracellulaire et du premier segment transmembranaire de M pour l'ajout des chaînes de polylactosamines

Nos précédentes analyses nous ont permis d'identifier que les résidus acides du domaine N-terminal de la protéine M étaient nécessaires pour le processus de glycosylation. En revanche, nous ne sommes pas parvenus à identifier spécifiquement une séquence en acide-aminé pouvant guider l'ajout des chaînes de polylactosamines.

Pour poursuivre nos investigations sur l'ajout de ces chaînes, nous avons construit des protéines chimériques. Dans un premier temps nous avons transféré l'extrémité N-terminale de la protéine d'enveloppe E1 du virus de l'hépatite C à la place de celle de SARS-CoV-2-M. En effet, il a été montré que la protéine HCV-E1 est fortement glycosylée sur 4 sites accepteurs, dont un se situe dans l'extrémité N-terminale (E1N1) (Meunier et al. 1999). En revanche, aucun N-glycane complexe n'a été détecté sur ces sites, puisque la protéine E1 est retenue dans le RE (Dubuisson et al. 1994). Ici, l'objectif est d'analyser si des chaînes de polylactosamines peuvent être ajoutées sur le premier site accepteur de HCV-E1 dans le contexte de la chimère E1-M.

Pour cela, les 9 premiers résidus de SARS-CoV-2-M (MADSNGTIT) ont été remplacés par les 11 résidus (MAQVKNSSGLY) du domaine N-terminal de HCV-E1. Cette construction a été réalisée suite à la publication de la structure cryo-EM de SARS-CoV-2-

M de Zhang *et al.*, montrant que les 9 premiers résidus constituent le domaine N-terminal de la protéine (Zhang et al. 2022).

Des cellules Huh-7 ont été transfectées par un vecteur pcDNA3.1 vide (mock) ou exprimant les protéines V5-M-WT, V5-N5Q-M, V5-E1-M ou HCV-E1. 24h post-transfection, les lysats cellulaires ont été récoltés et déposés sur SDS-PAGE 12 %. Les protéines ont ensuite été analysées par western-blot avec un anticorps anti-V5 ou A6 ciblant un épitope dans la région 230-239 de HCV-E1 (Mesalam et al. 2018). Alors que la protéine E1 migre sous la forme d'une espèce unique vers 30 kDa, la protéine chimère E1-M présente un profil de migration identique à celui de la protéine M-WT (**Figure 25A**). Ainsi, la détection d'une bande diffuse indique la présence de chaînes de polylectosamines sur le site accepteur contenu dans le domaine N-terminal de E1-M. Ces résultats confirment qu'aucun signal n'est présent dans le domaine N-terminal pour l'ajout de chaînes de polylectosamines.

La distribution intracellulaire des protéines a ensuite été analysée par immunofluorescence en cellules Huh-7. Alors que la protéine HCV-E1 est principalement retenue dans le RE, comme le montre la co-localisation avec la calnexine (CNX), la chimère E1-M co-localise, quant à elle, avec le marqueur TGN46 (**Figure 25B**). Par conséquent, ces résultats suggèrent que le trafic de la protéine jusqu'à l'appareil de Golgi, plutôt qu'un signal, est nécessaire pour l'ajout des chaînes de polylectosamines.

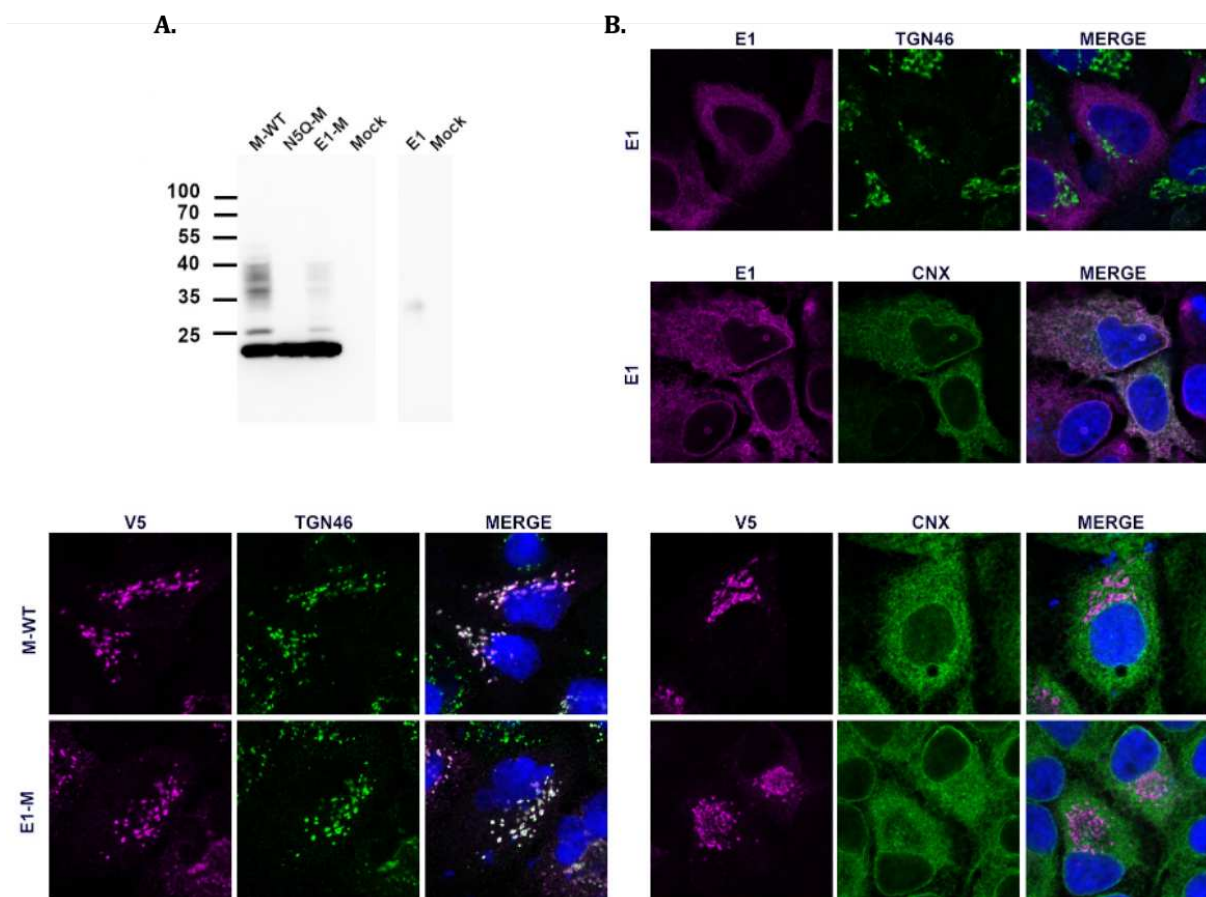


Figure 25. Le premier site de glycosylation de la protéine HCV-E1 est modifié par des chaînes de polylectosamines lorsqu'il est transféré sur SARS-CoV-2-M.

(A). Des cellules Huh-7 ont exprimé pendant 24 h les différentes constructions mentionnées. Les lysats cellulaires ont été collectés et les protéines ont été analysées par western-blot à l'aide d'un anticorps anti-V5 ou anti-E1. Les poids moléculaires sont exprimés en kDa. (B). Des cellules Huh-7 exprimant M-WT, E1 ou E1-M ont été fixées et perméabilisées. Des co-marquages ont été réalisés avec un anticorps anti-V5 ou anti-E1 et anti-TGN46 ou anti-calnexine (CNX). Les noyaux ont été marqué au DAPI (bleu).

Pendant la réalisation de ce projet, la structure tri-dimensionnelle de SARS-CoV-2-M a été résolue par cryo-EM (Dolan et al. 2022 ; Zhang et al. 2022). Il s'avère que la protéine M du SARS-CoV-2 partage une forte homologie structurale avec la protéine accessoire ORF3a, suggérant une possible origine commune (Kern et al. 2021). Selon les données structurales, le domaine N-terminal d'ORF3a est constitué de 42 résidus. Concernant la protéine M, son domaine N-terminal n'est pas structuré mais sa longueur diffère entre les deux structures qui ont été résolues. Ainsi, dans celle de Zhang *et al.*, le domaine N-terminal de la protéine M est constitué de seulement 9 résidus, contre 16 résidus dans celle de Dolan *et al.*, Par conséquent, selon la structure, les résidus acides identifiés précédemment et importants pour la glycosylation de la protéine M sont

localisés ou non dans l'hélice- α formant le premier segment transmembranaire. Ces données structurales montrent également que le premier segment transmembranaire d'ORF3a est plus court que celui de M, et ne contient aucun résidu acide. Par conséquent, nous avons construit de nouvelles protéines chimères en remplaçant, soit uniquement le domaine N-terminal d'ORF3a par les 9 premiers résidus de la protéine M (Nt-M/ORF3a), soit le domaine N-terminal et le premier segment transmembranaire de ORF3a par ceux de M (TM1-M/ORF3a). Les protéines chimères construites sont représentées en **Figure 26A**. En western-blot, une seule bande est visible après migration de la protéine ORF3a et de la chimère Nt-M/ORF3a, suggérant que la protéine n'est pas glycosylée (**Figure 26B**). Par ailleurs, il est connu que la protéine ORF3a ne possède aucun site consensus pour la N-glycosylation. En revanche, il a été montré que la protéine ORF3a du SARS-CoV était modifiée par des O-glycanes (Oostru et al. 2006). Concernant la chimère TM1-M/ORF3a, différentes espèces sont observées en western-blot, y compris une bande diffuse, suggérant la présence de chaînes de polylectosamines.

La distribution intracellulaire des protéines sauvages et chimériques a ensuite été analysée par immunofluorescence. Des cellules Huh-7 ont exprimé pendant 24h les protéines mentionnées avant d'être fixées et perméabilisées. Les protéines d'intérêt ont ensuite été immuno-marquées avec un anticorps anti-V5. Le compartiment TGN a été marqué avec un anticorps anti-TGN46 (**Figure 26C**). Des analyses précédentes ont mis en évidence que la protéine ORF3a est essentiellement localisée dans la voie endosomale et à la surface cellulaire (Miller et al. 2023). Ces données sont cohérentes avec l'absence de co-localisation de la protéine ORF3a avec le marqueur TGN46. En revanche, le marquage ponctué visible avec ORF3a est caractéristique d'un marquage de la voie endosomale. Concernant les chimères Nt-M/ORF3a et TM1-M/ORF3a, nous avons observé un marquage fluorescent identique à celui d'ORF3a. Par conséquent, les chimères Nt-M/ORF3a et TM1-M/ORF3a se distribuent de manière similaire à ORF3a.

En conclusion, l'ensemble de ces résultats suggère que le trafic de la protéine jusqu'à l'appareil de Golgi est nécessaire pour l'addition des chaînes de polylactosamines, comme observé avec E1-M. Cependant, les chimères M-ORF3a montrent que la rétention dans le TGN n'est pas obligatoire pour l'ajout de ces chaînes. En effet, la chimère TM1-M/ORF3a est modifiée par des polylactosaminoglycane, malgré une probable distribution de la protéine dans la voie endosomale. De plus, le premier segment transmembranaire de M, contenant les résidus acides, est indispensable pour la glycosylation du domaine N-terminal de la protéine.

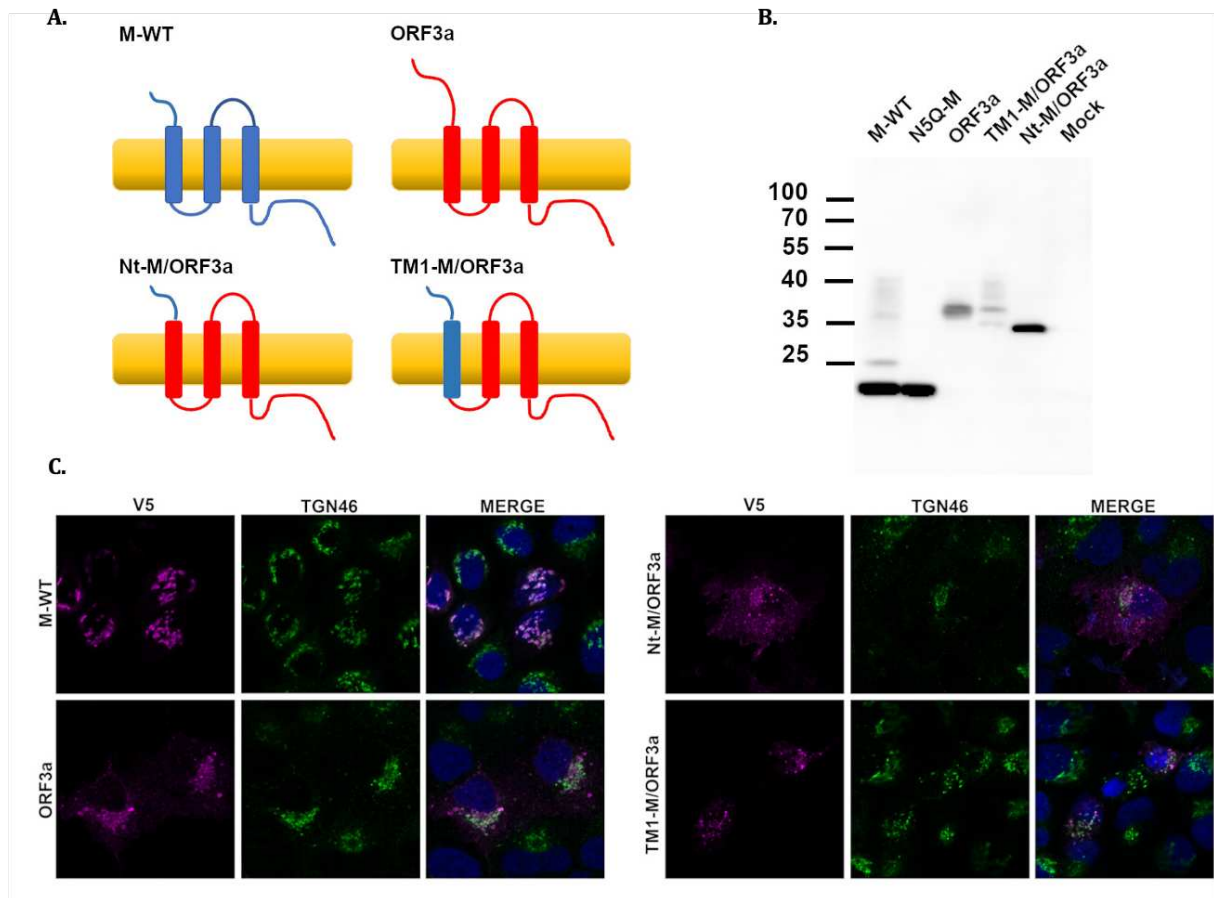


Figure 26. Le premier segment transmembranaire de SARS-CoV-2-M est requis pour la glycosylation de la chimère M/ORF3a.

(A). Représentation schématique des protéines chimères construites. Les domaines de la protéine M sont représentés en bleu et ceux de ORF3a en rouge. (B). Des cellules Huh-7 ont exprimé pendant 24 h les protéines mentionnées avant de collecter les lysats cellulaires. Les protéines ont été analysés à l'aide d'un anticorps anti-V5. Les poids moléculaires sont exprimés en kDa. (C). Après 24 h d'expression des protéines indiquées, les cellules Huh-7 ont été fixées et perméabilisées. Des co-marquages ont été réalisés avec un anticorps anti-V5 (magenta) et anti-TGN46 (vert). Les noyaux ont été marqués au DAPI (bleu).

4.6 La glycosylation de la protéine M est dépendante du complexe OST-B

La N-glycosylation des protéines est initiée dans le lumen du RE par l'action du complexe oligosaccharyltransférase (OST). Deux isoformes du complexe OST ont été identifiées, nommées OST-A et OST-B. Le complexe OST-A est associé au translocon Sec61 et catalyse le transfert du précurseur $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ sur les protéines de manière co-translationnelle (Shrimal et al. 2015, 2017). En revanche, OST-B n'est pas associé au translocon et est plutôt impliqué dans le transfert post-translationnel du précurseur oligosaccharidique (Shrimal et al. 2015). Cependant, il a été montré qu'OST-B peut remplacer l'activité OST-A lorsque ce dernier n'est pas fonctionnel (Ruiz-Canada et al. 2009). La principale différence entre ces deux isoformes est la sous unité catalytique STT3. Alors qu'OST-A contient la sous-unité STT3A, OST-B contient la sous-unité STT3B.

Afin d'identifier la machinerie responsable de la glycosylation de la protéine M du MERS-CoV et du SARS-CoV-2, nous avons généré des lignées cellulaires Huh-7 déficientes en STT3A et/ou STT3B par CRISPR/Cas9. Après sélection, les populations obtenues ont directement été transfectées pendant 24h afin d'exprimer les protéines M et S. Le statut de glycosylation a ensuite été étudié par western-blot. La protéine spike a été utilisée comme contrôle, puisqu'elle est largement glycosylée. De plus, il a été montré récemment que la glycosylation de spike était principalement dépendante de STT3A (Huang et al. 2021). Le niveau d'expression des sous-unités STT3A et STT3B a également été analysée par western-blot (**Figure 27A**). Les expériences ayant été menées sur des populations cellulaires, il n'est pas surprenant d'observer une expression résiduelle de STT3A et STT3B. Concernant la protéine M, nous avons remarqué que les bandes associées aux formes glycoconjuguées sont de plus faibles intensités dans les populations cellulaires *STT3B-KO* et *STT3A+B-KO* (**Figure 27B**). En revanche, leur niveau d'intensité est inchangé entre les populations pX330 et *STT3A-KO*. De plus, la bande correspondant à la forme non-glycosylée de MERS-CoV-M est plus intense dans les populations *STT3B-KO* et *STT3A+B-KO*, confirmant une déficience en glycosylation de la protéine. Concernant MERS-CoV-S et SARS-CoV-2-S, une légère bande diffuse est visible dans la population *STT3A+B-KO*, indiquant des modifications dans la glycosylation de la protéine (**Figure 27C**). De plus la quantité de protéine est fortement diminuée probablement à cause d'une dégradation de

la protéine suite à un défaut de glycosylation. Ainsi, l'ensemble de ces résultats permet de mettre en évidence que la glycosylation de la protéine M du MERS-CoV et du SARS-CoV-2 est dépendante du complexe OST-B.

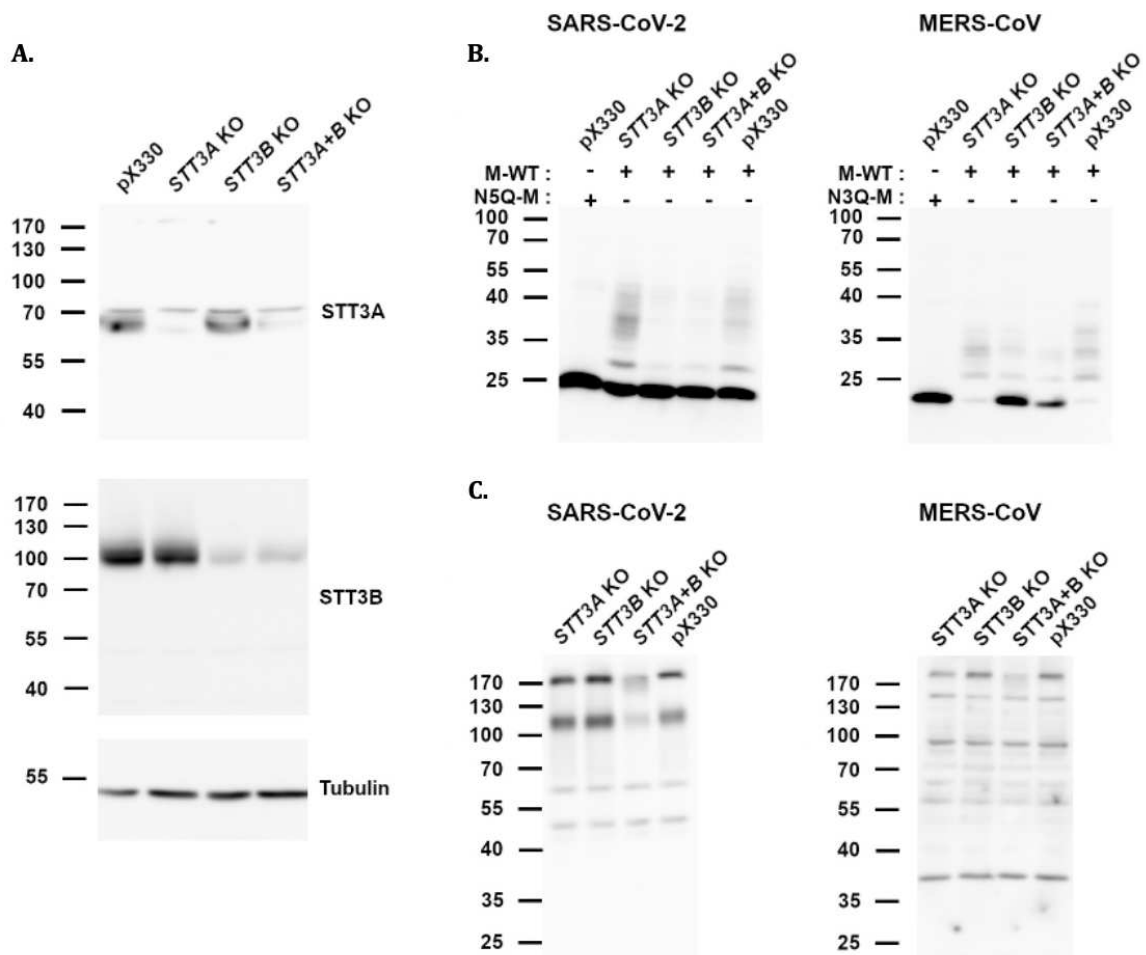


Figure 27. La glycosylation de la protéine M du MERS-CoV et du SARS-CoV-2 est dépendante de la machinerie OST-B.

(A). Niveau d'expression des protéines STT3A et STT3B dans les populations cellulaires manipulées génétiquement. (B). Des cellules Huh-7 contrôles (pX330) et déficientes (KO) en STT3A et/ou STT3B ont exprimé pendant 24 h les protéines M indiquées. Les lysats cellulaires ont été collectés et les protéines ont été révélées par un anticorps anti-V5. (C). Les cellules contrôles pX330 et déficientes pour STT3A et/ou STT3B ont exprimé pendant 24 h la protéine S du MERS-CoV ou du SARS-CoV-2. Les protéines ont été révélées par un anticorps polyclonal anti-S spécifique au MERS-CoV ou au SARS-CoV-2.

4.7 Analyse de la protéine M en cellules infectées et incorporée aux particules virales

Nos précédents résultats ont permis d'identifier et de valider que des chaînes de polylectosamines étaient associées à la protéine M du MERS-CoV et du SARS-CoV-2, en cellules transfectées transitoirement. De plus, la localisation intracellulaire de la protéine M au niveau du TGN est cohérente avec l'ajout de ces chaînes. En revanche, la protéine M est distribuée dans l'ensemble de la voie de sécrétion en cellule infectée. Par conséquent, nous avons testé si des polylectosaminoglycanes étaient également ajoutés sur la protéine M en cellule infectée.

Ainsi, nous avons tiré profit de notre lignée cellulaire Huh-7 $\beta 3GNT2$ -KO utilisée précédemment (**Figure 28A-B**). Cette lignée a été infectée par le MERS-CoV ou le SARS-CoV-2 à une M.O.I de 0,1 pendant 16h, avant de collecter les lysats cellulaires. Puisque les cellules Huh-7 n'expriment pas de façon ubiquitaire le récepteur ACE2 nécessaire pour l'infection par le SARS-CoV-2, elles ont été transduites par un lentivirus avant l'infection virale. La protéine M a ensuite été analysée par western-blot avec un anticorps anti-M spécifique. La bande diffuse pouvant être difficilement visible en western-blot, deux temps d'exposition sont présentés en Figure 28A. Dans les cellules infectées pX330, nous remarquons que le profil de migration de la protéine M ressemble à celui observé en condition de transfection transitoire. Cependant, il est nécessaire d'augmenter le temps d'exposition pour pouvoir observer la bande diffuse, suggérant que les formes associées à cette bande sont faibles dans ce contexte.

Dans les cellules Huh-7 $\beta 3GNT2$ -KO, cette bande diffuse n'est pas détectée, même après une longue exposition. A la place, une bande migrant plus lentement que les autres espèces est détectée. Ces résultats étant identiques à ceux observés précédemment en Figure 1E, nous pouvons conclure que des chaînes de polylectosamines sont également associées à la protéine M du MERS-CoV et du SARS-CoV-2 en cellule infectée.

L'assemblage et le bourgeonnement des coronavirus s'effectuent dans le compartiment ERGIC. Ce compartiment étant situé en-amont du Golgi, il est probable que la protéine M incorporée dans les virions ne soit pas modifiée par des chaînes de polylectosamines. Après cette étape, les virions transitent à travers l'appareil de Golgi, où

la protéine M peut subir davantage de modifications. Ainsi, pour étudier les formes de M associées aux particules virales, nous avons eu recours au système de VLP. Pour produire ces particules, des cellules Huh-7 ont été co-transfectées par des vecteurs exprimant la protéine M seule ou en combinaison avec un vecteur exprimant la protéine E. 48h post-transfection, les VLP sécrétées dans le milieu de culture ont été concentrées par ultracentrifugation sur un coussin de sucrose. Les lysats cellulaires ont également été collectés. Les échantillons ont ensuite été analysés par western-blot. Comme attendu, nous pouvons observer que les trois formes de M sont présentes dans les lysats cellulaires (Figure 9B). En revanche, nous observons un enrichissement de la forme conjuguée à des N-glycanes riches en mannose dans les VLP (**Figure 28C**). De plus, la co-expression des formes mutées MERS-CoV-N3Q-M ou SARS-CoV-2-N5Q-M avec la protéine E, permet la formation de VLP, confirmant à nouveau que la glycosylation de M n'est pas nécessaire pour l'assemblage.

Pour conclure, l'ensemble de ces résultats ont permis de confirmer que seule la forme conjuguée à des N-glycanes riches en mannose est incorporée dans les VLP. Par conséquent, la protéine M ne semble subir aucune modification additionnelle dans l'appareil de Golgi après avoir été incorporée dans les particules virales. De plus, l'incorporation de la forme associée à des N-glycanes riches en mannose est cohérente avec un assemblage et un bourgeonnement des coronavirus se déroulant au niveau du compartiment ERGIC.

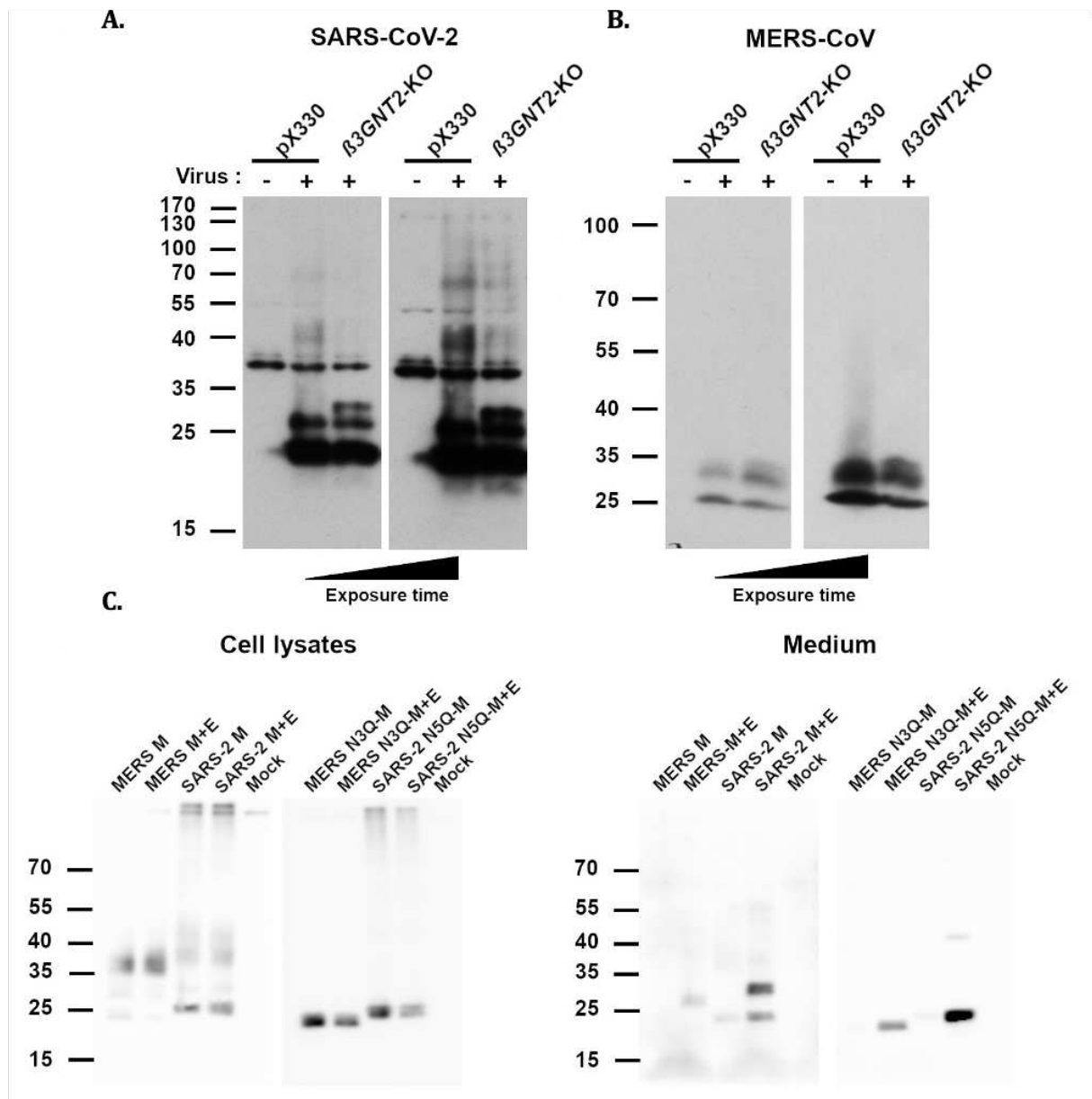


Figure 28. Les protéines M du MERS-CoV et du SARS-CoV-2 sont modifiées par des chaînes de polylectosamines en cellules infectées mais cette forme est absente des particules sub-virales.

(A). Des cellules Huh-7 déficientes (KO) en β 3GnT2 et contrôle (pX330) ont été infectées par le SARS-CoV-2 (panel de gauche) ou le MERS-CoV (panel de droite) à une M.O.I de 0,1 pendant 24 h. Les lysats cellulaires ont été collectés et analysés par western-blot. Les protéines ont été révélées à l'aide d'anticorps anti-M spécifiques. Deux temps d'exposition sont présentés. (B). Des cellules Huh-7 ont été co-transfectées pendant 24 h par des vecteurs d'expression contenant les séquences codants les protéines M sauvages ou mutées dans leur site de glycosylation, avec un vecteur d'expression vide ou contenant la séquence de la protéine E du MERS-CoV ou du SARS-CoV-2. Les lysats cellulaires ont été collectés et les protéines analysées par western-blot avec un anticorps anti-V5. (C). Les surnageants des cellules transfectées ont été collectés et ultracentrifugés sur un coussin de sucrose 20 % pour concentrer les VLP sécrétées dans le milieu. Les VLP ont ensuite été resuspendues et analysées par western-blot. La protéine M a été détectée avec un anticorps anti-V5.

5 Discussion et perspectives

Au moment de la synthèse protéique, des modifications post-traductionnelles peuvent être ajoutées sur les polypeptides naissant afin de rendre ces-derniers fonctionnels. Parmi-elles, la glycosylation est la modification post-traductionnelle majeure des protéines, et est commune à tous les êtres vivants. Elle consiste en la liaison covalente de glucides sur des protéines membranaires ou sécrétées, grâce à l'action de glycosyltransférases résidant dans le réticulum endoplasmique et l'appareil de Golgi. Lors d'infection virale, les virus peuvent détourner la machinerie de glycosylation cellulaire afin de promouvoir la glycosylation des protéines virales. Chez les coronavirus, la protéine spike est la protéine la plus glycosylée.

Ce projet de thèse s'est intéressé particulièrement à la protéine de membrane (M) des coronavirus humains hautement pathogènes MERS-CoV et SARS-CoV-2, où des sites potentiels de N-glycosylation ont été identifiés aux positions N3 pour MERS-CoV-M et N5 pour SARS-CoV-2-M. La glycosylation de la protéine M dans son domaine N-terminal est un processus conservé au cours de l'évolution, cependant, le rôle de ces modifications dans le cycle infectieux reste encore à définir. L'objectif du projet était de caractériser les mécanismes impliqués dans la glycosylation de MERS-CoV-M et SARS-CoV-2-M et de déterminer la nature des glycanes arborant ces protéines.

Avant le début de ce projet, il avait déjà été montré que le site N3 de MERS-CoV-M était fonctionnel, puisque des N-glycanes riches en mannose et complexes avaient été identifiés (Perrier et al. 2019). Nos travaux ont permis de mettre en évidence que le site N5 de SARS-CoV-2-M est également fonctionnel, puisque cette protéine présente un profil de migration similaire à celui de MERS-CoV-M en western-blot. De plus, la mutation du site N5 en glutamine (N5Q) a eu un effet sur le profil de migration, où plus qu'une seule espèce était présente. Nous avons également démontré que N5 est modifié à la fois par des N-glycanes riches en mannose (sensibles à l'Endo-H) et complexes.

Il est intéressant d'observer qu'une espèce non-glycosylée de M est également visible en western-blot, migrant à la même taille que les mutants respectifs MERS-CoV-N3Q-M et SARS-CoV-2-N5Q-M. De plus, l'intensité de la bande correspondant à M non-

glycosylée semble indiquer que cette espèce est majoritaire par rapport aux formes de M conjuguées à des N-glycanes. Nous avons émis l'hypothèse que cette espèce et l'espèce sensible à l'Endo-H étaient des intermédiaires de glycosylation. Ainsi, nous avons suivi la glycosylation de M au cours du temps grâce à un traitement par la cycloheximide, un inhibiteur de la machinerie de traduction. Les données recueillies ont montré que la forme sensible à l'Endo-H avait totalement disparu après 2 h de traitement, signifiant que cette espèce est un précurseur pour la synthèse des formes conjuguées à des glycanes complexes. En revanche, l'espèce non-glycosylée est restée stable dans le temps, suggérant qu'une proportion de M n'est jamais modifiée par des N-glycanes. Par ailleurs, nous avons détecté cette forme dans les particules VLP, ce qui indique qu'elle n'est probablement pas retenue dans le RE.

Nous pensons alors que la glycosylation de M n'était pas efficace. Mais étonnamment, nos expériences de quantification ont indiqué le contraire, puisque M non-glycosylée représente environ 3 % de la quantité totale de MERS-CoV-M et environ 6 % de celle de SARS-CoV-2-M.

Il se pourrait que le temps de résidence de la protéine M dans le RE puisse expliquer pourquoi une proportion de M ne subit pas de modifications. Il a été observé que la mutation en alanine du motif di-acidique DxE dans le domaine C-terminal de MERS-CoV-M, perturbe l'export de la protéine depuis le RE vers l'appareil de Golgi (Perrier et al. 2019). En western-blot, presque aucune forme non-glycosylée n'est visible, alors que la bande intermédiaire sensible à l'Endo-H est d'une plus forte intensité par rapport à la condition sauvage. Par conséquent, il semblerait que l'allongement du temps de résidence de la protéine dans le RE ait amélioré l'efficacité de glycosylation.

Deux machineries ont été décrites pour la glycosylation des protéines eucaryotes, nommées OST-A et OST-B. Ce sont des macro-complexes moléculaires, dont la principale différence est la sous-unité catalytique STT3A (OST-A) et STT3B (OST-B). Nos travaux ont permis d'identifier que la glycosylation des sites N3 et N5 respectifs à MERS-CoV-M et à SARS-CoV-2-M dépendait uniquement du complexe OST-B. Cette machinerie semble transférer un précurseur oligosaccharidique préférentiellement sur des polypeptides repliés et insérés dans les membranes, contrairement à OST-A, qui transfère ce précurseur lors de la synthèse du polypeptide à travers le canal de translocation. Par conséquent, la glycosylation de M semble s'effectuer de manière post-traductionnelle. Ainsi, la glycosylation post-traductionnelle de M couplé à un possible export rapide de la

protéine du RE vers le Golgi pourrait expliquer pourquoi une proportion de M n'est jamais modifiée.

Nous avons également mis en évidence que les N-glycanes complexes arborés par MERS-CoV-M et SARS-CoV-2-M sont probablement des chaînes de polylactosamines, puisque la lectine LEL a reconnu spécifiquement les formes de M glycoconjuguées à des N-glycanes complexes. L'absence de reconnaissance par la lectine DSL pourrait indiquer que les chaînes de polylactosamines associées à M seraient linéaires et porteraient des fucosylations et/ou des sialylations à leurs extrémités, en se basant sur les propriétés d'affinité de DSL (Yamashita et al. 1987 ; Bojar et al. 2022). De plus, le profil de migration de MERS-CoV-M et SARS-CoV-2-M est modifiée lorsque les protéines sont exprimées en cellules Huh-7 déficientes pour la β 1,3-N-acétylglucosaminyltransférase 2 (β 3GnT2), décrite comme la polylactosamine synthase cellulaire principale (Togayachi et al. 2010). Cette glycosyltransférase intervient dans l'élongation des chaînes de polylactosamines en transférant un résidu N-acétylglucosamine sur un résidu galactose par une liaison osidique β 1-3 (**Figure 17**). Ainsi, l'absence de bande diffuse visible en western-blot lorsque M est exprimée dans les cellules Huh-7 modifiées génétiquement, indique que β 3GnT2 participe à la maturation des glycanes complexes de M, confirmant ainsi la présence de chaînes de polylactosamines.

Malgré la complexité technique, il serait intéressant de définir la structure de ces polylactosaminoglycanes par spectrométrie de masse. Une lignée cellulaire déficiente pour la β 1,6-N-acétylglucosaminyltransférase 2 (GCNT2), impliquée dans la ramification des chaînes de polylactosamines, pourrait aussi être générée par manipulation génétique afin d'évaluer son rôle dans la maturation de N-glycanes des protéines M et mieux caractérisés ces derniers.

Nous avons également réalisé des infections par le MERS-CoV et le SARS-CoV-2 de notre lignée β 3GNT2-KO et contrôle pX330. Dans le contexte de la lignée contrôle pX330, la bande diffuse est difficilement détectable en western-blot, suggérant que la quantité de protéine M modifiée par des glycanes complexes serait plus faible qu'en condition de transfection transitoire. Il est fort possible que la présence des autres protéines virales soit à l'origine de cet effet, puisque les interactions protéine-protéines vont réguler le trafic intracellulaire des protéines afin de les retenir principalement au niveau du site

d'assemblage (ERGIC). Dans la lignée déficiente pour la β 3GnT2, la faible bande diffuse n'est plus détectée, malgré des longs temps d'exposition. Nous avons observé une espèce à haut poids moléculaire, comparable à celle détectée en cellules transfectées. Ainsi, β 3GnT2 est également impliquée dans la maturation des glycanes de M en cellules infectées, signifiant que des chaînes de polylactosamines ont été ajoutées.

D'autres protéines virales possèdent des chaînes de polylactosamines, comme la protéine NB du virus *Influenza B* et la petite protéine hydrophobe (SH) du virus respiratoire syncytial (Williams et Lamb 1988; Anderson et al. 1992). Il a été décrit que les formes glycosylées de ces protéines sont incorporées dans les particules virales (Betakova et al. 1996; Huong et al. 2023). En revanche, la fonction de leurs chaînes de polylactosamines reste encore inconnue.

Les mécanismes régulant l'ajout de ces structures sont peu connus. Il avait été suggéré par Williams et Lamb que la proximité du site de glycosylation avec les membranes cellulaires servirait de signal pour l'ajout des polylactosaminoglycanes (Williams et Lamb 1988). Cependant, la protéine HA des virus *Influenza* porte des sites de glycosylation proches de la membrane, mais ne sont pas modifiés par des chaînes de polylactosamines, signifiant que l'ajout de ces chaînes est régulé par des mécanismes plus complexes. De plus, des chaînes de polylactosamines ont été détectées seulement sur quelques sites accepteurs de la protéine S du SARS-CoV-2 et de la protéine membranaire associée aux lysosomes (LAMP) 1 et 2, démontrant davantage la complexité des mécanismes régulant l'ajout de ces chaînes (Carlsson et Fukuda 1990; Sanda et al. 2021).

Nous avons alors cherché à mieux comprendre ces mécanismes de régulation, en essayant d'identifier une séquence qui serait impliquée dans l'addition de chaînes de polylactosamines. Pour cela, nous avons inséré des mutations dans les 20 premiers résidus de la protéine, qui, selon des prédictions bio-informatiques, constituaient le domaine N-terminal de M. La structure de la protéine résolue par cryo-EM n'était pas encore disponible à ce moment. Nous n'avons, cependant, pas identifié de séquence spécifiquement responsable de l'ajout de ces chaînes, mais avons remarqué que les résidus acides proches du site accepteur (E9 et D18 pour MERS-CoV-M ; E11, E12 et E18 pour SARS-CoV-2-M) sont cruciaux pour le processus de glycosylation de M, puisque leur

mutation en alanine a totalement aboli la glycosylation. Curieusement, la mutation en alanine du motif $_{14}\text{KKL}_{16}$ de SARS-CoV-2-M semble avoir stimulé la glycosylation de la protéine, puisqu'aucune forme non-glycosylée n'est visible. De la même façon, pour les constructions de M contenant la séquence linker, nous pouvons observer sur le western-blot, que la bande majoritaire correspond à la forme oligomannosylée de M (**Figure 24B**). Une légère bande diffuse pouvant correspondre à l'ajout de chaînes de polylectosamines peut aussi être visible, selon le niveau d'expression des constructions. En revanche, nous ne détectons pas d'espèce non glycosylée en présence de la séquence linker. Ainsi, cela semble indiquer que l'ajout de ce linker a eu un impact positif sur l'efficacité de glycosylation de M. Il serait intéressant de mieux comprendre comment cette mutation/insertion peut améliorer l'efficacité de N-glycosylation. Il n'est pas impossible que ces protéines soient glycosylées de façon co-translationnelle par OST-A, ce qui résulterait en une glycosylation totale de M. L'expression de ces constructions dans nos lignées cellulaires Huh-7 *STT3A-KO* et/ou *STT3B-KO* permettrait de vérifier cette hypothèse.

Il est aussi possible que son temps de résidence dans le RE pourrait soit modifié, ralenti. Il pourrait être comparé à celui de la protéine sauvage par western-blot et immunofluorescence à l'aide d'un système d'expression inductible.

La fonction principale des N-glycosylations est d'assurer le contrôle qualité du repliement des protéines néosynthétisées. L'insertion de mutations peut être à l'origine d'un défaut du repliement protéique, c'est pourquoi, nous avons analysé le trafic intracellulaire des mutants E9D18A de MERS-CoV-M et E11,12,18A de SARS-CoV-2-M. Leur co-localisation avec le marqueur TGN46 suggèrait que les mutants étaient correctement repliés, où dans le cas contraire, ils auraient été retenus dans le RE avant d'être envoyé vers la voie de dégradation (ERAD). Nous nous sommes également assurés par (i) des expériences de perméabilisation sélective des membranes cellulaires et par (ii) l'insertion d'une séquence linker contenant un site de glycosylation, que l'ajout des mutations n'a pas impacté la topologie membranaire de M, puisqu'il a notamment été rapporté que la mutation de résidus chargés inversait la topologie de la protéase Lep de *Escherichia coli* (Nilsson et Von Heijne 1990). En effet, il est primordial que l'asparagine acceptrice soit orientée côté luminal pour être accessible à OST. Les résultats obtenus par

ces expériences ont indiqué qu'aucun changement de la topologie membranaires n'a eu lieu.

La séquence linker que nous avons utilisé pour nos constructions M-linker se compose de 31 résidus hydrophiles (YAYNRYRIGNYKL**N****T****T****I****K****D****L****D****Y****K****D****D****D****D****K****P****K****E**), dont 7 résidus correspondent à des résidus acides (6 acides aspartiques, 1 acide glutamique). Il n'est pas à exclure que la présence de ces résidus acides, en nombre important, puisse agir sur la glycosylation de M. Cette séquence a été construite et utilisée par Voss et ses collègues pour étudier la topologie membranaire de la protéine M du SARS-CoV (Voß et al. 2009). Nous pourrions remplacer cette séquence par une séquence contenant davantage de résidus polaires neutres, afin d'évaluer l'effet des résidus acides présents dans la séquence utilisée, sur la glycosylation de M. Il serait aussi intéressant de regarder la machinerie qui est responsable de la glycosylation du site N88.

Récemment, la structure de SARS-CoV-2-M a été résolue par cryo-EM (Dolan et al. 2022 ; Zhang et al. 2022). Selon la structure de Dolan *et al.*, l'ectodomaine de M est composé des résidus 1-16, contre les résidus 1-8 selon la structure de Zhang *et al.*, Il se pourrait donc que les résidus acides identifiés précédemment fassent partis du premier segment transmembranaire de M.

Si tel n'est pas le cas, il est alors envisageable que ces résidus soient engagés dans des interactions électrostatiques avec le complexe OST-B. Des expériences de co-immunoprécipitation du complexe M-STT3B pourraient être réalisées afin de vérifier cette hypothèse. L'interactions entre ces deux protéines devant être relativement brève, une étape de « *cross-linking* » serait nécessaire au-préalable afin de stabiliser le complexe.

Chez les eucaryotes, le potentiel électrostatique des membranes dépend des phosphatidylsérines (PS), un lipide chargé négativement (Bigay et Antonny 2012). De plus, le feuillet interne du RE semble être davantage enrichi en PS que le feuillet externe. Ainsi, il serait également possible que les résidus acides identifiés pour la glycosylation de MERS-CoV-M et SARS-CoV-2-M puissent induire une répulsion électrostatique du domaine N-terminal de la protéine avec la membrane du RE, ce qui pourrait rendre le site de glycosylation plus facilement accessible à OST-B.

Ci-dessous est représenté un alignement de séquences correspondants aux domaines N-terminaux de plusieurs protéines M de coronavirus modifiées par des N-

IBV	MSNET NCTLDFEQSVELFKEY NLFIATFLLFLTIIILQYGYATR SKFIYILKMI VLWCFWP	60
229E	---MSN-D NCTGD IVTHLKNWNFGWNVILTI FIVILQFGHYKYSR LFYGLKMLVLWLLWP	56
NL63	---MS NSSVPLSE VYVHLRNWNFSWNLILTVFIVVLQYGHYKYSRLLYGLKMSVLWCLWP	57
SARS	--MAD NGTITV EELKQLLE QWNLVIGFLFLAWIMLLQFAYSNNRNFYI IKLV FLWLLWP	58
COV2	-MAD SNGTITV EELKLLLE QWNLVIGFLFLTWICLLQFAYANRRNFYI IKLI FLWLLWP	59
HKU1	MNKS FLPQFTSDQAVTFLK EW NFSLGVILLFITIIILQFGYTSRSMFVYLIKMIILWLMWP	60
MERS	--MS NMTQLTE AQIIAIK DWN FAWSLIFLLITIVLQYGYPSRSM TVYVFKM FVLWLLWP	58
HKU4	--M SSNGSLTK DE VVNI KDWN FSWSIIIFLLITIVLQYGYPSRSM MVYVFKM FILWLLWP	58
HKU5	-M ASNVTLSN DE VLRLVK DWN FTWSVVFLITIVLQYGYPSRSMFVYVI KMFV LWLLWP	59
	: . . . : * : . : : * : : . . . * : * : . * * : * *	

glycanes (**Figure 29**). Nous pouvons observer que des résidus acides sont conservés au niveau de deux régions situées en aval des sites de glycosylation.

La conservation de ces résidus suggère une importance pour la protéine. Ainsi, la mutation de ces résidus pourrait être envisagée afin d’analyser si les effets observés sur la protéine M du MERS-CoV et du SARS-CoV-2 sont également observés sur d’autres protéines M.

Figure 29. Alignements de séquences en acide-aminé du domaine N-terminal de plusieurs protéines M de coronavirus modifiées par des N-glycanes.

Le virus correspondant à la séquence protéique est indiqué sur la gauche (SARS = SARS-CoV ; COV2 = SARS-CoV-2). Le site consensus de glycosylation est marqué en gras et les résidus acides sont écrits en rouge. Les séquences sont issues de la banque de données Uniprot et les alignements ont été réalisés avec l’algorithme CLUSTAL O(1.2.4).

Dans notre étude, nous avons construit une protéine chimère où le domaine N-terminal de la protéine E1 du virus HCV (MAQVK**NSSGLY**) a remplacé celui de SARS-CoV-2-M (MAD**SNGTIT**) (E1/M). Cette protéine possède 5 sites de glycosylation, mais seulement 4 sont fonctionnels et portent des N-glycanes riches en mannose, dû à la rétention de E1 dans le RE (Dubuisson et al. 1994 ; Meunier et al. 1999). Nous nous sommes assurés de cette rétention en réalisant un co-marquage de E1 avec la calnexine, une lectine chaperonne du RE. Également, l’accessibilité à la structure de la protéine ORF3a du SARS-CoV-2, résolue expérimentalement par cryo-EM, nous a permis de générer des protéines chimères M/ORF3a, où le domaine N-terminal de M (Nt-M/ORF3a) ou le domaine N-terminal et le premier segment transmembranaire de M (TM1-M/ORF3a) ont remplacé ceux de ORF3a (Kern et al. 2021). Par ailleurs, il s’avère que ORF3a présente de très fortes homologies structurales avec SARS-CoV-2-M, suggérant une possible origine commune. Son domaine N-terminal est plus long (1-39) que celui de

M, en revanche, son premier segment transmembranaire est plus court (1-18) et ne contient aucun résidu chargé.

En western-blot, le profil de migration de la chimère E1/M était semblable à celui de la protéine M sauvage, ce qui pourrait indiquer que des chaînes de polylectosamines ont été ajoutées sur le premier site accepteur de E1. Un test de précipitation par les lectines LEL et DSL pourrait le confirmer. Par conséquent, ces résultats confirment qu'aucun signal pour l'ajout des chaînes de polylectosamines n'est présent dans le domaine N-terminal de M.

Concernant la chimère Nt-M/ORF3a, le profil de migration observé semblait indiquer qu'aucune glycosylation n'avait eu lieu. Ce résultat n'était pas inattendu puisque le premier segment transmembranaire de ORF3a ne possède aucun résidu acide. En revanche, des espèces glycosylées ont été observées avec la chimère TM1-M/ORF3a, où, ici, les résidus acides sont présents. La détection d'une bande diffuse en western-blot indique que des polylectosaminoglycanes ont été ajoutées sur le site accepteur N5 de M. Cela confirme davantage l'importance des résidus acides pour la glycosylation de M. Cependant, nous ne pouvons pas exclure la possibilité qu'un motif régulant l'ajout des chaînes de polylectosamines soit présent dans TM1-M. En effet, des polylectosaminoglycanes ont été ajoutés à TM1-M/ORF3a puisque la glycosylation de la protéine a pu être initiée dans le RE par la présence des résidus acides.

Les β 4GalT et B3GnT sont des familles de glycosyltransférases de l'appareil de Golgi/TGN responsables de la synthèse des chaînes de polylectosamines (Nilsson et al. 1991; Lee et al. 2009; Kimber 2021). Il ne serait alors pas absurde de penser que la rétention des protéines dans ce compartiment favoriserait l'addition des chaînes de polylectosamines, puisqu'elles seraient davantage exposées aux glycosyltransférases. Cependant, il s'avère que ce n'est pas réellement le cas. En effet, la chimère TM1-M/ORF3a portent des chaînes de polylectosamines et n'est pas retenue dans le TGN, puisqu'aucune co-localisation n'a été observée avec le marqueur TGN46. Le signal de fluorescence observé pour cette chimère était semblable à celui d'ORF3a, à savoir, un marquage ponctué, plutôt caractéristique d'une localisation intravésiculaire.

Certaines études ont démontré que ORF3a est localisée dans le compartiment endolysosomal et à la surface cellulaire (Miller et al. 2023). Nous avons cherché à confirmer l'adressage d'ORF3a vers les endosomes/lysosomes par des expériences de co-

marquages avec les GTPases Rab5 (endosomes précoces), Rab7 (endosomes tardifs), Rab11 (endosomes de recyclage) et LAMP1 (lysosomes). Nos essais n'ont cependant pas été concluants.

La chimère E1/M est très probablement modifiée par des chaînes de polylectosamines et est retenue dans le TGN. Cependant, l'intensité de la bande diffuse observée en western-blot n'est pas plus intense que celle de TM1-M/ORF3, ce qui indique que E1-M ne serait pas davantage modifiée par des chaînes de polylectosamines, malgré sa rétention dans le RE. Évidemment, le niveau d'expression de ces deux chimères devra être déterminé afin de pouvoir comparer l'intensité des bandes de manière significative.

MERS-CoV-M possède un motif dans sa queue C-terminale responsable de la rétention de la protéine dans le TGN ($_{199}\text{KxGxYR}_{204}$) (Perrier et al. 2019). Dans cette étude, il a été montré que la mutation de ce motif engendrait une relocalisation d'environ 15 % de la quantité totale de M, à la surface cellulaire, contre moins de 1 % pour la protéine sauvage. En western-blot, la bande diffuse du mutant M-KxGxYR était plus intense que celle de M sauvage. Bien que la quantité de protéine non-glycosylée n'ait pas été calculée, le profil de migration laisse penser que M-KxGxYR serait davantage modifié que M sauvage par des chaînes de polylectosamines. Ainsi, le trafic jusqu'à la membrane plasmique pourrait être lié à cet effet.

Mentionné précédemment, les protéines LAMP1 et LAMP2 sont des protéines qui résident dans les lysosomes et sont modifiées par des polylectosaminoglycanes sur quelques sites accepteurs. En revanche, il a été montré, pour LAMP2, que la rétention de la protéine dans le Golgi n'était pas à l'origine de l'ajout des chaînes de polylectosamines, mais plutôt son temps de résidence dans ce compartiment (Nabi et Dennis 1998).

Ainsi, toutes ces données semblent indiquer que la rétention protéique dans le Golgi/TGN ne favoriserait pas davantage l'extension par des chaînes de polylectosamines.

Il est admis que la glycosylation de M n'est pas requise pour l'assemblage des coronavirus (De Haan et al. 1998; Liang et al. 2019). Nous l'avons confirmés pour le MERS-CoV et le SARS-CoV-2, où des particules VLP produites avec MERS-CoV-N3Q-M ou SARS-CoV-2-N5Q-M, ont été sécrétées dans le milieu de culture. Nous avons aussi constaté que la forme de M glycoconjuguée à des N-glycanes riches en mannose est davantage incorporée dans les VLP que les autres formes. L'étape d'assemblage des coronavirus se déroulant au sein du compartiment ERGIC, il n'est pas surprenant que cette forme soit

majoritaire dans les VLP. En revanche, l'absence de formes associées à des chaînes de polylectosamines indiquent que la protéine M ne subit plus de modifications de ces glycanes après le bourgeonnement des particules virales.

Une récente étude a démontré que les *betacoronavirus* pouvaient détourner la voie lysosomale pour sortir des cellules par exocytose (Ghosh et al. 2020). Cependant, la route empruntée par les particules virales pour rejoindre les lysosomes n'est pas encore clairement définie. En effet, les particules pourraient atteindre les lysosomes après avoir transité à travers l'appareil de Golgi, ou directement après leur bourgeonnement au niveau du compartiment ERGIC. Ce dernier cas expliquerait pourquoi les glycanes de M ne sont pas davantage modifiés par des glycanes complexes. Cependant, il est judicieux de prendre en considération que des variations peuvent exister entre les VLP et les particules virales infectieuses. Par exemple, seule la forme IBV-M non-glycosylée est incorporée dans les VLP, alors qu'une espèce glycosylée est observée dans les particules infectieuses (Liang et al. 2019). La synthèse de VLP nécessite, à minima, la co-expression des protéines M et E dans des cultures de cellules. Il est fort probable que la forte expression de ces protéines provoque une saturation au niveau du RE, pouvant amener à un bourgeonnement précoce des particules. Ainsi, il serait intéressant d'étudier les formes de M associées aux particules virales infectieuses. Des tests ont été initiés pendant ce projet, mais n'ont pu aboutir pour des raisons techniques.

Pour conclure ce projet de thèse, nos travaux ont permis d'approfondir nos connaissances sur la glycosylation de la protéine M du MERS-CoV et du SARS-CoV-2. Nous avons caractérisé les N-glycanes qui arborent MERS-CoV-M et SARS-CoV-2-M, ainsi que les mécanismes impliqués dans la glycosylation de ces protéines.

Cette modification est conservée chez tous les coronavirus, sous forme de N- ou O-glycosylation selon le genre du virus, indiquant un rôle pour le cycle infectieux. Cependant, peu d'études permettent de définir réellement la fonction de ces glycanes. Liang *et al.*, ont observé que la glycosylation de IBV-M n'influçait pas le trafic intracellulaire de la protéine, et ne modifiait pas la réplication *in vitro* (Liang et al. 2019). En revanche, ils ont observé une activité apoptotique plus importante des cellules Vero infectées par du virus recombinant sauvage comparé à ces mêmes cellules infectées par du virus recombinant déficient en N-glycosylation sur IBV-M. L'augmentation de l'apoptose s'est également accompagnée par une augmentation du stress du RE. Il a été suggéré par les auteurs que

l'activation de la voie UPR (*Unfolded protein response*) est à l'origine de l'induction de l'apoptose et du stress du RE, à cause d'une synthèse protéique forte au niveau du réticulum. Une seconde étude, consacrée à MHV, a observé que le statut de glycosylation de M pouvait influencer la réplication virale dans le foie mais pas le cerveau (De Haan et al. 2003). Il est alors possible que les glycosylations de MHV-M puisse interagir avec des facteurs d'attachement types lectines exposés à la surface des cellules hépatiques, favorisant l'entrée virale dans ces cellules. Enfin, plusieurs études ont mis en évidence une activité interférogénique de la protéine M du TGEV sur des cellules PBMC (*Peripheral blood mononuclear cells*) qui serait liée à la présence de glycanes complexes liés à TGEV-M (Charley et al. 1991; Laude et al. 1992). Il serait intéressant de déterminer si cette protéine est aussi modifiée par des chaînes de polylactosamines. Cette activité a aussi été retrouvée avec la protéine MHV-M (De Haan et al. 2003). Par ailleurs, cette activité était plus importante lorsque MHV-M était modifiée afin de porter des N-glycanes à la place de O-glycanes. Des chaînes de polylactosamines peuvent aussi être ajoutées sur des O-glycanes, il serait donc intéressant de caractériser aussi les glycanes portés par MHV-M. Étudier cette activité interférogénique sur les protéines MERS-CoV-M et SARS-CoV-2-M pourrait être une piste à envisager. En revanche, cette activité n'a jamais été confirmée *in vivo*.

Toutes ces données semblent indiquer que les glycosylations de la protéine M des coronavirus interviennent dans des interactions avec la cellule hôte. Compte-tenu du caractère émergent des coronavirus, la compréhension de leur biologie s'avère crucial.

BIBLIOGRAPHIE

Allen JD, Ivory DP, Song SG, He W ting, Capozzola T, Yong P, et al. The diversity of the glycan shield of sarbecoviruses related to SARS-CoV-2. *Cell Reports*. avr 2023;42(4):112307.

Amraei R, Yin W, Napoleon MA, Suder EL, Berrigan J, Zhao Q, et al. CD209L/L-SIGN and CD209/DC-SIGN Act as Receptors for SARS-CoV-2. *ACS Cent Sci*. 28 juill 2021;7(7):1156-65.

Anderson K, King AMQ, Lerch RA, Wertz GW. Polylactosaminoglycan modification of the respiratory syncytial virus small hydrophobic (SH) protein: A conserved feature among human and bovine respiratory syncytial viruses. *Virology*. nov 1992;191(1):417-30.

Angelini MM, Akhlaghpour M, Neuman BW, Buchmeier MJ. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Proteins 3, 4, and 6 Induce Double-Membrane Vesicles. *Moscona A, éditeur. mBio [Internet]*. 30 août 2013 [cité 10 nov 2021];4(4). Disponible sur: <https://journals.asm.org/doi/10.1128/mBio.00524-13>

Ann Yeh E, Collins A, Cohen ME, Duffner PK, Faden H. Detection of Coronavirus in the Central Nervous System of a Child With Acute Disseminated Encephalomyelitis. *Pediatrics*. 1 janv 2004;113(1):e73-6.

Araf Y, Akter F, Tang Y, Fatemi R, Parvez MdSA, Zheng C, et al. Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. *Journal of Medical Virology*. mai 2022;94(5):1825-32.

Arden KE, Nissen MD, Sloots TP, Mackay IM. New human coronavirus, HCoV-NL63, associated with severe lower respiratory tract disease in Australia. *J Med Virol*. mars 2005;75(3):455-62.

Armstrong J, Niemann H, Smeekens S, Rottiert P, Warren G. Sequence and topology of a

model intracellular membrane protein, El glycoprotein, from a coronavirus. 1984;2.

Arndt AL, Larson BJ, Hogue BG. A Conserved Domain in the Coronavirus Membrane Protein Tail Is Important for Virus Assembly. *J Virol.* nov 2010;84(21):11418-28.

Bai Z, Cao Y, Liu W, Li J. The SARS-CoV-2 Nucleocapsid Protein and Its Role in Viral Structure, Biological Functions, and a Potential Target for Drug or Vaccine Mitigation. *Viruses.* 10 juin 2021;13(6):1115.

Bastien N, Anderson K, Hart L, Van Caesele P, Brandt K, Milley D, et al. Human Coronavirus NL63 Infection in Canada. *J INFECT DIS.* 15 févr 2005;191(4):503-6.

Baudoux P, Besnardeau L, Carrat C, Rottier P, Charley B, Laude H. Interferon Alpha Inducing Property of Coronavirus Particles and Pseudoparticles. In: Enjuanes L, Siddell SG, Spaan W, éditeurs. *Coronaviruses and Arteriviruses* [Internet]. Boston, MA: Springer US; 1998 [cité 10 nov 2021]. p. 377-86. (Advances in Experimental Medicine and Biology; vol. 440). Disponible sur: http://link.springer.com/10.1007/978-1-4615-5331-1_49

Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci USA.* 7 avr 2009;106(14):5871-6.

Belouzard S, Millet JK, Licitra BN, Whittaker GR. Mechanisms of Coronavirus Cell Entry Mediated by the Viral Spike Protein. *Viruses.* 20 juin 2012;4(6):1011-33.

Berndsen ZT, Chakraborty S, Wang X, Cottrell CA, Torres JL, Diedrich JK, et al. Visualization of the HIV-1 Env glycan shield across scales. *Proc Natl Acad Sci USA.* 10 nov 2020;117(45):28014-25.

Bestle D, Heindl MR, Limburg H, Van Lam Van T, Pilgram O, Moulton H, et al. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci Alliance.* sept 2020;3(9):e202000786.

Betakova T, Nermut MV, Hay AJ. The NB protein is an integral component of the

membrane of influenza B virus. *Journal of General Virology*. 1 nov 1996;77(11):2689-94.
Bigay J, Antony B. Curvature, Lipid Packing, and Electrostatics of Membrane Organelles: Defining Cellular Territories in Determining Specificity. *Developmental Cell*. nov 2012;23(5):886-95.

Bojar D, Meche L, Meng G, Eng W, Smith DF, Cummings RD, et al. A Useful Guide to Lectin Binding: Machine-Learning Directed Annotation of 57 Unique Lectin Specificities. *ACS Chem Biol*. 18 nov 2022;17(11):2993-3012.

Bolles M, Deming D, Long K, Agnihothram S, Whitmore A, Ferris M, et al. A Double-Inactivated Severe Acute Respiratory Syndrome Coronavirus Vaccine Provides Incomplete Protection in Mice and Induces Increased Eosinophilic Proinflammatory Pulmonary Response upon Challenge. *J Virol*. déc 2011;85(23):12201-15.

Bonnin A, Danneels A, Dubuisson J, Goffard A, Belouzard S. HCoV-229E spike protein fusion activation by trypsin-like serine proteases is mediated by proteolytic processing in the S2' region. *Journal of General Virology*. 1 juill 2018;99(7):908-12.

Booth CM. Clinical Features and Short-term Outcomes of 144 Patients With SARS in the Greater Toronto Area. *JAMA*. 4 juin 2003;289(21):2801.

Booth TF, Kournikakis B, Bastien N, Ho J, Kobasa D, Stadnyk L, et al. Detection of Airborne Severe Acute Respiratory Syndrome (SARS) Coronavirus and Environmental Contamination in SARS Outbreak Units. *The Journal of Infectious Diseases*. 1 mai 2005;191(9):1472-7.

Boscarino JA, Logan HL, Lacny JJ, Gallagher TM. Envelope Protein Palmitoylations Are Crucial for Murine Coronavirus Assembly. *J Virol*. 15 mars 2008;82(6):2989-99.

Boson B, Legros V, Zhou B, Siret E, Mathieu C, Cosset FL, et al. The SARS-CoV-2 envelope and membrane proteins modulate maturation and retention of the spike protein, allowing assembly of virus-like particles. *Journal of Biological Chemistry*. janv 2021;296:100111.

Bosshart H, Humphrey J, Deignan E, Davidson J, Drazba J, Yuan L, et al. The cytoplasmic domain mediates localization of furin to the trans-Golgi network en route to the endosomal/lysosomal system. *The Journal of cell biology*. 1 sept 1994;126(5):1157-72.

Bouvet M, Debarnot C, Imbert I, Selisko B, Snijder EJ, Canard B, et al. In Vitro Reconstitution of SARS-Coronavirus mRNA Cap Methylation. Buchmeier MJ, éditeur. *PLoS Pathog*. 22 avr 2010;6(4):e1000863.

Brown AJ, Won JJ, Graham RL, Dinnon KH, Sims AC, Feng JY, et al. Broad spectrum antiviral remdesivir inhibits human endemic and zoonotic deltacoronaviruses with a highly divergent RNA dependent RNA polymerase. *Antiviral Research*. sept 2019;169:104541.

Cabeça TK, Granato C, Bellei N. Epidemiological and clinical features of human coronavirus infections among different subsets of patients. *Influenza Other Respi Viruses*. nov 2013;7(6):1040-7.

Cabrera-Garcia D, Bekdash R, Abbott GW, Yazawa M, Harrison NL. The envelope protein of SARS-CoV-2 increases intra-Golgi pH and forms a cation channel that is regulated by pH. *J Physiol*. juin 2021;599(11):2851-68.

Callaway E. The coronavirus is mutating — does it matter? *Nature*. 10 sept 2020;585(7824):174-7.

Cao Z, Xia H, Rajsbaum R, Xia X, Wang H, Shi PY. Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response. *Cell Mol Immunol*. mars 2021;18(3):746-8.

Carlsson SR, Fukuda M. The polylactosaminoglycans of human lysosomal membrane glycoproteins lamp-1 and lamp-2. Localization on the peptide backbones. *Journal of Biological Chemistry*. nov 1990;265(33):20488-95.

Casais R, Dove B, Cavanagh D, Britton P. Recombinant Avian Infectious Bronchitis Virus Expressing a Heterologous Spike Gene Demonstrates that the Spike Protein Is a Determinant of Cell Tropism. *J Virol*. 15 août 2003;77(16):9084-9.

Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, Harbison AM, et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent Sci.* 28 oct 2020;6(10):1722-34.

Chan KS, Lai ST, Chu CM, Tsui E, Tam CY, Wong MML, et al. Treatment of severe acute respiratory syndrome with lopinavir/ritonavir: a multicentre retrospective matched cohort study. *Hong Kong Med J.* déc 2003;9(6):399-406.

Charley B, Laude H. Induction of Alpha Interferon by Transmissible Gastroenteritis Coronavirus: Role of Transmembrane Glycoprotein El. 1988;62:4.

Charley B, Lavenant L, Delmas B. Glycosylation is Required for Coronavirus TGEV to Induce an Efficient Production of IFNalpha by Blood Mononuclear Cells. *Scand J Immunol.* avr 1991;33(4):435-40.

Chen D, Zheng Q, Sun L, Ji M, Li Y, Deng H, et al. ORF3a of SARS-CoV-2 promotes lysosomal exocytosis-mediated viral egress. *Developmental Cell.* déc 2021;56(23):3250-3263.e5.

Chiu SS, Hung Chan K, Wing Chu K, Kwan SW, Guan Y, Man Poon LL, et al. Human Coronavirus NL63 Infection and Other Coronavirus Infections in Children Hospitalized with Acute Respiratory Disease in Hong Kong, China. *Clinical Infectious Diseases.* 15 juin 2005;40(12):1721-9.

Choi JY, Smith DM. SARS-CoV-2 Variants of Concern. *Yonsei Med J.* 2021;62(11):961.

Chu VC, McElroy LJ, Chu V, Bauman BE, Whittaker GR. The Avian Coronavirus Infectious Bronchitis Virus Undergoes Direct Low-pH-Dependent Fusion Activation during Entry into Host Cells. *J Virol.* avr 2006;80(7):3180-8.

Cinaud A, Sorbets E, Blachier V, Vallee A, Kretz S, Lelong H, et al. Hypertension artérielle et COVID-19. *La Presse Médicale Formation.* févr 2021;2(1):25-32.

Clausen TM, Sandoval DR, Spliid CB, Pihl J, Perrett HR, Painter CD, et al. SARS-CoV-2

Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell*. nov 2020;183(4):1043-1057.e15.

Clementz MA, Kanjanahaluethai A, O'Brien TE, Baker SC. Mutation in murine coronavirus replication protein nsp4 alters assembly of double membrane vesicles. *Virology*. mai 2008;375(1):118-29.

Cong Y, Ulasli M, Schepers H, Mauthe M, V'kovski P, Kriegenburg F, et al. Nucleocapsid Protein Recruitment to Replication-Transcription Complexes Plays a Crucial Role in Coronaviral Life Cycle. *Dutch RE, éditeur. J Virol*. 31 janv 2020;94(4):e01925-19.

Cornillez-Ty CT, Liao L, Yates JR, Kuhn P, Buchmeier MJ. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling. *J Virol*. oct 2009;83(19):10314-8.

Corse E, Machamer CE. Infectious Bronchitis Virus E Protein Is Targeted to the Golgi Complex and Directs Release of Virus-Like Particles. *J Virol*. mai 2000;74(9):4319-26.

Corse E, Machamer CE. The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction. *Virology*. juill 2003;312(1):25-34.

Cottam EM, Whelband MC, Wileman T. Coronavirus NSP6 restricts autophagosome expansion. *Autophagy*. 20 août 2014;10(8):1426-41.

Cui L, Wang H, Ji Y, Yang J, Xu S, Huang X, et al. The Nucleocapsid Protein of Coronaviruses Acts as a Viral Suppressor of RNA Silencing in Mammalian Cells. *Perlman S, éditeur. J Virol*. sept 2015;89(17):9029-43.

De Haan CAM, Kuo L, Masters PS, Vennema H, Rottier PJM. Coronavirus Particle Assembly: Primary Structure Requirements of the Membrane Protein. *J Virol*. août 1998;72(8):6838-50.

De Haan CAM, Vennema H, Rottier PJM. Assembly of the Coronavirus Envelope: Homotypic Interactions between the M Proteins. *J Virol.* juin 2000;74(11):4967-78.

DeDiego ML, Álvarez E, Almazán F, Rejas MT, Lamirande E, Roberts A, et al. A Severe Acute Respiratory Syndrome Coronavirus That Lacks the E Gene Is Attenuated In Vitro and In Vivo. *J Virol.* 15 févr 2007;81(4):1701-13.

DeDiego ML, Nieto-Torres JL, Jiménez-Guardeño JM, Regla-Nava JA, Álvarez E, Oliveros JC, et al. Severe Acute Respiratory Syndrome Coronavirus Envelope Protein Regulates Cell Stress Response and Apoptosis. Thiel V, éditeur. *PLoS Pathog.* 20 oct 2011;7(10):e1002315.

Delmas B, Laude H. Assembly of coronavirus spike protein into trimers and its role in epitope expression. *J Virol.* nov 1990;64(11):5367-75.

Deng X, Hackbart M, Mettelman RC, O'Brien A, Mielech AM, Yi G, et al. Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc Natl Acad Sci USA* [Internet]. 23 mai 2017 [cité 9 mai 2023];114(21). Disponible sur: <https://pnas.org/doi/full/10.1073/pnas.1618310114>

Desforges M, Desjardins J, Zhang C, Talbot PJ. The Acetyl-Esterase Activity of the Hemagglutinin-Esterase Protein of Human Coronavirus OC43 Strongly Enhances the Production of Infectious Virus. *J Virol.* 15 mars 2013;87(6):3097-107.

Dijkman R, Jebbink MF, El Idrissi NB, Pyrc K, Müller MA, Kuijpers TW, et al. Human Coronavirus NL63 and 229E Seroconversion in Children. *J Clin Microbiol.* juill 2008;46(7):2368-73.

Dolan KA, Dutta M, Kern DM, Kotecha A, Voth GA, Brohawn SG. Structure of SARS-CoV-2 M protein in lipid nanodiscs [Internet]. *Biophysics*; 2022 juin. Disponible sur: <http://biorxiv.org/lookup/doi/10.1101/2022.06.12.495841>

Dubuisson J, Hsu HH, Cheung RC, Greenberg HB, Russell DG, Rice CM. Formation and

intracellular localization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia and Sindbis viruses. *J Virol.* oct 1994;68(10):6147-60.

Ducatelle R, Hoorens J. Significance of lysosomes in the morphogenesis of coronaviruses. *Archives of Virology.* mars 1984;79(1-2):1-12.

Dveksler GS, Pensiero MN, Dieffenbach CW, Cardellichio CB, Basile AA, Elia PE, et al. Mouse hepatitis virus strain A59 and blocking antireceptor monoclonal antibody bind to the N-terminal domain of cellular receptor. *Proc Natl Acad Sci USA.* mars 1993;90(5):1716-20.

Ebihara T, Endo R, Ma X, Ishiguro N, Kikuta H. Detection of human coronavirus NL63 in young children with bronchiolitis. *J Med Virol.* mars 2005;75(3):463-5.

Egloff MP, Ferron F, Campanacci V, Longhi S, Rancurel C, Dutartre H, et al. The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc Natl Acad Sci USA.* 16 mars 2004;101(11):3792-6.

Enjuanes L, Zuñiga S, Castaño-Rodríguez C, Gutierrez-Alvarez J, Canton J, Sola I. Molecular Basis of Coronavirus Virulence and Vaccine Development. In: *Advances in Virus Research* [Internet]. Elsevier; 2016 [cité 21 avr 2023]. p. 245-86. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S0065352716300422>

Esper F, Weibel C, Ferguson D, Landry ML, Kahn JS. Coronavirus HKU1 Infection in the United States. *Emerg Infect Dis.* mai 2006;12(5):775-9.

Eymieux S, Uzbekov R, Rouillé Y, Blanchard E, Hourieux C, Dubuisson J, et al. Secretory Vesicles Are the Principal Means of SARS-CoV-2 Egress. *Cells.* 10 août 2021;10(8):2047.

Finkel Y, Mizrahi O, Nachshon A, Weingarten-Gabbay S, Morgenstern D, Yahalom-Ronen Y, et al. The coding capacity of SARS-CoV-2. *Nature.* 7 janv 2021;589(7840):125-30.

Fischer F, Stegen CF, Masters PS, Samsonoff WA. Analysis of Constructed E Gene Mutants

of Mouse Hepatitis Virus Confirms a Pivotal Role for E Protein in Coronavirus Assembly. *J Virol.* oct 1998;72(10):7885-94.

Fouchier RAM, Hartwig NG, Bestebroer TM, Niemeyer B, de Jong JC, Simon JH, et al. A previously undescribed coronavirus associated with respiratory disease in humans. *Proc Natl Acad Sci USA.* 20 avr 2004;101(16):6212-6.

Fu YZ, Wang SY, Zheng ZQ, Yi Huang, Li WW, Xu ZS, et al. SARS-CoV-2 membrane glycoprotein M antagonizes the MAVS-mediated innate antiviral response. *Cell Mol Immunol.* mars 2021;18(3):613-20.

Gadlage MJ, Sparks JS, Beachboard DC, Cox RG, Doyle JD, Stobart CC, et al. Murine Hepatitis Virus Nonstructural Protein 4 Regulates Virus-Induced Membrane Modifications and Replication Complex Function. *J Virol.* janv 2010;84(1):280-90.

Ghosh S, Dellibovi-Ragheb TA, Kerviel A, Pak E, Qiu Q, Fisher M, et al. β -Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell.* déc 2020;183(6):1520-1535.e14.

Glowacka I, Bertram S, Müller MA, Allen P, Soilleux E, Pfefferle S, et al. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *J Virol.* mai 2011;85(9):4122-34.

Gong Y, Qin S, Dai L, Tian Z. The glycosylation in SARS-CoV-2 and its receptor ACE2. *Sig Transduct Target Ther.* déc 2021;6(1):396.

Graham RL, Sims AC, Brockway SM, Baric RS, Denison MR. The nsp2 Replicase Proteins of Murine Hepatitis Virus and Severe Acute Respiratory Syndrome Coronavirus Are Dispensable for Viral Replication. *J Virol.* nov 2005;79(21):13399-411.

Granovsky M, Fata J, Pawling J, Muller WJ, Khokha R, Dennis JW. Suppression of tumor growth and metastasis in Mgat5-deficient mice. *Nat Med.* mars 2000;6(3):306-12.

Grant OC, Montgomery D, Ito K, Woods RJ. Analysis of the SARS-CoV-2 spike protein glycan shield: implications for immune recognition [Internet]. Immunology; 2020 avr. Disponible sur: <http://biorxiv.org/lookup/doi/10.1101/2020.04.07.030445>

Grant SN, Lester HA. Regulation of epithelial sodium channel activity by SARS-CoV-1 and SARS-CoV-2 proteins. Biophysical Journal. juill 2021;120(14):2805-13.

Haagmans BL, Van Den Brand JMA, Raj VS, Volz A, Wohlsein P, Smits SL, et al. An orthopoxvirus-based vaccine reduces virus excretion after MERS-CoV infection in dromedary camels. Science. janv 2016;351(6268):77-81.

de Haan CAM, Masters PS, Shen X, Weiss S, Rottier PJM. The Group-Specific Murine Coronavirus Genes Are Not Essential, but Their Deletion, by Reverse Genetics, Is Attenuating in the Natural Host. Virology. avr 2002;296(1):177-89.

de Haan CAM, de Wit M, Kuo L, Montalto-Morrison C, Haagmans BL, Weiss SR, et al. The glycosylation status of the murine hepatitis coronavirus M protein affects the interferogenic capacity of the virus in vitro and its ability to replicate in the liver but not the brain. Virology. août 2003;312(2):395-406.

Hadj Hassine I. Covid-19 vaccines and variants of concern: A review. Reviews in Medical Virology [Internet]. juill 2022 [cité 25 avr 2023];32(4). Disponible sur: <https://onlinelibrary.wiley.com/doi/10.1002/rmv.2313>

Hamre D, Procknow JJ. A New Virus Isolated from the Human Respiratory Tract. Experimental Biology and Medicine. 1 janv 1966;121(1):190-3.

Han DP, Lohani M, Cho MW. Specific Asparagine-Linked Glycosylation Sites Are Critical for DC-SIGN- and L-SIGN-Mediated Severe Acute Respiratory Syndrome Coronavirus Entry. J Virol. nov 2007a;81(21):12029-39.

Han TH, Chung JY, Kim SW, Hwang ES. Human Coronavirus-NL63 infections in Korean children, 2004–2006. Journal of Clinical Virology. janv 2007b;38(1):27-31.

Heath PT, Galiza EP, Baxter DN, Boffito M, Browne D, Burns F, et al. Safety and Efficacy of NVX-CoV2373 Covid-19 Vaccine. *N Engl J Med*. 23 sept 2021;385(13):1172-83.

Henion TR, Faden AA, Knott TK, Schwarting GA. 3GnT2 Maintains Adenylyl Cyclase-3 Signaling and Axon Guidance Molecule Expression in the Olfactory Epithelium. *Journal of Neuroscience*. 27 avr 2011;31(17):6576-86.

Hillen HS. Structure and function of SARS-CoV-2 polymerase. *Current Opinion in Virology*. juin 2021;48:82-90.

van der Hoek L, Pyrc K, Berkhout B. Human coronavirus NL63, a new respiratory virus. *FEMS Microbiol Rev*. sept 2006;30(5):760-73.

Hoffmann M, Kleine-Weber H, Pöhlmann S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Molecular Cell*. mai 2020;78(4):779-784.e5.

Hofmann H, Pyrc K, Van Der Hoek L, Geier M, Berkhout B, Pöhlmann S. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc Natl Acad Sci USA*. 31 mai 2005;102(22):7988-93.

Hofmann H, Simmons G, Rennekamp AJ, Chaipan C, Gramberg T, Heck E, et al. Highly Conserved Regions within the Spike Proteins of Human Coronaviruses 229E and NL63 Determine Recognition of Their Respective Cellular Receptors. *J Virol*. sept 2006;80(17):8639-52.

Hsieh PK, Chang SC, Huang CC, Lee TT, Hsiao CW, Kou YH, et al. Assembly of Severe Acute Respiratory Syndrome Coronavirus RNA Packaging Signal into Virus-Like Particles Is Nucleocapsid Dependent. *J Virol*. 15 nov 2005;79(22):13848-55.

Hu Q, He L, Zhang Y. Community Transmission via Indirect Media-To-Person Route: A Missing Link in the Rapid Spread of COVID-19. *Front Public Health*. 28 juill 2021;9:687937.

Huang C, Ito N, Tseng CTK, Makino S. Severe Acute Respiratory Syndrome Coronavirus 7a Accessory Protein Is a Viral Structural Protein. *J Virol.* août 2006a;80(15):7287-94.

Huang C, Narayanan K, Ito N, Peters CJ, Makino S. Severe Acute Respiratory Syndrome Coronavirus 3a Protein Is Released in Membranous Structures from 3a Protein-Expressing Cells and Infected Cells. *J Virol.* janv 2006b;80(1):210-7.

Huang C, Peters CJ, Makino S. Severe Acute Respiratory Syndrome Coronavirus Accessory Protein 6 Is a Virion-Associated Protein and Is Released from 6 Protein-Expressing Cells. *J Virol.* 15 mai 2007;81(10):5423-6.

Huang HC, Lai YJ, Liao CC, Wang FY, Huang KB, Lee IJ, et al. Targeting conserved N-glycosylation blocks SARS-CoV-2 variant infection in vitro. *eBioMedicine.* déc 2021;74:103712.

Hulswit RJG, Lang Y, Bakkers MJG, Li W, Li Z, Schouten A, et al. Human coronaviruses OC43 and HKU1 bind to 9- *O* -acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proc Natl Acad Sci USA.* 12 févr 2019;116(7):2681-90.

Huong TN, Ravi Iyer L, Lui J, Wang DY, Tan BH, Sugrue RJ. The respiratory syncytial virus SH protein is incorporated into infectious virus particles that form on virus-infected cells. *Virology.* mars 2023;580:28-40.

Hurst KR, Koetzner CA, Masters PS. Characterization of a Critical Interaction between the Coronavirus Nucleocapsid Protein and Nonstructural Protein 3 of the Viral Replicase-Transcriptase Complex. *J Virol.* 15 août 2013;87(16):9159-72.

Irigoyen N, Firth AE, Jones JD, Chung BYW, Siddell SG, Brierley I. High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling. Frieman MB, éditeur. *PLoS Pathog.* 26 févr 2016;12(2):e1005473.

Ito N, Mossel EC, Narayanan K, Popov VL, Huang C, Inoue T, et al. Severe acute respiratory syndrome coronavirus 3a protein is a viral structural protein. *J Virol.* mars

2005;79(5):3182-6.

Jackson CB, Farzan M, Chen B, Choe H. Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol.* janv 2022;23(1):3-20.

Jacomy H, Fragoso G, Almazan G, Mushynski WE, Talbot PJ. Human coronavirus OC43 infection induces chronic encephalitis leading to disabilities in BALB/C mice. *Virology.* juin 2006;349(2):335-46.

Johnson BA, Xie X, Bailey AL, Kalveram B, Lokugamage KG, Muruato A, et al. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature.* 11 mars 2021;591(7849):293-9.

Juckel D, Dubuisson J, Belouzard S. Les coronavirus, ennemis incertains. *Med Sci (Paris).* juin 2020;36(6-7):633-41.

Jung K, Saif LJ. Porcine epidemic diarrhea virus infection: Etiology, epidemiology, pathogenesis and immunoprophylaxis. *The Veterinary Journal.* mai 2015;204(2):134-43.

Kamitani W, Narayanan K, Huang C, Lokugamage K, Ikegami T, Ito N, et al. Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. *Proc Natl Acad Sci USA.* 22 août 2006;103(34):12885-90.

Kanjanahaluethai A, Chen Z, Jukneliene D, Baker SC. Membrane topology of murine coronavirus replicase nonstructural protein 3. *Virology.* mai 2007;361(2):391-401.

Kanzawa N, Nishigaki K, Hayashi T, Ishii Y, Furukawa S, Niino A, et al. Augmentation of chemokine production by severe acute respiratory syndrome coronavirus 3a/X1 and 7a/X4 proteins through NF- κ B activation. *FEBS Letters.* 22 déc 2006;580(30):6807-12.

Kern DM, Sorum B, Mali SS, Hoel CM, Sridharan S, Remis JP, et al. Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. *Nat Struct Mol Biol.* juill 2021;28(7):573-82.

Kim SH, Chang SY, Sung M, Park JH, Bin Kim H, Lee H, et al. Extensive Viable Middle East Respiratory Syndrome (MERS) Coronavirus Contamination in Air and Surrounding Environment in MERS Isolation Wards. *Clin Infect Dis*. 1 août 2016;63(3):363-9.

Kimber MS. How to extend your (polylactosamine) antennae. *Journal of Biological Chemistry*. janv 2021;296:100212.

Kirchdoerfer RN, Cottrell CA, Wang N, Pallesen J, Yassine HM, Turner HL, et al. Pre-fusion structure of a human coronavirus spike protein. *Nature*. mars 2016;531(7592):118-21.

Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun*. 28 mai 2019;10(1):2342.

Klumperman J, Locker JK, Meijer A, Horzinek MC, Geuze HJ, Rottier PJ. Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding. *J Virol*. oct 1994;68(10):6523-34.

Knoops K, Kikkert M, Worm SHEVD, Zevenhoven-Dobbe JC, Van Der Meer Y, Koster AJ, et al. SARS-Coronavirus Replication Is Supported by a Reticulovesicular Network of Modified Endoplasmic Reticulum. Emerman M, éditeur. *PLoS Biol*. 16 sept 2008;6(9):e226.

Kobayashi Y, Suzuki Y. Evidence for N-Glycan Shielding of Antigenic Sites during Evolution of Human Influenza A Virus Hemagglutinin. *J Virol*. avr 2012;86(7):3446-51.

Koch G, Kant A. Binding of Antibodies that Strongly Neutralise Infectious Bronchitis Virus is Dependent on the Glycosylation of the Viral Peplomer Protein. In: Cavanagh D, Brown TDK, éditeurs. *Coronaviruses and their Diseases* [Internet]. Boston, MA: Springer US; 1990 [cité 30 juill 2023]. p. 143-50. (Advances in Experimental Medicine and Biology; vol. 276). Disponible sur: http://link.springer.com/10.1007/978-1-4684-5823-7_21

Koch T, Dahlke C, Fathi A, Kupke A, Krähling V, Okba NMA, et al. Safety and immunogenicity of a modified vaccinia virus Ankara vector vaccine candidate for Middle

East respiratory syndrome: an open-label, phase 1 trial. *The Lancet Infectious Diseases*. juill 2020;20(7):827-38.

Kopecky-Bromberg SA, Martínez-Sobrido L, Frieman M, Baric RA, Palese P. Severe Acute Respiratory Syndrome Coronavirus Open Reading Frame (ORF) 3b, ORF 6, and Nucleocapsid Proteins Function as Interferon Antagonists. *J Virol*. 15 janv 2007;81(2):548-57.

Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*. août 2020;182(4):812-827.e19.

Krokhin O, Li Y, Andonov A, Feldmann H, Flick R, Jones S, et al. Mass Spectrometric Characterization of Proteins from the SARS Virus. *Molecular & Cellular Proteomics*. mai 2003;2(5):346-56.

Kuo L, Masters PS. The Small Envelope Protein E Is Not Essential for Murine Coronavirus Replication. *J Virol*. 15 avr 2003;77(8):4597-608.

Lang J, Yang N, Deng J, Liu K, Yang P, Zhang G, et al. Inhibition of SARS Pseudovirus Cell Entry by Lactoferrin Binding to Heparan Sulfate Proteoglycans. Geraghty RJ, éditeur. *PLoS ONE*. 22 août 2011;6(8):e23710.

Lau SKP, Woo PCY, Yip CCY, Tse H, Tsoi H wah, Cheng VCC, et al. Coronavirus HKU1 and Other Coronavirus Infections in Hong Kong. *J Clin Microbiol*. juin 2006;44(6):2063-71.

Laude H, Gelfi J, Lavenant L, Charley B. Single amino acid changes in the viral glycoprotein M affect induction of alpha interferon by the coronavirus transmissible gastroenteritis virus. *J Virol*. févr 1992;66(2):743-9.

Lavie M, Hanouille X, Dubuisson J. Glycan Shielding and Modulation of Hepatitis C Virus Neutralizing Antibodies. *Front Immunol*. 27 avr 2018;9:910.

Law PTW, Wong CH, Au TCC, Chuck CP, Kong SK, Chan PKS, et al. The 3a protein of severe acute respiratory syndrome-associated coronavirus induces apoptosis in Vero E6 cells. *Journal of General Virology*. 1 juill 2005;86(7):1921-30.

Ledford H. Coronavirus breakthrough: dexamethasone is first drug shown to save lives. *Nature*. 25 juin 2020;582(7813):469-469.

Lednicky JA, Lauzardo M, Fan ZH, Jutla A, Tilly TB, Gangwar M, et al. Viable SARS-CoV-2 in the air of a hospital room with COVID-19 patients. *International Journal of Infectious Diseases*. nov 2020;100:476-82.

Lee PL, Kohler JJ, Pfeffer SR. Association of beta-1,3-N-acetylglucosaminyltransferase 1 and beta-1,4-galactosyltransferase 1, trans-Golgi enzymes involved in coupled poly-N-acetyllactosamine synthesis. *Glycobiology*. juin 2009;19(6):655-64.

Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell*. sept 2020;182(5):1284-1294.e9.

Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*. nov 2003;426(6965):450-4.

Liang JQ, Fang S, Yuan Q, Huang M, Chen RA, Fung TS, et al. N-Linked glycosylation of the membrane protein ectodomain regulates infectious bronchitis virus-induced ER stress response, apoptosis and pathogenesis. *Virology*. mai 2019;531:48-56.

Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*. févr 2020;395(10224):565-74.

Lu W, Zheng BJ, Xu K, Schwarz W, Du L, Wong CKL, et al. Severe acute respiratory syndrome-associated coronavirus 3a protein forms an ion channel and modulates virus release. *Proc Natl Acad Sci USA*. 15 août 2006;103(33):12540-5.

Lui PY, Wong LYR, Fung CL, Siu KL, Yeung ML, Yuen KS, et al. Middle East respiratory syndrome coronavirus M protein suppresses type I interferon expression through the inhibition of TBK1-dependent phosphorylation of IRF3. *Emerging Microbes & Infections*. janv 2016;5(1):1-9.

Luo H, Wu D, Shen C, Chen K, Shen X, Jiang H. Severe acute respiratory syndrome coronavirus membrane protein interacts with nucleocapsid protein mostly through their carboxyl termini by electrostatic attraction. *The International Journal of Biochemistry & Cell Biology*. 2006;38(4):589-99.

Ma C, Li Y, Wang L, Zhao G, Tao X, Tseng CTK, et al. Intranasal vaccination with recombinant receptor-binding domain of MERS-CoV spike protein induces much stronger local mucosal immune responses than subcutaneous immunization: Implication for designing novel mucosal MERS vaccines. *Vaccine*. avr 2014;32(18):2100-8.

Machamer CE, Mentone, Sue Ann, Rose, John K. JK, Farquhar, Marilyn G. The E1 glycoprotein of an avian coronavirus is targeted to the cis Golgi complex. *Cell Biology*. sept 1990;87(18):6944-8.

Madu IG, Roth SL, Belouzard S, Whittaker GR. Characterization of a Highly Conserved Domain within the Severe Acute Respiratory Syndrome Coronavirus Spike Protein S2 Domain with Characteristics of a Viral Fusion Peptide. *J Virol*. août 2009;83(15):7411-21.

Mardones G, González A. Selective plasma membrane permeabilization by freeze-thawing and immunofluorescence epitope access to determine the topology of intracellular membrane proteins. *Journal of Immunological Methods*. avr 2003;275(1-2):169-77.

Masters PS. *The Molecular Biology of Coronaviruses*. In: *Advances in Virus Research* [Internet]. Elsevier; 2006 [cité 10 nov 2021]. p. 193-292. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S0065352706660053>

McBride CE, Li J, Machamer CE. The Cytoplasmic Tail of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein Contains a Novel Endoplasmic Reticulum Retrieval

Signal That Binds COPI and Promotes Interaction with Membrane Protein. *J Virol.* mars 2007;81(5):2418-28.

McIntosh K, Dees JH, Becker WB, Kapikian AZ, Chanock RM. Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc Natl Acad Sci USA.* avr 1967;57(4):933-40.

Meinhardt J, Radke J, Dittmayer C, Franz J, Thomas C, Mothes R, et al. Olfactory transmucosal SARS-CoV-2 invasion as a port of central nervous system entry in individuals with COVID-19. *Nat Neurosci.* févr 2021;24(2):168-75.

de Melo GD, Lazarini F, Levallois S, Hautefort C, Michel V, Larrous F, et al. COVID-19-related anosmia is associated with viral persistence and inflammation in human olfactory epithelium and brain infection in hamsters. *Sci Transl Med.* 2 juin 2021;13(596):eabf8396.

Menachery VD, Mitchell HD, Cockrell AS, Gralinski LE, Yount BL, Graham RL, et al. MERS-CoV Accessory ORFs Play Key Role for Infection and Pathogenesis. *Racaniello VR, éditeur. mBio [Internet].* 6 sept 2017 [cité 10 nov 2021];8(4). Disponible sur: <https://journals.asm.org/doi/10.1128/mBio.00665-17>

Menni C, Valdes AM, Polidori L, Antonelli M, Penamakuri S, Nogal A, et al. Symptom prevalence, duration, and risk of hospital admission in individuals infected with SARS-CoV-2 during periods of omicron and delta variant dominance: a prospective observational study from the ZOE COVID Study. *The Lancet.* avr 2022;399(10335):1618-24.

Mesalam AA, Desombere I, Farhoudi A, Van Houtte F, Verhoye L, Ball J, et al. Development and characterization of a human monoclonal antibody targeting the N-terminal region of hepatitis C virus envelope glycoprotein E1. *Virology.* janv 2018;514:30-41.

Meunier JC, Fournillier A, Choukhi A, Cahour A, Cocquerel L, Dubuisson J, et al. Analysis of the glycosylation sites of hepatitis C virus (HCV) glycoprotein E1 and the influence of E1

glycans on the formation of the HCV glycoprotein complex. *Journal of General Virology*. 1 avr 1999;80(4):887-96.

Miao G, Zhao H, Li Y, Ji M, Chen Y, Shi Y, et al. ORF3a of the COVID-19 virus SARS-CoV-2 blocks HOPS complex-mediated assembly of the SNARE complex required for autolysosome formation. *Developmental Cell*. févr 2021;56(4):427-442.e5.

Milewska A, Zarebski M, Nowak P, Stozek K, Potempa J, Pyrc K. Human Coronavirus NL63 Utilizes Heparan Sulfate Proteoglycans for Attachment to Target Cells. Perlman S, éditeur. *J Virol*. 15 nov 2014;88(22):13221-30.

Miller AN, Houlihan PR, Matamala E, Cabezas-Bratesco D, Lee GY, Cristofori-Armstrong B, et al. The SARS-CoV-2 accessory protein Orf3a is not an ion channel, but does interact with trafficking proteins. *eLife*. 25 janv 2023;12:e84477.

Miorin L, Kehrer T, Sanchez-Aparicio MT, Zhang K, Cohen P, Patel RS, et al. SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *Proc Natl Acad Sci USA*. 10 nov 2020;117(45):28344-54.

Miura TA, Travanty EA, Oko L, Bielefeldt-Ohmann H, Weiss SR, Beauchemin N, et al. The Spike Glycoprotein of Murine Coronavirus MHV-JHM Mediates Receptor-Independent Infection and Spread in the Central Nervous Systems of *Ceacam1a*^{-/-} Mice. *J Virol*. 15 janv 2008;82(2):755-63.

Modjarrad K, Roberts CC, Mills KT, Castellano AR, Paolino K, Muthumani K, et al. Safety and immunogenicity of an anti-Middle East respiratory syndrome coronavirus DNA vaccine: a phase 1, open-label, single-arm, dose-escalation trial. *The Lancet Infectious Diseases*. sept 2019;19(9):1013-22.

Moës E, Vijgen L, Keyaerts E, Zlateva K, Li S, Maes P, et al. A novel pancoronavirus RT-PCR assay: frequent detection of human coronavirus NL63 in children hospitalized with respiratory tract infections in Belgium. *BMC Infect Dis*. déc 2005;5(1):6.

Molloy SS, Thomas L, VanSlyke JK, Stenberg PE, Thomas G. Intracellular trafficking and activation of the furin proprotein convertase: localization to the TGN and recycling from the cell surface. *The EMBO Journal*. janv 1994;13(1):18-33.

Motozono C, Toyoda M, Zahradnik J, Saito A, Nasser H, Tan TS, et al. SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell Host & Microbe*. juill 2021;29(7):1124-1136.e11.

Muthumani K, Falzarano D, Reuschel EL, Tingey C, Flingai S, Villarreal DO, et al. A synthetic consensus anti-spike protein DNA vaccine induces protective immunity against Middle East respiratory syndrome coronavirus in nonhuman primates. *Sci Transl Med [Internet]*. 19 août 2015 [cité 25 avr 2023];7(301). Disponible sur: <https://www.science.org/doi/10.1126/scitranslmed.aac7462>

Nabi IR, Dennis JW. The extent of poly(lactosamine) glycosylation of MDCK LAMP-2 is determined by its Golgi residence time. *Glycobiology*. 1 sept 1998;8(9):947-53.

Naskalska A, Dabrowska A, Szczepanski A, Milewska A, Jasik KP, Pyrc K. Membrane Protein of Human Coronavirus NL63 Is Responsible for Interaction with the Adhesion Receptor. Gallagher T, éditeur. *J Virol [Internet]*. oct 2019 [cité 10 nov 2021];93(19). Disponible sur: <https://journals.asm.org/doi/10.1128/JVI.00355-19>

Navas S, Seo SH, Chua MM, Sarma JD, Lavi E, Hingley ST, et al. Murine Coronavirus Spike Protein Determines the Ability of the Virus To Replicate in the Liver and Cause Hepatitis. *J Virol*. mars 2001;75(5):2452-7.

Navas S, Weiss SR. Murine Coronavirus-Induced Hepatitis: JHM Genetic Background Eliminates A59 Spike-Determined Hepatotropism. *J Virol*. 15 avr 2003;77(8):4972-8.

Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, et al. A structural analysis of M protein in coronavirus assembly and morphology. *Journal of Structural Biology*. avr 2011;174(1):11-22.

Nilsson I, Von Heijne G. Fine-tuning the topology of a polytopic membrane protein: Role of positively and negatively charged amino acids. *Cell*. sept 1990;62(6):1135-41.

Nilsson T, Lucocq JM, Mackay D, Warren G. The membrane spanning domain of beta-1,4-galactosyltransferase specifies trans Golgi localization. *The EMBO Journal*. déc 1991;10(12):3567-75.

Nishimura N, Balch WE. A Di-Acidic Signal Required for Selective Export from the Endoplasmic Reticulum. *Science*. 25 juill 1997;277(5325):556-8.

Oguri S. Analysis of sugar chain-binding specificity of tomato lectin using lectin blot: recognition of high mannose-type N-glycans produced by plants and yeast. *Glycoconj J*. nov 2005;22(7-9):453-61.

Oostra M, de Haan CAM, de Groot RJ, Rottier PJM. Glycosylation of the Severe Acute Respiratory Syndrome Coronavirus Triple-Spanning Membrane Proteins 3a and M. *J Virol*. mars 2006;80(5):2326-36.

Ortego J, Ceriani JE, Patiño C, Plana J, Enjuanes L. Absence of E protein arrests transmissible gastroenteritis coronavirus maturation in the secretory pathway. *Virology*. nov 2007;368(2):296-308.

Papa A, Papadimitriou E, Luna LK de S, Al Masri M, Souliou E, Eboriadou M, et al. Coronaviruses in Children, Greece. *Emerg Infect Dis*. juin 2007;13(6):947-9.

Pardo J, Shukla AM, Chamarthi G, Gupte A. The journey of remdesivir: from Ebola to COVID-19. *DIC*. 22 mai 2020;9:1-9.

Parsons LM, Bouwman KM, Azurmendi H, De Vries RP, Cipollo JF, Verheije MH. Glycosylation of the viral attachment protein of avian coronavirus is essential for host cell and receptor binding. *Journal of Biological Chemistry*. mai 2019;294(19):7797-809.

Peiris J, Chu C, Cheng V, Chan K, Hung I, Poon L, et al. Clinical progression and viral load in

a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *The Lancet*. mai 2003;361(9371):1767-72.

Pepe A, Pietropaoli S, Vos M, Barba-Spaeth G, Zurzolo C. Tunneling nanotubes provide a route for SARS-CoV-2 spreading. *Sci Adv*. 22 juill 2022;8(29):eabo0171.

Perrier A, Bonnin A, Desmarets L, Danneels A, Goffard A, Rouillé Y, et al. The C-terminal domain of the MERS coronavirus M protein contains a trans-Golgi network localization signal. *Journal of Biological Chemistry*. sept 2019;294(39):14406-21.

Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med*. 31 déc 2020;383(27):2603-15.

Prabakaran P, Gan J, Feng Y, Zhu Z, Choudhry V, Xiao X, et al. Structure of Severe Acute Respiratory Syndrome Coronavirus Receptor-binding Domain Complexed with Neutralizing Antibody. *Journal of Biological Chemistry*. juin 2006;281(23):15829-36.

Qin L, Xiong B, Luo C, Guo ZM, Hao P, Su J, et al. Identification of probable genomic packaging signal sequence from SARS-CoV genome by bioinformatics analysis. *Acta Pharmacol Sin*. juin 2003;24(6):489-96.

Raamsman MJB, Locker JK, De Hooge A, De Vries AAF, Griffiths G, Vennema H, et al. Characterization of the Coronavirus Mouse Hepatitis Virus Strain A59 Small Membrane Protein E. *J Virol*. mars 2000;74(5):2333-42.

Raj VS, Mou H, Smits SL, Dekkers DHW, Müller MA, Dijkman R, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature*. mars 2013;495(7440):251-4.

Ramírez AS, Kowal J, Locher KP. Cryo-electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B. *Science*. 13 déc 2019;366(6471):1372-5.

Redondo N, Zaldívar-López S, Garrido JJ, Montoya M. SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. *Front Immunol.* 7 juill 2021;12:708264.

Regan AD, Whittaker GR. Utilization of DC-SIGN for Entry of Feline Coronaviruses into Host Cells. *J Virol.* déc 2008;82(23):11992-6.

Reusken CB, Haagmans BL, Müller MA, Gutierrez C, Godeke GJ, Meyer B, et al. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *The Lancet Infectious Diseases.* oct 2013;13(10):859-66.

Risco C, Antón IM, Suñé C, Pedregosa AM, Martín-Alonso JM, Parra F, et al. Membrane protein molecules of transmissible gastroenteritis coronavirus also expose the carboxy-terminal region on the external surface of the virion. *J Virol.* sept 1995;69(9):5269-77.

Ruch TR, Machamer CE. The Hydrophobic Domain of Infectious Bronchitis Virus E Protein Alters the Host Secretory Pathway and Is Important for Release of Infectious Virus. *J Virol.* 15 janv 2011;85(2):675-85.

Ruiz-Canada C, Kelleher DJ, Gilmore R. Cotranslational and Posttranslational N-Glycosylation of Polypeptides by Distinct Mammalian OST Isoforms. *Cell.* janv 2009;136(2):272-83.

Sabir JSM, Lam TTY, Ahmed MMM, Li L, Shen Y, E. M. Abo-Aba S, et al. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science.* janv 2016;351(6268):81-4.

Sakai Y, Kawachi K, Terada Y, Omori H, Matsuura Y, Kamitani W. Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. *Virology.* oct 2017;510:165-74.

Sanda M, Morrison L, Goldman R. N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Anal Chem.* 2 févr 2021;93(4):2003-9.

Schaefer SR, Mackenzie JM, Pekosz A. The ORF7b Protein of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) Is Expressed in Virus-Infected Cells and Incorporated into SARS-CoV Particles. *J Virol*. 15 janv 2007;81(2):718-31.

Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature*. 14 mai 2020;581(7807):221-4.

Shang J, Zheng Y, Yang Y, Liu C, Geng Q, Luo C, et al. Cryo-EM structure of infectious bronchitis coronavirus spike protein reveals structural and functional evolution of coronavirus spike proteins. Gallagher T, éditeur. *PLoS Pathog*. 23 avr 2018;14(4):e1007009.

Shao X, Guo X, Esper F, Weibel C, Kahn JS. Seroepidemiology of group I human coronaviruses in children. *Journal of Clinical Virology*. nov 2007;40(3):207-13.

Sheahan TP, Sims AC, Graham RL, Menachery VD, Gralinski LE, Case JB, et al. Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Sci Transl Med*. 28 juin 2017;9(396):eaal3653.

Shen S, Lin PS, Chao YC, Zhang A, Yang X, Lim SG, et al. The severe acute respiratory syndrome coronavirus 3a is a novel structural protein. *Biochemical and Biophysical Research Communications*. avr 2005;330(1):286-92.

Shi CS, Qi HY, Boularan C, Huang NN, Abu-Asab M, Shelhamer JH, et al. SARS-Coronavirus Open Reading Frame-9b Suppresses Innate Immunity by Targeting Mitochondria and the MAVS/TRAF3/TRAF6 Signalosome. *The Journal of Immunology*. 15 sept 2014;193(6):3080-9.

Shim BS, Stadler K, Nguyen HH, Yun CH, Kim DW, Chang J, et al. Sublingual immunization with recombinant adenovirus encoding SARS-CoV spike protein induces systemic and mucosal immunity without redirection of the virus to the brain. *Virol J*. déc 2012;9(1):215.

Shirato K, Kawase M, Matsuyama S. Middle East respiratory syndrome coronavirus

infection mediated by the transmembrane serine protease TMPRSS2. *J Virol.* déc 2013;87(23):12552-61.

Shrimal S, Cherepanova NA, Gilmore R. Cotranslational and posttranslocational N-glycosylation of proteins in the endoplasmic reticulum. *Seminars in Cell & Developmental Biology.* mai 2015;41:71-8.

Shrimal S, Cherepanova NA, Gilmore R. DC2 and KCP2 mediate the interaction between the oligosaccharyltransferase and the ER translocon. *Journal of Cell Biology.* 6 nov 2017;216(11):3625-38.

Simmons G, Gosalia DN, Rennekamp AJ, Reeves JD, Diamond SL, Bates P. Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci USA.* 16 août 2005;102(33):11876-81.

Sims AC, Baric RS, Yount B, Burkett SE, Collins PL, Pickles RJ. Severe Acute Respiratory Syndrome Coronavirus Infection of Human Ciliated Airway Epithelia: Role of Ciliated Cells in Viral Spread in the Conducting Airways of the Lungs. *J Virol.* 15 déc 2005;79(24):15511-24.

Siu KL, Chan CP, Kok KH, Chiu-Yat Woo P, Jin DY. Suppression of innate antiviral response by severe acute respiratory syndrome coronavirus M protein is mediated through the first transmembrane domain. *Cell Mol Immunol.* mars 2014;11(2):141-9.

Siu KL, Kok KH, Ng MHJ, Poon VKM, Yuen KY, Zheng BJ, et al. Severe Acute Respiratory Syndrome Coronavirus M Protein Inhibits Type I Interferon Production by Impeding the Formation of TRAF3·TANK·TBK1/IKK ϵ Complex. *Journal of Biological Chemistry.* juin 2009;284(24):16202-9.

Siu YL, Teoh KT, Lo J, Chan CM, Kien F, Escriou N, et al. The M, E, and N Structural Proteins of the Severe Acute Respiratory Syndrome Coronavirus Are Required for Efficient Assembly, Trafficking, and Release of Virus-Like Particles. *J Virol.* 15 nov 2008;82(22):11318-30.

Sloots T, McErlean P, Speicher D, Arden K, Nissen M, Mackay I. Evidence of human coronavirus HKU1 and human bocavirus in Australian children. *Journal of Clinical Virology*. janv 2006;35(1):99-102.

Song HC, Seo MY, Stadler K, Yoo BJ, Choo QL, Coates SR, et al. Synthesis and Characterization of a Native, Oligomeric Form of Recombinant Severe Acute Respiratory Syndrome Coronavirus Spike Glycoprotein. *J Virol*. oct 2004;78(19):10328-35.

Stertz S, Reichelt M, Spiegel M, Kuri T, Martínez-Sobrido L, García-Sastre A, et al. The intracellular sites of early replication and budding of SARS-coronavirus. *Virology*. mai 2007;361(2):304-15.

Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, et al. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends in Microbiology*. juin 2016;24(6):490-502.

Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E, et al. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc Natl Acad Sci USA* [Internet]. 16 sept 2014 [cité 5 mai 2023];111(37). Disponible sur: <https://pnas.org/doi/full/10.1073/pnas.1323705111>

Talbot HKB, Crowe JE, Edwards KM, Griffin MR, Zhu Y, Weinberg GA, et al. Coronavirus infection and hospitalizations for acute respiratory illness in young children. *J Med Virol*. mai 2009;81(5):853-6.

Tan YJ, Fielding BC, Goh PY, Shen S, Tan THP, Lim SG, et al. Overexpression of 7a, a Protein Specifically Encoded by the Severe Acute Respiratory Syndrome Coronavirus, Induces Apoptosis via a Caspase-Dependent Pathway. *J Virol*. 15 déc 2004;78(24):14043-7.

Thorp EB, Boscarino JA, Logan HL, Goletz JT, Gallagher TM. Palmitoylations on Murine Coronavirus Spike Proteins Are Essential for Virion Assembly and Infectivity. *J Virol*. févr 2006;80(3):1280-9.

Togayachi A, Kozono Y, Ishida H, Abe S, Suzuki N, Tsunoda Y, et al. Polylysosamine on glycoproteins influences basal levels of lymphocyte and macrophage activation. *Proceedings of the National Academy of Sciences*. 2 oct 2007;104(40):15829-34.

Togayachi A, Kozono Y, Kuno A, Ohkura T, Sato T, Hirabayashi J, et al. β 3GnT2 (B3GNT2), a Major Polylysosamine Synthase: Analysis of B3gnt2-Deficient Mice. In: *Methods in Enzymology* [Internet]. Elsevier; 2010 [cité 10 nov 2021]. p. 185-204. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/S007668791079011X>

Tu C, Cramer G, Kong X, Chen J, Sun Y, Yu M, et al. Antibodies to SARS Coronavirus in Civets. *Emerg Infect Dis*. déc 2004;10(12):2244-8.

Tyrrell DAJ, Cohen S, Schilarb JE. Signs and symptoms in common colds. *Epidemiol Infect*. août 1993;111(1):143-56.

Vabret A, Dina J, Gouarin S, Petitjean J, Corbet S, Freymuth F. Detection of the New Human Coronavirus HKU1: A Report of 6 Cases. *Clinical Infectious Diseases*. 1 mars 2006;42(5):634-9.

Vabret A, Mourez T, Dina J, van der Hoek L, Gouarin S, Petitjean J, et al. Human Coronavirus NL63, France. *Emerg Infect Dis*. août 2005;11(8):1225-9.

Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., éditeurs. *Essentials of Glycobiology* [Internet]. 4th éd. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022 [cité 26 août 2023]. Disponible sur: <http://www.ncbi.nlm.nih.gov/books/NBK579918/>

te Velthuis AJW, van den Worm SHE, Snijder EJ. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic Acids Research*. févr 2012;40(4):1737-47.

Venkatagopalan P, Daskalova SM, Lopez LA, Dolezal KA, Hogue BG. Coronavirus envelope (E) protein remains at the site of assembly. *Virology*. avr 2015;478:75-85.

Vennema H, Godeke GJ, Rossen JW, Voorhout WF, Horzinek MC, Opstelten DJ, et al. Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *The EMBO Journal*. avr 1996;15(8):2020-8.

Voß D, Kern A, Traggiai E, Eickmann M, Stadler K, Lanzavecchia A, et al. Characterization of severe acute respiratory syndrome coronavirus membrane protein. *FEBS Letters*. 6 févr 2006;580(3):968-73.

Voß D, Pfefferle S, Drosten C, Stevermann L, Traggiai E, Lanzavecchia A, et al. Studies on membrane topology, N-glycosylation and functionality of SARS-CoV membrane protein. *Virology*. déc 2009;6(1):79.

Walls AC, Tortorici MA, Bosch BJ, Frenz B, Rottier PJM, DiMaio F, et al. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature*. 3 mars 2016a;531(7592):114-7.

Walls AC, Tortorici MA, Frenz B, Snijder J, Li W, Rey FA, et al. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat Struct Mol Biol*. oct 2016b;23(10):899-905.

Wang H, Paulson KR, Pease SA, Watson S, Comfort H, Zheng P, et al. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *The Lancet*. avr 2022;399(10334):1513-36.

Wang K, Lu W, Chen J, Xie S, Shi H, Hsu H, et al. PEDV ORF3 encodes an ion channel protein and regulates virus production. *FEBS Lett*. 17 févr 2012;586(4):384-91.

Wang N, Shi X, Jiang L, Zhang S, Wang D, Tong P, et al. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res*. août 2013;23(8):986-93.

Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet*. mai

2020;395(10236):1569-78.

Ward JM. Morphogenesis of a virus in cats with experimental feline infectious peritonitis. *Virology*. mai 1970;41(1):191-4.

Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*. 17 juill 2020a;369(6501):330-3.

Watanabe Y, Berndsen ZT, Raghwani J, Seabright GE, Allen JD, Pybus OG, et al. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat Commun*. 27 mai 2020b;11(1):2688.

Westerbeck JW, Machamer CE. A Coronavirus E Protein Is Present in Two Distinct Pools with Different Effects on Assembly and the Secretory Pathway. Perlman S, éditeur. *J Virol*. 15 sept 2015;89(18):9313-23.

Williams MA, Lamb RA. Polylactosaminoglycan Modification of a Small Integral Membrane Glycoprotein, Influenza B Virus NB. *MOL CELL BIOL*. 1988;8.

Williamson BN, Feldmann F, Schwarz B, Meade-White K, Porter DP, Schulz J, et al. Clinical benefit of remdesivir in rhesus macaques infected with SARS-CoV-2. *Nature*. 10 sept 2020a;585(7824):273-6.

Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*. 20 août 2020b;584(7821):430-6.

de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol*. août 2016;14(8):523-34.

Wolff G, Limpens RWAL, Zevenhoven-Dobbe JC, Laugks U, Zheng S, de Jong AWM, et al. A molecular pore spans the double membrane of the coronavirus replication organelle. *Science*. 11 sept 2020;369(6509):1395-8.

Woo PCY, Lau SKP, Chu C ming, Chan K hung, Tsoi H wah, Huang Y, et al. Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *J Virol.* 15 janv 2005;79(2):884-95.

Worobey M, Levy JI, Malpica Serrano L, Crits-Christoph A, Pekar JE, Goldstein SA, et al. The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science.* 26 août 2022;377(6609):951-9.

Xiao F, Sun J, Xu Y, Li F, Huang X, Li H, et al. Infectious SARS-CoV-2 in Feces of Patient with Severe COVID-19. *Emerg Infect Dis.* août 2020;26(8):1920-2.

Yadav R, Chaudhary JK, Jain N, Chaudhary PK, Khanra S, Dhamija P, et al. Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. *Cells.* 6 avr 2021;10(4):821.

Yamashita K, Totani K, Ohkura T, Takasaki S, Goldstein IJ, Kobata A. Carbohydrate binding properties of complex-type oligosaccharides on immobilized *Datura stramonium* lectin. *Journal of Biological Chemistry.* févr 1987;262(4):1602-7.

Yang Y, Zhang L, Geng H, Deng Y, Huang B, Guo Y, et al. The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists. *Protein Cell.* déc 2013;4(12):951-61.

Ye Y, Hogue BG. Role of the Coronavirus E Viroporin Protein Transmembrane Domain in Virus Assembly. *J Virol.* avr 2007;81(7):3597-607.

Yount B, Roberts RS, Sims AC, Deming D, Frieman MB, Sparks J, et al. Severe Acute Respiratory Syndrome Coronavirus Group-Specific Open Reading Frames Encode Nonessential Functions for Replication in Cell Cultures and Mice. *J Virol.* 15 déc 2005;79(23):14909-22.

Yu ITS, Li Y, Wong TW, Tam W, Chan AT, Lee JHW, et al. Evidence of Airborne Transmission of the Severe Acute Respiratory Syndrome Virus. *N Engl J Med.* 22 avr

2004;350(17):1731-9.

Yuan X, Wu J, Shan Y, Yao Z, Dong B, Chen B, et al. SARS coronavirus 7a protein blocks cell cycle progression at G0/G1 phase via the cyclin D3/pRb pathway. *Virology*. mars 2006;346(1):74-85.

Yuan X, Yao Z, Wu J, Zhou Y, Shan Y, Dong B, et al. G1 Phase Cell Cycle Arrest Induced by SARS-CoV 3a Protein via the Cyclin D3/pRb Pathway. *Am J Respir Cell Mol Biol*. juill 2007;37(1):9-19.

Yuan Y, Cao D, Zhang Y, Ma J, Qi J, Wang Q, et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat Commun*. 10 avr 2017;8(1):15092.

Yuan Z, Hu B, Xiao H, Tan X, Li Y, Tang K, et al. The E3 Ubiquitin Ligase RNF5 Facilitates SARS-CoV-2 Membrane Protein-Mediated Virion Release. Goff SP, éditeur. *mBio*. 22 févr 2022;13(1):e03168-21.

Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N Engl J Med*. 8 nov 2012;367(19):1814-20.

Zhang J, Xiao T, Cai Y, Chen B. Structure of SARS-CoV-2 spike protein. *Current Opinion in Virology*. oct 2021;50:173-82.

Zhang R, Wang K, Lv W, Yu W, Xie S, Xu K, et al. The ORF4a protein of human coronavirus 229E functions as a viroporin that regulates viral production. *Biochim Biophys Acta*. avr 2014;1838(4):1088-95.

Zhang Y, Chen C, Zhu S, Shu C, Wang D, Song J, et al. Isolation of 2019-nCoV from a Stool Specimen of a Laboratory-Confirmed Case of the Coronavirus Disease 2019 (COVID-19). *China CDC Weekly*. 2020;2(8):123-4.

Zhang Z, Nomura N, Muramoto Y, Ekimoto T, Uemura T, Liu K, et al. Structure of SARS-CoV-2 membrane protein essential for virus assembly. *Nat Commun.* 5 août 2022;13(1):4399.

Zhao MM, Yang WL, Yang FY, Zhang L, Huang WJ, Hou W, et al. Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. *Sig Transduct Target Ther.* 27 mars 2021;6(1):134.

Zheng J, Yamada Y, Fung TS, Huang M, Chia R, Liu DX. Identification of N-linked glycosylation sites in the spike protein and their functional impact on the replication and infectivity of coronavirus infectious bronchitis virus in cell culture. *Virology.* janv 2018;513:65-74.

Zheng Y, Zhuang MW, Han L, Zhang J, Nan ML, Zhan P, et al. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) membrane (M) protein inhibits type I and III interferon production by targeting RIG-I/MDA-5 signaling. *Sig Transduct Target Ther.* déc 2020;5(1):299.

Zhong N, Zheng B, Li Y, Poon L, Xie Z, Chan K, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *The Lancet.* oct 2003;362(9393):1353-8.

Zhou D. Why are Glycoproteins Modified by Poly-N-Acetyllactosamine Glycoconjugates? *CPPS.* 1 févr 2003;4(1):1-9.

Zhou Y, Lu K, Pfefferle S, Bertram S, Glowacka I, Drosten C, et al. A Single Asparagine-Linked Glycosylation Site of the Severe Acute Respiratory Syndrome Coronavirus Spike Glycoprotein Facilitates Inhibition by Mannose-Binding Lectin through Multiple Mechanisms. *J Virol.* sept 2010;84(17):8753-64.

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med.* 20 févr 2020;382(8):727-33.

Zumla A, Chan JFW, Azhar EI, Hui DSC, Yuen KY. Coronaviruses — drug discovery and therapeutic options. *Nat Rev Drug Discov.* mai 2016;15(5):327-47.

Züst R, Cervantes-Barragan L, Habjan M, Maier R, Neuman BW, Ziebuhr J, et al. Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat Immunol.* févr 2011;12(2):137-43.

ANNEXES

1 **MERS-CoV and SARS-CoV-2 membrane proteins are** 2 **modified with polylactosamine chains.**

4 **1.1 Author names**

5 **Dylan Juckel¹, Lowiese Desmarets^{1,2}, Adeline Danneels^{1,2}, Yves Rouillé¹, Jean Dubuisson¹**
6 **and Sandrine Belouzard*¹**

8 **1.2 Affiliation(s)**

9 ¹Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 9017 -
10 CILL- Center for Infection and Immunity of Lille, F-59000 Lille, France

11 ² Equally contributed to the work

14 **1.3 Corresponding author and email address**

15 *To whom correspondence should be addressed : sandrine.belouzard@ibl.cnrs.fr

17 **1.4 Keywords**

18 N-linked glycosylation, polylactosamines, coronavirus, membrane protein

21 **Abstract**

22 Coronaviruses are positive-stranded RNA enveloped viruses. The helical nucleocapsid is
23 surrounded by a lipid bilayer in which are anchored three viral proteins: the spike (S), the
24 membrane (M) and the envelope (E) proteins. The membrane protein is the major component
25 of the viral envelope and is believed to be its building block. The membrane protein of MERS-
26 CoV and SARS-CoV-2 contains a short N-terminal domain with a N-glycosylation site. We
27 investigated their N-glycosylation and show that polylactosamine chains are conjugated to
28 SARS-CoV-2 and MERS-CoV M proteins in transfected and infected cells. Acidic residues
29 present in the first transmembrane segments of the proteins are required for their
30 glycosylation. No specific signal to specify polylactosamine conjugation could be identified and
31 high mannose-conjugated protein were incorporated into virus-like particles.

32

33 **Introduction**

34 Coronaviruses are positive-stranded RNA enveloped viruses. The helical ribonucleocapsid is
35 surrounded by a lipid bilayer in which are anchored three viral proteins: the spike (S), the
36 membrane (M) and the envelope (E) proteins. The membrane protein is the major component
37 of the viral envelope and is believed to be its building block. Indeed, the M protein has been
38 shown to be able to undertake molecular interactions with all the other structural proteins of
39 the viral envelope and has an essential role in virus-like particle (VLP) formation [1]. The M
40 protein ranges from 219 to 262 residues and is composed of a small N-terminal glycosylated
41 domain, three transmembrane segments and a long C-terminal domain. The N-terminal
42 domain of the protein is exposed at the surface of the virion and the C-terminal domain inside
43 the virion (N-exo/C-endo topology). However a dual topology was demonstrated for the M
44 protein of transmissible gastroenteritis virus (TGEV) with roughly one-third of the protein
45 having its C-terminal domain exposed at the virion surface (N-exo/C-exo) suggesting the
46 presence of a fourth transmembrane segment [2]. TGEV and feline coronavirus (FCoV) M
47 proteins have a signal peptide sequence. The structure of the SARS-CoV-2 M protein was
48 recently solved by cryo-EM [3, 4]. M protein forms homodimers and is structurally related to
49 the ORF3a protein suggesting a shared ancestral origin. Depending on the structure, the N-
50 terminal domain is composed of 9 or 16 residues and is followed by three transmembrane
51 helices (amino acids 10/17-36, 43-71, and 79-105) spaced by short linkers. The C-terminal
52 domain is highly structured and rich in beta-strands.

53 The short N-terminal domain of M proteins can be either N- or O-glycosylated depending on
54 the virus. Mouse hepatitis virus (MHV) and other betacoronaviruses such as HCoV-OC43 M
55 proteins are O-glycosylated whereas M proteins of MERS-CoV, SARS-CoV or SARS-CoV-2 are N-
56 glycosylated. It is well established that M protein glycosylation is dispensable for virus
57 assembly but a role in viral pathogenesis has been suggested. Notably, the glycosylation of M
58 proteins of MHV and TGEV has been linked to its interferogenic capacity and for MHV-A59,
59 abolition of M glycosylation decreased the replication of the virus in the liver whereas the
60 replication in the brain was not affected [5, 6]. Furthermore, the glycosylation of infectious
61 bronchitis virus (IBV) M protein induced ER stress and apoptosis [7].

62 In this study, we investigated the N-glycosylation of M proteins of the highly pathogenic
63 viruses MERS-CoV and SARS-CoV-2. Our results demonstrated that M proteins are modified
64 by polylysosamine chain in transfected and infected cells. Moreover, we found that not only
65 the N-linked glycosylation site but also acidic residues in the first transmembrane segment are
66 important for the glycosylation of the protein.

67

68 **Methods**

69 **Cell culture**

70 Huh-7 cells were grown at 37°C with 5% CO₂ in Dulbecco's Modified Eagle medium (DMEM;
71 Life Technologies) supplemented with 10% fetal bovine serum and 2 mM glutamax (Life
72 Technologies).

73

74 **Plasmids**

75 cDNA was obtained from total RNA extracted from SARS-CoV-2 (strain Lille
76 hCoV-19/France/HDF-IPL/2020, genbank MW575140) infected cells. Then the coding
77 sequence corresponding to the M protein gene was amplified with the forward primer : 5'-
78 CGgatccgcagattccaacggtactatta-3' and the following reverse primer: 5'-

79 CAgaattcttactgtacaagcaagcaata-3'. The PCR product was then inserted between BamHI and
80 EcoRI restriction sites of the pcDNA3.1 vector containing a coding sequence for the V5 tag.
81 Mutations were introduced by PCR-directed mutagenesis. For pcDNA3.1-V5-N5Q-M, the
82 coding sequence of the M protein was amplified by using the following primers: Fw-1 5'-
83 CGgatccgcagattccCAAggtactattaccgtt-3' and Rv-1 5'-CAgaattcttactgtacaagcaagcaata-3'.
84 For mutation of 3 residues of the N-terminus into alanine, the coding sequence were amplified
85 in two fragments to insert the mutations. Then, for SARS-CoV-2 M mutants, the two fragments
86 were fused by using the same Fw-2 (5'-ATTGACGTCAATGGGAGTTT-3') and Rv-2 (5'-
87 AGCGGCAGCagtaccgttgaatctgCGGatccCG-3') primers. The following primers were used for the
88 first round of PCR for SARS-CoV-2₈ITV₁₀-M: Fw-2 with Rv (5'-
89 AGCGGCAGCagtaccgttgaatctgCGGatccCG-3') and Rv-2 with (5'-
90 ctGCTGCCGTgaagagcttaaaaagctcc-3'), for₁₁EEL₁₃-M: Fw-2 with Rv (5'-
91 AGCCGCTGCaacggttaagtaccgttgg-3') and Rv-2 with Fw (5'-
92 tGCAGCGCTaaaaagctcctgaacaatgg -3'), for₁₄KKL₁₆-M: Fw-2 with Rv (5'-
93 AGCCGCTGCaagctctcaacggttaagtacc -3') and Rv-2 with Fw (5'-
94 tGCAGCGCTcttgaacaatggaacctagtaaatagg -3'), for₁₇LEQ₁₉-M Fw-2 with Rv (5'-
95 TGCTGCAGCgagcttttaagctcttcaac-3') and Rv-2 with Fw (5'-
96 cGCTGCAGCAtggaacctagtaaatagttcc -3') and for E11,12,18A-M Fw-2 with Rv (5'-
97 attgCGCaaggagcttttaagCGCTGCaac -3') and Rv-2 with Fw (5'-aaaagctccttGCGcaatgg-3').
98 Similarly, two mutated MERS-CoV M fragments were generated by PCR using pcDNA3.1-V5-
99 MERS-M as a template. The following primers were used for₆QLT₈-M: Fw2 with Rv (5'-
100 tcAGCGGCTGCcgtcatattggatccCGTAGAATC-3') and Rv-4 (5'-aaaagctccttGCGcaatgg-3') with Fw
101 (5'- CAGCCGCTgagcgcagattattgccattataaa - 3'), for₉EAQ₁₁-M: Fw2 with Rv (5'-
102 atCGCgcCGCagtgagttgctcatattggatcc -3') and Rv-4 with Fw (5'-
103 GCGgcgGCGattattgccattataaagactggaa - 3'), for₁₂IIA₁₄-M: Fw2 with Rv (5'-
104 aatggcTGCAgctcctcagtgagttgctcatatt -3') and Rv-4 with Fw (5'-
105 gcagGCTGCAgccattataaagactggaactttgcatgg - 3'), for₁₅IKD₁₈-M: Fw2 with Rv (5'-
106 GGCTGCAGCAGCggcaataatctgCGcctcagtg -3') and Rv-4 with Fw (5'-
107 CAgaattcctaagctcgaagcaatgcaa - 3'), for E9A-M: Fw2 with Rv (5'-
108 tctgcgCGCagtgagttgctcatattgg -3') and Rv-4 with Fw (5'- tGCGgcgagattattgccattataaaga -
109 3'), for E9D-M Fw2 with Rv (5'- cgcATCagtgagttgctcata -3') and Rv-4 with Fw (5'-
110 cactGATgcgagattattgcc - 3'), for E9D18A-M Fw2 with Rv (5'-
111 caaagtccaCGCtttaataatggcaataatctgCGcCGCagtg -3') and Rv-4 with Fw (5'-
112 cactGCGgcgagattattgccattataaaGCGtggaactttg - 3'). Then the two PCR products were fused
113 by using the primers Fw-3 (5'-CGgatccaatgatgagcaactcact-3') and Rv-4.
114 To generate the M-linker construct (pcDNA3.1-MERS-M-linker), two PCR products were
115 generated by using the primers Fw-3 with Rv (5'-
116 CTTAATAGTTGTATTTAATTTATAGTTTCCAATACGGTAGCGGTTGTATGCGTAtagatcaattggataaacg-3')
117 and Rv-4 with Fw (5'-
118 TAAATTAATACTATTAAGGACCTGGACTACAAGGACGACGATGACAAGCCAAAAGAGgcttccag
119 ataactc-3'). Then the two PCR products were assembled by a second PCR with Fw-3 and Rv-4
120 primers. Then the pcDNA3.1-MERS-M-linker was used as a template to introduce the N3Q and
121 N88Q mutations. For the N3Q mutations, only one round of PCR was used to insert the
122 mutation. The M-linker coding sequence was amplified with Fw-4 (5'-
123 CGgatccCAAatgacgcaactcactga-3') and Rv-4. For the N88Q mutations, the two first PCR were
124 performed with Fw-3 and Rv (5'- CCTAATAGTTGTTGTAATTTATAGTTTCC-3') and with Fw (5'-
125 GGAAACTATAAATTACAAACAATAAAGG-3') and Rv-4. The two PCR products were assembled

126 by using Fw-3 and Rv-4 for the single mutation or with Fw-4 and Rv-4 for the double mutation
127 (N3Q,N88Q). The E9,D18-A mutation were introduced as done for MERS-M without linker and
128 by using pcDNA3.1-V5 -M-linker or pcDNA3.1-V5-N3Q-M-linker as a template for the first
129 round of PCR.

130 For the E1-M chimera, E1 coding sequence was added into the forward primer used for PCR.
131 The SARS-CoV-2 M protein coding sequence was amplified with 5'-
132 ACggatccGCCACcatgGCTCAGGTGAAGAATAGCAGTGGCCTCTACgttgaagagcttaaaaagct-3' and Rv-
133 1.

134 ORF3a coding sequence was amplified from infected cells cDNA by using 5'-
135 actctgagccagtgtctcaaag-3' and 5'-gaaacgaatgagtacataagt-3' primers. Then a second PCR was
136 performed to introduce the restriction site with the following primers:
137 TCggatccGCCACCATGGATTTGTTTATGAGAA-3' and 5'-ccgaattcCAAAGGCACGCTAGTAGTCG-3'.

138 To produce the Nt-M/ORF3a chimera, the M sequence was amplified with Fw-2 and Rv (5'-
139 GCCAACAATAAGCCATCCgtaatagtagcttgaatc-3') and ORF3a with Fw (5'-
140 gattccaacggtagctattaccGGATGGCTTATTGTTGGC-3') and Rv-5 (5'-
141 ccgaattcCAAAGGCACGCTAGTAGTCG-3'). For the TM1-M/ORF3a chimera, the M protein
142 sequence was amplified with Fw-2 and Rv (5'-
143 CTCTTTTGGAGGGTTATGATaaattgtagaagacaaatcc-3') and ORF3a with Fw (5'-
144 ggattgtctctacaattATCATAACCCTCAAAAAGAG-3') and Rv-5. For both chimeras, the PCR
145 products were assembled by amplification with Fw-2 and Rv-5.

146 Finally, for all of the constructs, the final PCR products were inserted between the BamHI and
147 EcoRI restriction sites of the vector and the constructions were verified by DNA sequencing.

148
149 To generate the KO cell lines, targeted DNA sequence were inserted into the pX330 vector as
150 previously described [8]. Briefly, 10 μ M of oligonucleotides were annealed and
151 phosphorylated with T4 polynucleotide kinase (NEB). Then the oligonucleotides and the
152 vector were digested with BbsI and ligated in a single step. Residual linearized DNA was
153 removed by exonuclease treatment and the ligation product transformed in Top10 E. Coli
154 competent cells. The following sequences were used for β 3GNT2: 5'-
155 CACCGACTACTGCGAACCTGACCTG-3' and 5' AAACCAGGTCAGGTTTCGACAGTAGTC-3', for STT3A:
156 5'- CACCGTACTCACGGATCAAACCTCA-3' and 5'-AAACTGAGTTTGATCCGTGAGTAC-3' and for
157 STT3B: 5'- CACCGTACAGCAAAAAGAGTCTACAT-3' and 5'-AAACATGTAGACTCTTTTGCTGTA-3'.

158
159 **KO cell line.**

160 Huh-7 cells were seeded in 6-well plates and co-transfected with 1 μ g of pX330 vector coding
161 for small guide RNA targeting β 3GNT2, STT3A, STT3B or pX330 empty vector and 50 ng of
162 plasmid containing a puromycin resistance gene. The transfection was performed with
163 TransIT[®]-LT1 transfection reagent (Mirus Bio) according to the manufacturer's instruction.
164 Twenty-four hours after transfection, 5 μ g/ml of puromycin was added to the cell culture
165 medium. Selection was maintained for 4 days. For STT3A and STT3B, cell population were used
166 for expression of M proteins. STT3A and STT3B expression was analyzed in immunoblot with
167 polyclonal anti-STT3A (Fisher Scientific) and STT3B (Sigma)

168 For β 3GNT2 KO cells, clones were isolated by serial dilutions of the cell population obtained
169 after puromycin selection. After sequencing, a β 3GNT2 KO clone was selected showing a
170 truncated protein with stop codon at position 109.

171
172 **Glycosidase treatment**

173 Huh-7 cells plated into 6 well-plates were transfected with 1 µg of pcDNA3.1-V5-MERS-M,
174 pcDNA3.1-V5-N3Q-MERS-M, pcDNA3.1-V5-SARS-2-M or pcDNA3.1-V5-N5Q-SARS-2-M by
175 using TransIT®-LT1 transfection reagent (Mirus Bio) according to the manufacturer's
176 instruction. Twenty-four hours later, cells were lysed on ice with B1 buffer (50 mM Tris-HCl
177 pH7.5, 100 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS) containing protease inhibitors.
178 Cell lysates were cleared by centrifugation at 14,000 rpm for 5 min. Then, lysates were mock-
179 treated or treated with Endoglycosidase H or PNGase F according to the manufacturer's
180 recommendations. Proteins were separated by SDS-PAGE and transferred onto a
181 nitrocellulose membrane. Membrane proteins were detected by using an anti-V5 antibody
182 (Invitrogen, R96025) and horseradish peroxidase-conjugated anti-mouse secondary antibody
183 (Jackson Immunoresearch). Detection was carried out by chemiluminescence (Pierce).
184 For quantification of the unglycosylated form of the protein, cell lysates were either mock
185 treated or PNGase F treated and then samples were analyzed in western blot. The band
186 corresponding to the unglycosylated form was quantified in both samples using the band
187 quantification function of Image J. The PNGase F treated samples allows for the quantification
188 of the total amount of M protein in the sample.

189

190 **Lectin pulldown**

191 Huh-7 cells plated into 6 well-plates were transfected with 1 µg of pcDNA3.1-V5-MERS-M,
192 pcDNA3.1-V5-N3Q-MERS-M, pcDNA3.1-V5-SARS-2-M or pcDNA3.1-V5-N5Q-SARS-2-M by
193 using TransIT®-LT1 transfection reagent (Mirus Bio) according to the manufacturer's
194 instruction. Twenty-four hours later, the cells were lysed on ice with B1 buffer containing
195 protease inhibitors. Cell lysates were cleared by centrifugation at 14,000 rpm for 5 min. An
196 aliquot of lysates was kept for western blot analysis, the rest was incubated with DSL or LEL-
197 conjugated agarose beads (Eurobio) for 2h at 4°C on a wheel. Then, beads were washed three
198 times with B1 buffer and resuspended in 2X Laemmli loading buffer. Precipitated proteins
199 were detected by immunoblot. Proteins were separated by SDS-PAGE and transferred onto a
200 nitrocellulose membrane. Membrane proteins were detected by using an anti-V5 antibody
201 (Invitrogen, R96025) and horseradish peroxidase-conjugated anti-mouse secondary antibody
202 (Jackson Immunoresearch). Detection was carried out by chemiluminescence (Pierce).

203

204 **Infection**

205 β 3GNT2 KO and control cells were infected at a MOI of 0.1 with MERS-CoV or SARS-CoV-2. For
206 SARS-CoV-2 infection, cells were previously transduced with a lentivirus expressing ACE2 48h
207 before infection with SARS-CoV-2. 16h after infection, cells were lysed in B1 buffer and
208 centrifuged at 14,000 rpm. Cell lysates were then heated at 95°C for 20 min before removing
209 from the BSL3 facility. Proteins were separated on SDS-PAGE and transferred on a
210 nitrocellulose membrane. Membrane proteins were detected with anti-M antibodies
211 (proteogenix for MERS-CoV M and Abnova PAB31758 for SARS-CoV-2 M) and horseradish
212 peroxidase-conjugated anti-rabbit secondary antibodies (Jackson Immunoresearch).
213 Detection was carried out by chemiluminescence (Pierce).

214

215 **Immunofluorescence and selective permeabilization assay.**

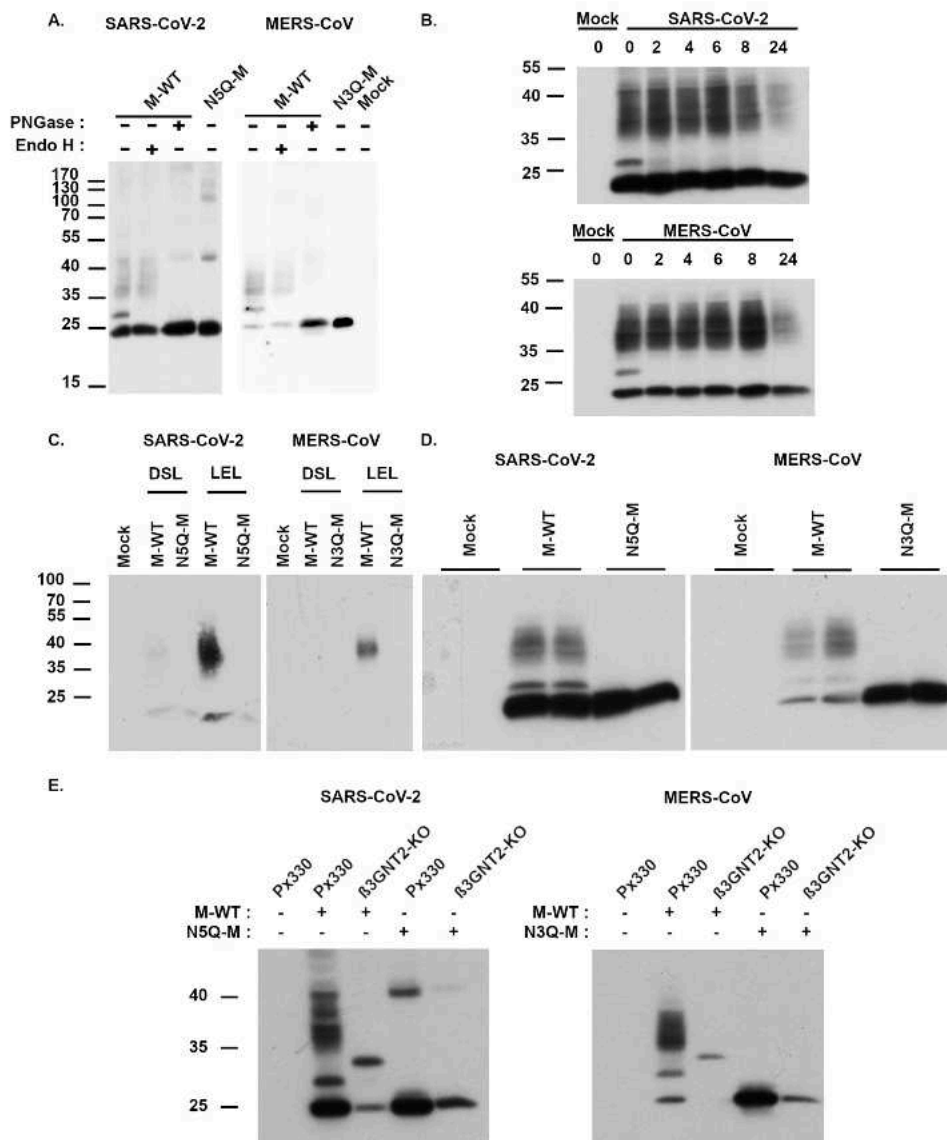
216 Huh-7 cells plated on coverslips were transfected with vector expressing the different
217 membrane protein mutants. The next day, cells were fixed with 3% paraformaldehyde, then
218 permeabilized with 0.1% Triton X-100 or by freeze-thaw cycle for selective permeabilization
219 assay. The cells were frozen at -80°C in PBS then allowed to thaw at room temperature. Then,

220 cells were blocked in 5% horse serum in PBS for double stainings or 5% goat serum for simple
221 labelling. Next, cells were incubated with anti-V5 mAbs (Invitrogen, mAb 2F11F7) or anti- E1
222 (A6 mAb) and a sheep anti-TGN46 pAbs (Biorad) or a rabbit anti-calnexin (Abcam) diluted in
223 blocking solution. Cells were rinsed 3 times with PBS and incubated with Cy3-conjugated
224 donkey anti-mouse (Jackson immunoresearch) and A488-conjugated donkey anti-sheep
225 secondaries antibodies for V5 and TGN-46 co-staining and Cy3-conjugated donkey anti-mouse
226 and A488-conjugated donkey anti-rabbit secondaries antibodies (Jackson immunoresearch)
227 for V5 and Calnexin co-staining. Anti-E1 was detected with A488-conjugated donkey anti-
228 human (Jackson immunoresearch) in combination with Cy3-conjugated donkey anti-rabbit
229 (Jackson immunoresearch) or Cy3-conjugated donkey anti-sheep for calnexin or TGN46 co-
230 staining. Nuclei were stained with 1 µg/ml 4,6-diamidino-2-phenylindole.
231 Images were acquired using a laser-scanning confocal microscope LSM 880 (Zeiss) using a
232 63x oil immersion objective.
233

234 **Results**

235 **SARS-CoV-2 and MERS-CoV M proteins are N-glycosylated.**

236 In a previous study, we noticed that the MERS-CoV M protein migrated as three bands in
237 immunoblot, corresponding to 3 M species differing in their glycosylation state [9]. When
238 analyzing the migration profile of the SARS-CoV-2 membrane protein in immunoblot after
239 transient expression in Huh-7 cells, a similar migration pattern was observed (Figure 1A). The
240 first band around 25 kDa corresponded to the non-glycosylated protein as shown by the
241 migration profile of the M proteins with a mutated N-glycosylation site (N5Q-M for SARS-CoV-
242 2 and N3Q-M for MERS-CoV) or the M proteins treated with PNGase F. A second band between
243 25 and 35 kDa was sensitive to Endo-H treatment and hence corresponded to a high-mannose
244 form of the glycans linked to the M protein (obtained in ER/cis-Golgi) that has not reached the
245 medial Golgi for further modification. The third species of M protein migrated as a diffuse
246 band (35-40kDa) that was insensitive to Endo-H but sensitive to PNGase F, confirming that this
247 band corresponded to M proteins modified with complex N-glycans.
248



249
250 **Figure 1**
251 **N-glycosylation of MERS-CoV and SARS-CoV-2 membrane proteins.** (A.) Lysates of cells transfected with
252 plasmids encoding the V5-tagged SARS-CoV-2 or MERS-CoV membrane proteins were treated with Endo-H or
253 PNGase F. M proteins were detected with an anti-V5 antibody. Lysates of cells expressing a N-glycosylation
254 defective mutant (N3Q-M for MERS-CoV or N5Q-M for SARS-CoV-2) were loaded as control. (B.) Cells were
255 transfected with vector expressing the SARS-CoV-2 or MERS-CoV WT-M. Then, cells were treated with 100 μM
256 cycloheximide to block protein synthesis for different period of time (2, 4, 6, 8 or 24h) before lysis. M proteins
257 were detected with an anti-V5 antibody. (C.) Lysates of cells transfected with vectors expressing the membrane
258 proteins were precipitated with *Datura stramonium* lectin (DSL) or *Lycopersicon esculentum* lectin (LEL)-
259 conjugated beads and membrane proteins were detected in western blot (precipitation, left panels). (D.) M
260 proteins expression in cell lysates (=input) was also analyzed by western blot. (E.) Control (px330) and β3GNT2

261 KO cells were transfected with vectors expressing the wild-type or the N-glycosylation deficient form of the
262 MERS-CoV and SARS-CoV-2 M proteins. Cells were lysed and M expression was detected in western blot.

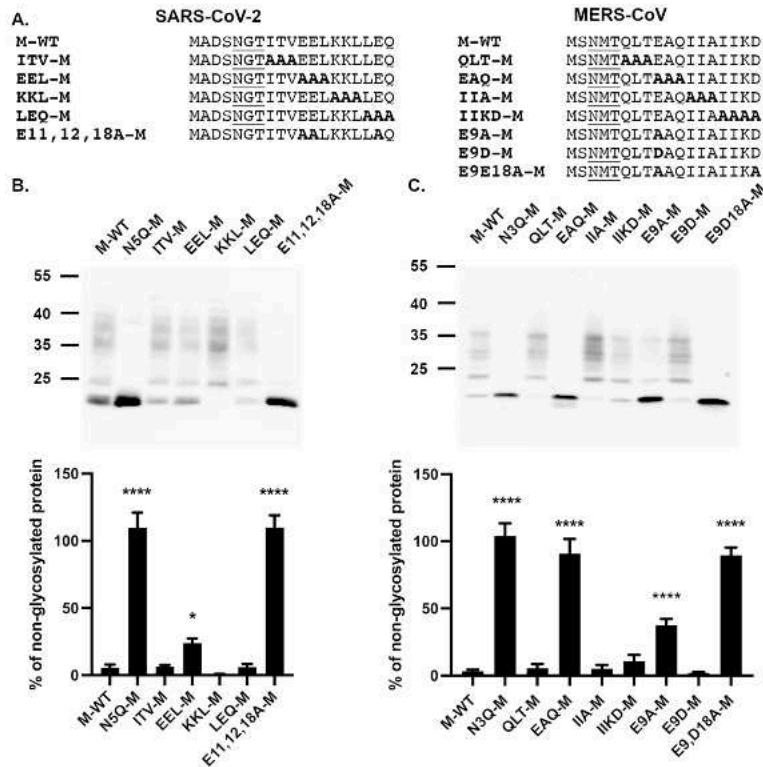
263
264 Then, we transfected the cells to express the SARS-CoV-2 and MERS-CoV M proteins and
265 treated them with cycloheximide to block protein synthesis for different period of time.
266 Starting at 2h of cycloheximide treatment, the band corresponding to high-mannose
267 conjugated M disappeared showing that the high-mannose conjugated M are then processed
268 into complex glycans-conjugated M (Figure 1B). The band corresponding to the unglycosylated
269 protein remains stable overtime suggesting that some proportion of the protein does not
270 undergo glycosylation and therefore is not a precursor of the glycosylated M.

271
272 **SARS-CoV-2 and MERS-CoV M proteins are modified with polylectosamines.**

273 We further investigated the nature of these complex N-glycans. It has been suggested that the
274 infectious bronchitis virus (IBV) membrane protein is modified by polylectosamine chains [10].
275 Polylectosamine chains are constituted by repeats of N-acetyllectosamine (LacNac). Each
276 LacNac is composed of galactose linked to N-acetylglucosamine through a β 1-4 bond (Gal β 1-
277 4GlcNac). In polylectosamine chains, LacNac repeats are linked by β 1-3 bonds.
278 Polylectosamine chains can occur on multiantennary glycans, the chains can be linear or may
279 also become branched by the addition of β 1-6GlcNac to internal galactose [11].

280 To test if SARS-CoV-2 and MERS-CoV M proteins could harbor such modifications, we
281 performed a protein pulldown with beads conjugated with lectins known for their affinity to
282 polylectosamine motifs (Figure 1C and 1D). We used two different lectins, the *Lycopersicon*
283 *esculentum* lectin (LEL) and the *Datura stramonium* lectin (DSL) to precipitate proteins from
284 cells transfected with vectors expressing either the wild-type or the non-glycosylated M
285 proteins of MERS-CoV and SARS-CoV-2. As shown in Figure 1C, only LEL-conjugated beads
286 precipitated the M protein modified with complex N-glycans. The lack of recognition by DSL
287 could be due to the fact that this lectin binds preferentially polyLacNac on branched glycans.
288 Furthermore, its affinity increases with higher branching and DSL binding is inhibited by α 2-6
289 sialic acid terminal modification whereas LEL tolerates terminal fucose and sialic acid [12]. To
290 confirm that M proteins are modified with polylectosamine glycans, we generated a β 1,3-N-
291 acetylglucosaminyltransferase 2 (β 3GNT2) knockout Huh-7 cell line by using CRISPR-Cas 9
292 mediated gene inactivation and expressed the M proteins in these β 3GNT2 KO cells, since
293 β 3Gnt2 is a major polylectosamine synthase [13]. In absence of an antibody recognizing
294 β 3Gnt2 in western blot, we isolated clones and sequenced the β 3GNT2 gene. We choose to
295 use a clone with a stop codon in the protein at position 109 for our experiment. As shown in
296 Figure 1E, the third M protein species appeared as a more defined band migrating slightly
297 slower than the high-mannose-conjugated M proteins in this cell line.

298 These results show that MERS-CoV and SARS-CoV-2 membrane proteins are modified with
299 polylectosamine chain.



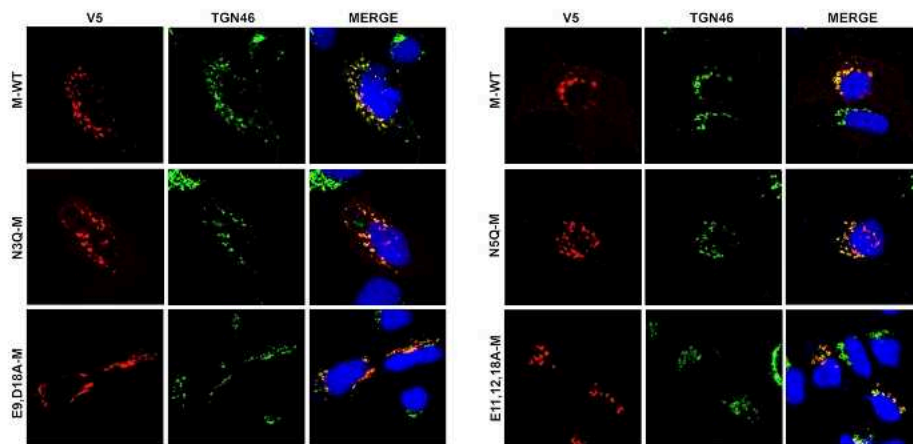
300
301
302
303
304
305
306
307
308
309
310
311
312

Figure 2.
Mutation of acidic residues of the SARS-CoV-2 and MERS-CoV M proteins abrogates N-glycosylation. (A.) Amino-acid sequences of the different mutants that were generated. (B.) Western blot analysis of the different mutants expression in transfected cells. (C.) Quantification of the non-glycosylated M protein for each mutant. Cells were transfected, then lysates were either mock or PNGase F-treated before detection in western blot. Then the intensity of the band corresponding to the non-glycosylated form of the protein was quantified and normalized with the PNGase F-treated sample reflecting the total amount of protein. Results are presented as the percentage of non-glycosylated proteins and are expressed as the mean of at least four independent experiments. Error bars represent the standard error of the mean (SEM). Results were analyzed by using an ANOVA test (* P<0.1; **** P< 0.0001).

313 **Nearby acidic residues in the first transmembrane segment are important for M protein**
314 **glycosylation.**
315 The N-terminus domain of M proteins is very short, so we investigated whether amino acid
316 residues in this region were important for the protein to become modified with
317 poly lactosamine chains. As we started this study before the release of the SARS-CoV-2 M
318 protein structure, we predicted that the N-terminus domain would be around 20 residues
319 long. Therefore, several residues located further downstream of the N-glycosylation site were
320 mutated 3 by 3 into alanine, generating ITV-M, EEL-M, KKL-M, LEQ-M mutants for SARS-CoV-
321 2 and QLT-M, EAQ-M, IIA-M, and IIKD-M mutants for MERS-CoV (Figure 2A). All the mutants
322 that were glycosylated, were modified with poly lactosamine chains (Figure 2B and C, top
323 panels). However, an accumulation of non-glycosylated proteins was noticed when the
324 residues ₁₁EEL₁₃ in the SARS-CoV-2 M protein and ₉EAQ₁₁ in the M protein of MERS-CoV were

325 mutated. Because of its heterogeneity, the complex glycan conjugated M protein can be
 326 difficult to detect in immunoblot, therefore to better compare the glycosylation state of the
 327 mutants with the wild-type protein, the non-glycosylated species were quantified for each
 328 mutant. The total amount of M protein was quantified after deglycosylation of the sample by
 329 PNGase F treatment (Figure 2C). We found that the majority of wild-type proteins became
 330 glycosylated as only 3.32 +/-1.13 % and 5.73 +/- 2.18 % of MERS-CoV and SARS-CoV-2 M
 331 proteins were found to be unglycosylated, respectively. As expected, MERS-CoV N3Q-M and
 332 SARS-CoV-2 N5Q-M remained completely unglycosylated. These quantifications also
 333 confirmed that the SARS-CoV-2 EEL-M mutant showed a slight defect in glycosylation, whereas
 334 the MERS-CoV EAQ-M mutant remained largely unglycosylated (90.9%+/-10.82). As both
 335 mutants involved glutamic acid residues, the role of such acidic residues was further
 336 investigated with additional mutant proteins, including, E11,12,18A-M for SARS-CoV-2 and
 337 E9A, E9D, and E9A,D18A-M for MERS-CoV. Mutation of E11,E12 and E18 in the SARS-CoV-2
 338 membrane protein completely abolished the glycosylation of the protein (Figure 2B). For
 339 MERS-CoV, the E9A,D18A-M mutant showed a strong defect of its glycosylation whereas the
 340 single mutant E9A had a partial effect and E9D mutation had no impact (Figure 2C). As the
 341 above mentioned ₉EAQ_{1,1} mutation had a stronger effect than the E9A mutation alone. We also
 342 constructed a Q11A-M mutant, but its glycosylation state was similar to the wild-type protein
 343 (data not shown). Nonetheless, these results suggest that apart from the intrinsic N-
 344 glycosylation site, acidic residues in the N-terminal part of SARS-CoV-2 and MERS-CoV M
 345 proteins are important for the glycosylation of the protein.

346



347

348

Figure 3

349

Mutants are located in the TGN. Cells were transfected with vector expressing the MERS-CoV or SARS-CoV-2 wild-type M protein or non-glycosylated proteins (MERS-CoV N3Q-M and E9D18A-M or SARS-CoV-2 N5Q-M and E11,12,18A-M) fixed and processed for immunofluorescence. Cells were stained with anti-V5 (red) and anti-TGN46 (green) antibodies.

350

351

352

353

354

To make sure that the absence of glycosylation was not due to a misfolding of the mutated protein, we investigated the intracellular location of the E9A,D18A-MERS-M and E11,12,18A-SARS-2-M, expecting that if the mutants were misfolded, then they would no longer localize in the *trans*-Golgi network but would be retained in the endoplasmic reticulum. We did not detect any change of the localization of the mutants, all of them colocalizing with TGN46, a

355

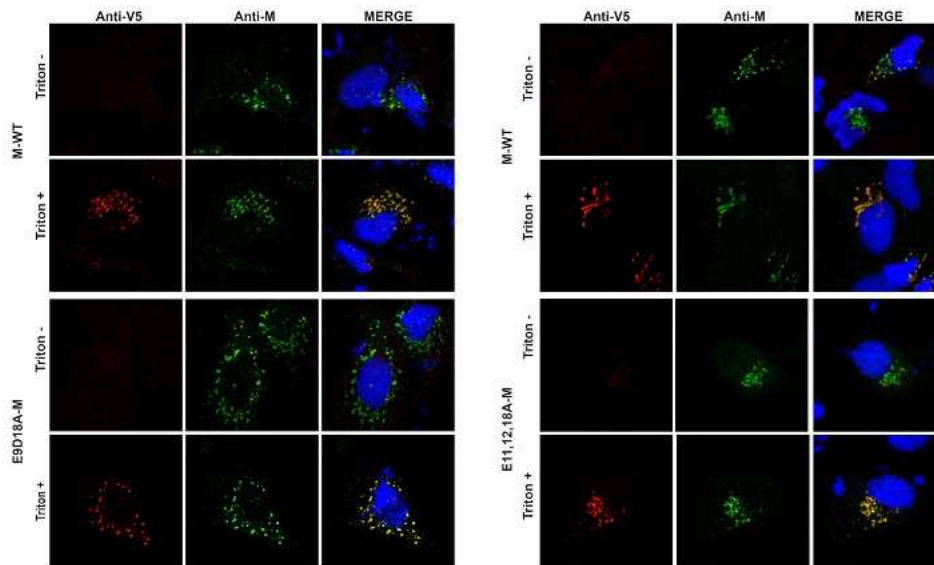
356

357

358

359 marker of the *trans*-Golgi network (Figure 3). These results suggest that the E9A,D18A-MERS-
360 M and E11,12,18A-SARS-2-M are correctly folded.

361 The M protein is a triple membrane-spanning protein with its N-terminus in the lumen of
362 intracellular compartments and the C-terminus exposed into the cytosol. This topology with
363 the N-terminus facing the lumen of the intracellular compartments is very important for the
364 glycosylation of the protein, therefore we tested if the N-terminus of our mutants is accessible
365 to antibodies when performing selective permeabilization of the cells. To specifically
366 permeabilize the plasma membrane of the cells, we performed freeze/thaw cycles of the cells
367 after fixation. As observed for the wild-type protein, the N-terminal V5-tagged extremity of
368 the wild-type or the mutated M proteins was not accessible to antibodies with selective
369 permeabilization but was detected upon permeabilization with 0.1% Triton X-100 (Figure 4).
370 In contrast, the protein was detected with an antibody directed against the C-terminus of the
371 protein with both permeabilization methods. These results suggest that the topology of the
372 protein is not affected when the acidic residues were mutated into alanine.
373



374
375

Figure 4

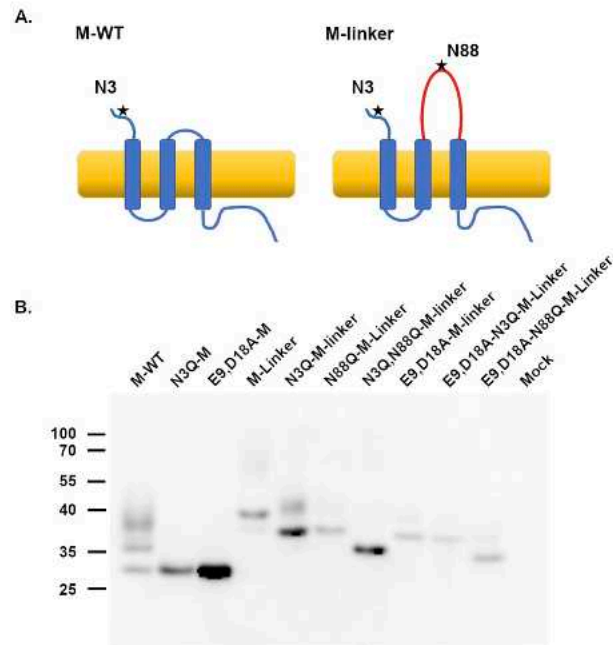
The N-terminal extremity of mutant proteins is exposed in the luminal side of intracellular compartments.

376 Cells expressing the N-terminal V5 tagged MERS-CoV wild-type or E9,D18A-M or the SARS-CoV-2 wild-type or
377 E11,12,18A-M proteins were fixed. Cells were either permeabilized by freeze/thaw (Triton-) or with 0.1% Triton
378 X-100 (Triton+) and stained with anti-V5 (red) or with an anti-MERS-CoV M or anti-SARS-CoV-2 M that are both
379 directed against the C-terminal domain of the proteins (green).
380

381

382 Finally, to confirm these findings we also inserted an N-glycosylation site in a linker that we
383 inserted between the second and third transmembrane segment of the MERS-CoV M protein
384 as previously reported for the SARS-CoV M protein [14] (Figure 5A). To have a better
385 understanding of the different migration profile of our construct with the linker, we also
386 constructed the M-linker protein with mutations of the two glycosylation sites (N3Q and/or
387 N88Q). As shown by the different migration profiles of N3Q-M, N88Q-M and N3Q,N88Q-M,
388 both sites were glycosylated and could be modified by polylectosamine chains (Figure 5B).
389 E9,D18 mutation abolished only the glycosylation of the N3 sites. Indeed E9D18A-M-linker

390 had a similar migration profile as the N3Q-M-linker or N88Q-M-linker mutants and the
 391 E9,D18A-N88Q-M-linker was similar to the double mutant N3Q,N88Q-M-linker.
 392 First, these results confirm the correct topology of the protein as N88 in the linker is accessible
 393 to glycosyl-transferases. Moreover, these results also show that the acidic residues act locally
 394 on the N-glycosylation of the protein as their mutation did not affect the glycosylation of the
 395 N-glycosylation site inserted into the linker.
 396

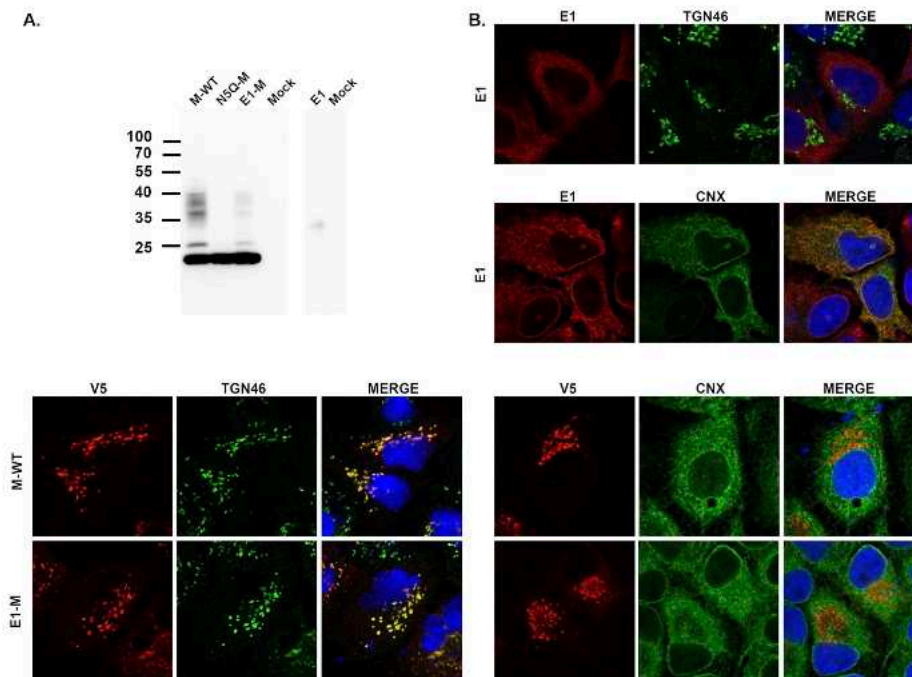


397
 398 **Figure 5**
 399 **N-glycosylation of a linker inserted between the second and third transmembrane segment of the MERS-CoV**
 400 **membrane protein is not dependent on the acidic residues of the N-terminal extremity.** (A.) Schematic
 401 representation of the wild-type M protein and the protein with a linker (red) inserted between the second
 402 and third transmembrane segment (M-linker). The positions of the glycosylation sites are depicted with stars. (B.)
 403 Cells were transfected with vectors expressing constructs containing the linker in combination with different
 404 mutations (E9,D18A and/or N3Q and/or N88Q). Proteins were analyzed in immunoblot with an anti-V5 antibody.
 405

406 **HCV E1-M and SARS-CoV-2 M-ORF3A chimeric proteins confirm the importance for correct**
 407 **subcellular trafficking and the role of the first transmembrane segment of M for its**
 408 **glycosylation.**

409 To further assess whether a motif in the N-terminal extremity of the M protein could be
 410 involved in the modification of the protein by poly-lactosamine chains, we constructed a
 411 chimera of the SARS-CoV-2 M protein in which the first 9 residues (MADSNNGTIT) were replaced
 412 by the 11 N-terminal residues of hepatitis C virus E1 protein (MAQVKNSSGLY). HCV envelope
 413 proteins are highly glycosylated but it has been shown that they do not contain complex N-
 414 glycans as they mainly localize in the endoplasmic reticulum. The E1 envelope protein also
 415 contains a N-glycosylation site in its N-terminus extremity [15]. When the N-terminus of E1
 416 was transferred on the SARS-CoV-2 M protein (E1-M), it was also modified by poly-lactosamine
 417 chains as shown by the migration profile in western blot (Figure 6A). This result suggests that

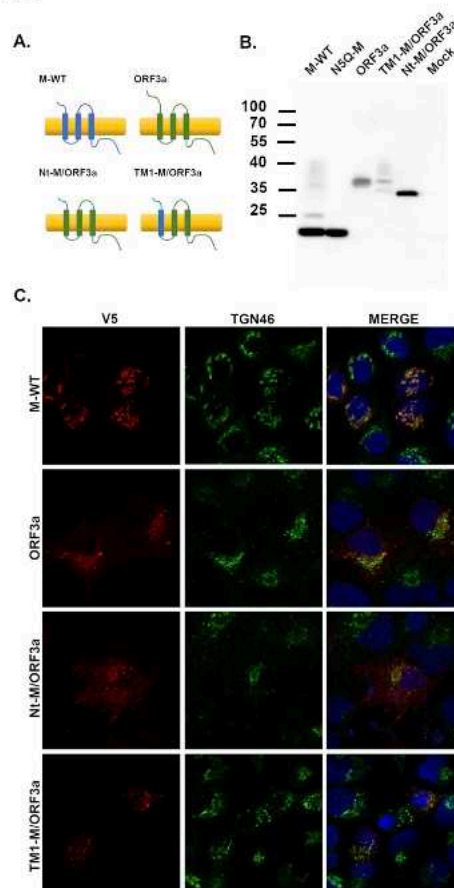
418 the trafficking of the protein through the Golgi is important for the modification of the protein
 419 with poly-lactosamine chains. We then analyzed the intracellular distribution of E1-M. E1 was
 420 mainly distributed in the endoplasmic reticulum, but the E1-M had the same localization as
 421 M-WT and co-localized with TGN46 (Figure 6B).
 422



423
 424 **Figure 6.**
 425 **HCV E1 N-terminal extremity is conjugated to complex-glycans when transferred on SARS-CoV-2 M protein.**
 426 (A.) Western blot analysis of SARS-CoV-2 M and the E1-M chimera with an anti-V5, E1 was detected with an anti-
 427 E1 (A6 mAb). (B.) Cells expressing V5-M, V5-E1-M and E1 were fixed and co-stained with either anti-V5 and anti-
 428 TGN46 or anti-calnexin (CNX) antibodies or anti-E1 and anti-TGN46 or anti-calnexin antibodies.
 429

430 The structure of SARS-CoV-2 M protein is very similar to the one of ORF3a [3, 4]. The N-
 431 terminal domain of ORF3a is composed of 42 residues, being much longer than the N-terminal
 432 domain of SARS-CoV-2 M. According to the M structure solved by Zhang and colleagues, the
 433 glutamic residues important for SARS-CoV-2 M protein glycosylation are located within the
 434 first transmembrane segment. As the first transmembrane segment of ORF3a is shorter and
 435 does not contain any acidic residues, we replaced its N-terminus by the one of SARS-CoV-2 M
 436 protein (Nt-M/ORF3a, Figure 7A). This chimera remained unglycosylated (Figure 7B). However,
 437 when we also replaced the first transmembrane segment of ORF3a with the one of SARS-CoV-
 438 2 M (TM1-M/ORF3a), then the protein was modified with poly-lactosamine chains as shown
 439 by its migration profile in western blot. ORF3a has been shown to localize in the endocytic
 440 pathway and at the cell surface [16]. As expected, we did not detect any co-localization of
 441 ORF3a with TGN46 in immunofluorescence and observed a punctuated staining consistent
 442 with endocytic compartments staining (Figure 7C). Both M-ORF3a chimeras had similar
 443 patterns as ORF3a staining and did not colocalize with TGN46. These results confirm the

444 important role played by the first transmembrane segment of the SARS-CoV-2 membrane
 445 protein for its glycosylation.

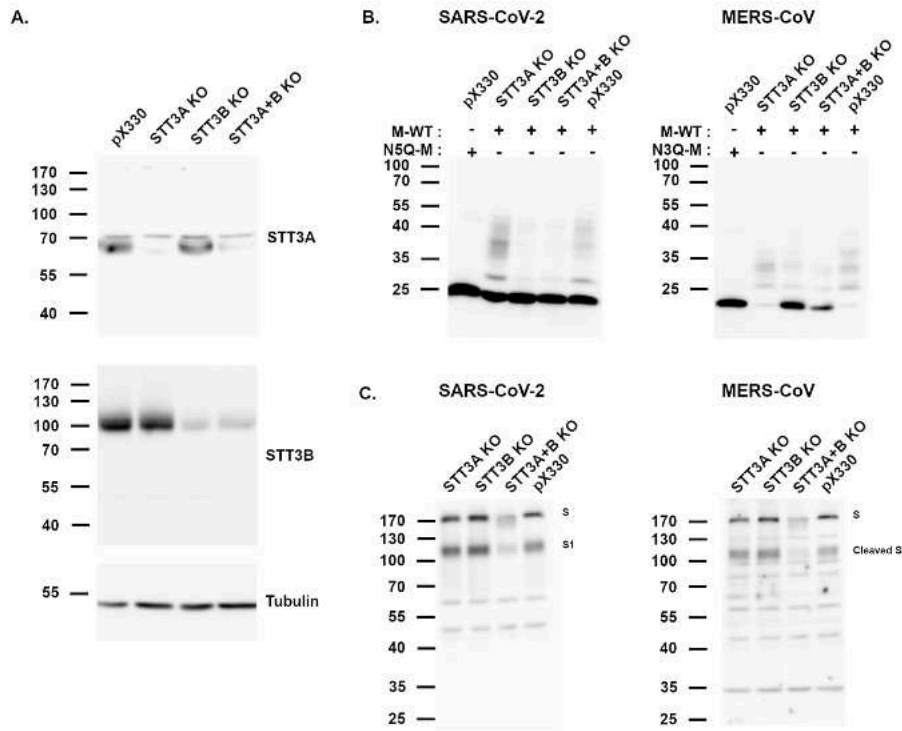


446
 447
 448 **Figure 7**
 449 **The first transmembrane segment of SARS-CoV-2 M is required for a chimera between M and ORF3a to be**
 450 **glycosylated.** (A.) Drawing of the different chimeras that were constructed between SARS-CoV-2 M (blue) and
 451 ORF3a (green). (B.) Western blot analysis of the different chimera. (C.) Cells expressing the V5-tagged SARS-CoV-
 452 2 M, ORF3a or the different chimeras were fixed and stained with anti-V5 (red) and anti-TGN46 (green).

453
 454 **OST-B is responsible for the glycosylation of M protein.**

455 Next, we investigated the machinery involved in the glycosylation of the M protein. Two
 456 oligosaccharyltransferases (OST) can mediate the transfer of the preassembled
 457 oligosaccharide onto the asparagine of N-glycosylation sequon. OST-A is associated to the
 458 protein translocon channel and acts co-translationally. OST-B is not associated with sec61
 459 complex but localizes to the rough endoplasmic reticulum and is considered as a proofreading
 460 machinery. The main difference between OST-A and B is in the catalytic subunit STT3. OST-A
 461 contains STT3A whereas OST-B contains STT3B. Therefore we invalidated the expression of
 462 STT3A or STT3B by CRISPR/Cas9 (Figure 8A) and expressed the M protein and the spike protein
 463 that is highly N-glycosylated in these knock-out cell populations [17]. SARS-CoV-2 M protein
 464 was no longer glycosylated in the STT3B-KO cells but fully glycosylated in the STT3A-KO cells

465 (Figure 8B). For MERS-CoV-M protein, we also detected an accumulation of non-glycosylated
 466 protein in STT3B-KO cells. A weak N-glycosylation could be detected that is likely due to
 467 residual expression of STT3B. In contrast, spike protein of both MERS-CoV and SARS-CoV-2
 468 were glycosylated in STT3A- and STT3B-KO cells (Figure 8C), however in the double knock-out
 469 cells, the expression of the protein was severely impaired suggesting a misfolding of the
 470 protein caused by the lack of glycosylation. For SARS-CoV-2 spike, a smear could be observed
 471 suggesting a modification of its glycosylation. These results suggest that M proteins are
 472 glycosylated by the OST-B complex.
 473

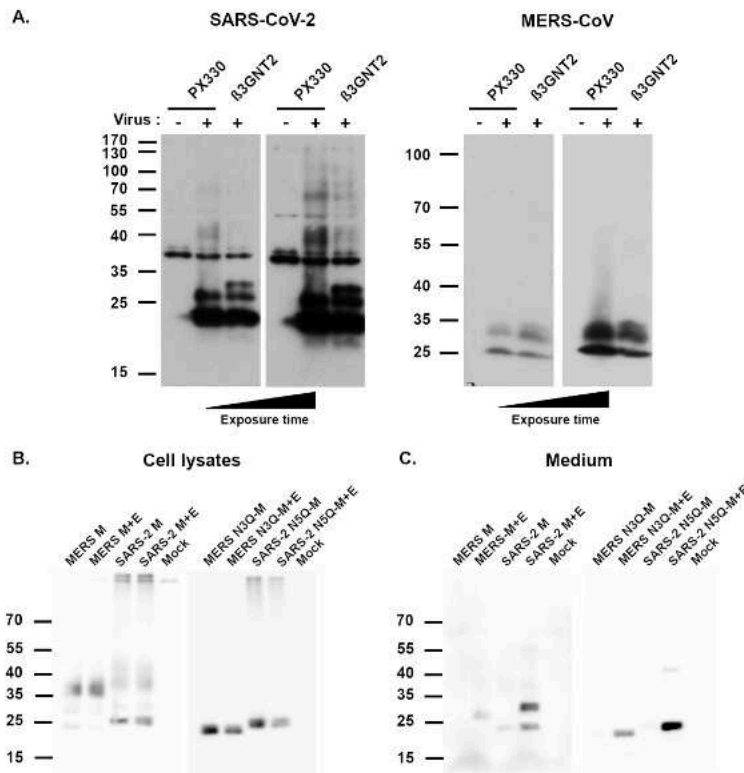


474 **Figure 8**
 475 **OST-B mediates M protein-glycosylation.** (A.) Expression level of STT3A and STT3B in knock-out cell population.
 476 (B.) STT3A-KO, STT3B-KO or control cells (pX330) were transfected with expression vector for M-WT or
 477 unglycosylated-M mutants (N3Q-M for MERS-CoV and N5Q-M for SARS-CoV-2). (C.) STT3A-KO, STT3B-KO or
 478 control cells (pX330) were transfected with expression vector for SARS-CoV-2 or MERS-CoV spike protein. SARS-
 479 CoV-2 spike protein was detected with an anti-S1 mAb and the MERS-CoV spike with a polyclonal anti-spike.
 480
 481

482 **Glycosylated proteins in infected cells and viral assembly.**

483 When expressed alone, the M protein is located in the TGN, however in infected cells, the M
 484 proteins is spread throughout the biosynthetic pathway, including the ERGIC where the new
 485 virions are formed. This difference in intracellular distribution of the protein may affect its
 486 glycosylation as glycosidases and glycosyl-transferases implicated in complex-glycan formation
 487 are located in the *trans*-Golgi. Because of the heterogeneity of the complex-glycan conjugated
 488 to the M proteins, these forms are difficult to observe in western blot. To confirm that the M
 489 protein is also modified with poly-lactosamine chain in infected cells, we took advantage of our
 490 β 3GNT2 KO cell line. Therefore, we infected this cell line with MERS-CoV or with SARS-CoV-2.

491 However, as Huh-7 cells do not express the receptor of SARS-CoV-2, they were transduced
 492 with a lentivirus expressing ACE2 prior to SARS-CoV-2 infection. Cells were infected and lysed
 493 24h later and the migration profile of membrane proteins were analyzed in immunoblots
 494 (Figure 9A). In control cells, mainly two bands could be observed for SARS-CoV-2 and MERS-
 495 CoV membrane protein. These bands correspond to the non-glycosylated protein and the
 496 high-mannose glycosylated protein. With longer exposure of the western blot, in the lanes
 497 corresponding to the control cells, a smear could be observed that is not present in β 3GNT2
 498 KO cells. In these cells, an additional band migrating slower than the high-mannose
 499 glycosylated protein was present. This result shows that M protein is also modified by
 500 polylactosamine chains in the context of infected cells.
 501



502
 503 **Figure 9 :**
 504 **M proteins are modified with polylactosamine chains in infected cells but high-mannose conjugated M**
 505 **proteins are incorporated in virus-like particles.** (A.) β 3GNT2 cells were infected with SARS-CoV-2 (left panel) or
 506 MERS-CoV (right panel) and lysed 24h later. Cell lysates were analyzed in western blot. M proteins were detected
 507 by using an anti-M antibody. (B.) Cells were transfected with plasmids encoding either the V5-tagged wild-type
 508 M proteins of SARS-CoV-2 and MERS-CoV (left panel) or their non-glycosylated forms (right panel), in absence or
 509 presence of an E-encoding plasmid. Cells lysates were loaded onto SDS-PAGE and M proteins were detected in
 510 immunoblot with an anti-V5 antibody. (C.) Supernatants were collected and ultracentrifuged on 20% sucrose
 511 cushion to concentrate the VLPs secreted in the medium. Then VLPs were resuspended and analyzed in western
 512 blot. The M protein was detected with an anti-V5 antibody.
 513
 514 As new virions bud in the ERGIC, it is likely that polylactosamines are not linked to the M
 515 proteins before the budding. After budding, virions transit through the Golgi and further

516 modifications of the proteins may occur. To identify the glycosylated forms of the protein that
517 are incorporated during viral assembly, we used a virus-like particle production assay. Huh-7
518 cells were co-transfected with a vector expressing the M protein alone or in combination with
519 E protein expression vector. VLPs released in the extracellular medium were concentrated by
520 ultracentrifugation on sucrose cushions. For both SARS-CoV-2 and MERS-CoV, the three M
521 glycoforms were detected in the cell lysates (Figure 9B). In contrast, the VLP fraction showed
522 a clear enrichment of the high-mannose form of the M protein for both viruses (Figure 9C).
523 When using both glycosylated and N3Q/N5Q unglycosylated M mutant proteins, VLP
524 formation was visible, confirming previous findings that coronaviruses do not use the N
525 glycans on the M proteins for assembly. Interestingly, the VLP formation was more efficient
526 for SARS-CoV-2 than MERS-CoV in the same experimental conditions. Taken together, these
527 results seem to be in accordance with virus budding occurring in the ERGIC compartment, but
528 also indicate that the M protein is not much modified after the formation of the new VLPs.

529 **Discussion**

530 N-glycosylation occurs on N-X-T/S sequon with X being any amino acid residue except proline,
531 however it is not known how the protein environment of the sequon influences its
532 glycosylation efficiency. Here, we show that nearby acidic residues in MERS-CoV and SARS-
533 CoV-2 membrane proteins are important for their glycosylation, their mutations completely
534 abolishing the glycosylation of the proteins. The structure of SARS-CoV-2 M protein has been
535 recently solved by two different groups [3, 4]. In the structure solved by Dolan et al, the N-
536 terminus domain structure of the protein comprising the 1-16 residues was not solved
537 suggesting that this domain was disordered and the first transmembrane segment was found
538 to start at residue 17. In the structure of Zhang et al, the first transmembrane segment was
539 found to start at residue 10, and only the structure of the first 9 residues was not solved.
540 Therefore, in this second structure, the acidic residues that are important for M glycosylation
541 are included in the first transmembrane segment. If the acidic residues are not part of the
542 transmembrane segment, then we can hypothesize that because of their charge, these
543 residues may push the N-terminus away from the surface of the ER membrane and facilitates
544 the recognition of the sequon by the OST. Membrane electrostatics largely depends on
545 phosphatidylserine (PS), a negatively charged lipid. This lipid only accounts for few percents
546 of the ER membrane composition, however, PS is mostly present on the luminal side of the ER
547 membrane [18]. On the other hand, these residues may also promote a direct interaction of
548 OST with the M protein. Two different machineries are responsible for the initial transfer of
549 the oligosaccharide precursor to the nascent protein. OST-A is directly associated with the
550 translocon channel whereas OST-B is not and is considered as a proofreading machinery [19].
551 We found that OST-B is likely responsible for SARS-CoV-2 and MERS-CoV M N-glycosylation.
552 Indeed, OST-B has been shown to be responsible for N-glycosylation of sites located near the
553 N-terminus of the protein such as the N-glycosylation site found in SARS-CoV-2 and MERS-CoV
554 M proteins. Acidic residues are relatively rare in transmembrane helices and their flanking
555 region, but it has been suggested that these negatively charged residues are more likely
556 depleted from the cytosolic side of the transmembrane region and enriched on the “outside”
557 face [20]. It has also been suggested that changing the flanking charges of a transmembrane
558 segment can reverse the topology of the TM [21], therefore we tested if the N-terminus of M
559 protein in which charged residues were mutated into alanine was still directed to the luminal
560 side of intracellular compartments. We did not see any change of protein topology either with
561 selective permeabilization and staining with an antibody recognizing the N-terminus of the

562 protein or by glycosylation of a linker inserted between the second and third transmembrane
563 segment of the MERS-CoV M protein. Moreover, mutation of the acidic residues in the first
564 transmembrane segment did not have any effect on the glycosylation of the linker inserted
565 between the second and third transmembrane segment of the protein, suggesting that the
566 acidic residues act locally on the structure of the M protein for its accessibility to OST.
567 We first thought that M protein must contain a signal that specifies the modification of the
568 protein with polylectosamines, that's why we did the serial mutations of the N-terminus
569 domain. We did not find any signal responsible for polylectosamine modification but residues
570 that are important for the glycosylation of the protein. We confirmed this absence of signal
571 by switching the N-terminal extremity of the M protein with the one of the E1 envelope
572 protein of HCV. Glycosylation of E1 has been extensively studied [15, 22]. E1 contains 5
573 potential N-glycosylation sites but only four of them are glycosylated. Of particular interest,
574 E1 contains an N-glycosylation site (E1N1) in its N-terminal extremity that is branched with
575 high-mannose glycans. Our E1-M chimera containing the E1N1 site is glycosylated with
576 complex glycans. This result suggests that trafficking through the Golgi is likely more important
577 for the modification with polylectosamines than a specific signal. Therefore, we constructed a
578 chimera between the M protein and ORF3a. ORF3a structure is very similar to the one of SARS-
579 CoV-2 M protein, however its N-terminal extremity with 42 residues is longer than the one of
580 SARS-CoV-2 M. The first transmembrane segment is 28 residues long for SARS-CoV-2 M and
581 20 for ORF3a, moreover, ORF3a TM1 does not contain any acidic residues. The chimera
582 containing only the N-terminal domain of M was not glycosylated, whereas addition of the
583 SARS-CoV-2 M TM1 rescues the glycosylation of the chimera that is modified with
584 polylectosaminoglycans. It has been reported that ORF3a colocalizes with markers of the
585 plasma membrane, the endocytic pathway and the Golgi. Therefore the protein is synthesized
586 in the ER and then crosses the Golgi before being sorted to the plasma membrane or
587 endosomes [16]. Similar to ORF3a, our chimeras did not colocalize with TGN46 suggesting that
588 these chimeras have a similar trafficking pattern as ORF3a. Taken together, our results suggest
589 that no signal specifies the modification with polylectosaminoglycans but the trafficking
590 through the Golgi that allows for the encountering of the proper glycosydases and glycosyl-
591 transferases is important. LAMP-2 is a lysosomal protein that contains 18 N-glycans some of
592 which are modified with polylectosaminoglycans [23]. It has been shown that the Golgi
593 residence time of the protein is associated with the extent of polylectosamines conjugation
594 [24]. It is interesting to note that it is not known for LAMP-2 why some N-glycosylation sites
595 are modified with polylectosamine glycans and the other ones with high-mannose. It has been
596 suggested that only membrane-anchored proteins can be modified with polylectosamines and
597 even if the proximity to the membrane has been evoked as a possible requirement for
598 polylectosamines conjugation, it has not been demonstrated [25].
599 The membrane proteins of MERS-CoV and SARS-CoV-2 are N-glycosylated and are processed
600 from a high-mannose form to a heterogeneous form of higher molecular weight. Because of
601 their heterogeneity, these forms can be difficult to visualize in immunoblot and can let to think
602 that the majority of the M protein is not glycosylated. However, when we quantified the non-
603 glycosylated form of the protein in cell lysates, only around 5% of the SARS-CoV-2 M and 3%
604 of the MERS-CoV M protein were not glycosylated. Our results show that these two proteins
605 are modified by polylectosamine chain. Other viral envelope proteins have been shown to
606 harbor similar complex glycans. Indeed, the respiratory syncytial virus small hydrophobic (SH)
607 protein has been shown to be modified with polylectosamines [26]. The function of this
608 protein still remains unclear, but it has been shown to be incorporated in viral particles [27].

609 Influenza B virus encodes a unique NB protein that has also been shown to harbor
610 polylysosamines modifications and to be incorporated into the virion [28, 29].
611 N-glycans have a wide variety of functions, they can affect their conformation, the solubility,
612 the activity or the antigenicity of glycosylated proteins. Functions of polylysosamine chains
613 remains unclear. β 3GnT2 is a key enzyme for the synthesis of polylysosamine chains. β 3GnT2
614 KO mice present an hyperactivation of lymphocytes, showing that polylysosamine chains are
615 important for immunological biofunctions. Polylysosamine chains are also involved in “i” and
616 “I” blood group antigens. Moreover, they can be recognized by glycan-binding proteins. For
617 example, polylysosamines conjugated to LAMP-2 can bind galectin-9. This binding is
618 important for lysosome homeostasis [30]. The role of the N-glycosylation of coronaviruses M
619 protein remains unclear. It is well-established that glycosylation of M protein is not necessary
620 for IBV or MHV virus assembly, and we obtained similar results with MERS-CoV and SARS-CoV-
621 2 M protein as we observed VLP formation with N-glycosylation defective M proteins.
622 However, we observed that the high-mannose glycoconjugated form of the M was mainly
623 incorporated in VLPs. For SARS-CoV-2, a small proportion of non-glycosylated protein was also
624 observed, suggesting that the non-glycosylated M protein could dimerize with N-glycosylated
625 M. As budding of coronaviruses occur in the ERGIC, it is not surprising that high-mannose
626 conjugated-M are incorporated into nascent particles. After assembly of the virus, it is not
627 clear how new virions are released from the cells, but our results suggest that the M protein
628 glycan does not undergo any further maturation after its incorporation into the viral particle.
629 Indeed, it has been suggested that β -coronaviruses may egress by lysosomal exocytosis [31],
630 but it is not clear where virions are sorted toward the lysosomal pathway, if it is directly from
631 the ERGIC or after crossing the Golgi.
632 The role of the glycosylation of coronaviruses membrane proteins is not clear. It has been
633 shown that M proteins of IBV, TGEV or MHV have interferogenic activity. The glycosylation
634 status of the protein is probably important for this activity. MHV M protein is O-glycosylated,
635 and recombinant viruses with a M protein either O-glycosylated, N-glycosylated or non-
636 glycosylated have been generated by De Haan *et al* [6]. The recombinant virus with a N-
637 glycosylated protein was a stronger inducer of type I IFN than the recombinant virus with a O-
638 glycosylated protein in *in vitro* assays. The recombinant virus with a non-glycosylated protein
639 was the poorer inducer. *In vivo*, the replication of the virus in the liver but not in the brain was
640 also dependent on the glycosylation status of the protein. Viral replication was the highest for
641 the virus with a N-glycosylated proteins. The pathogenicity of the different strains of MHV
642 varies. Whereas MHV-A59 causes an acute hepatitis, an acute encephalitis and chronic
643 demyelination in infected mice, MHV-2 produces only hepatitis and mild meningitis without
644 encephalitis or demyelination. Interestingly, it has been shown that the M protein of MHV-2
645 strain is N-glycosylated [32]. Although coronavirus cell tropism depends mainly on spike
646 protein, we cannot exclude the possible implication of M through its glycosylation status to
647 modulate cell tropism. Polylysosamines can also be linked to O-glycans, and we can
648 hypothesize that MHV M could also be modified with polylysosamines. Glycosylation of M
649 protein is conserved suggesting that it has a functional role. It is likely regulating host-
650 pathogen interactions however further investigations are necessary to characterize this
651 interaction and the role played by M protein polylysosaminoglycans.

652 **Author statements**

653 **1.5 Author contributions**

654 SB and JD conceived and designed the experiments. DJ, LD, AD and YR performed the
655 experiments. SB, DJ, LD, YR, JD analyzed the data. SB wrote the manuscript. All authors read
656 and approved the final manuscript.

657

658 **1.6 Conflicts of interest**

659 The authors declare that they have no conflicts of interest with the contents of this article
660

661 **1.7 Funding information**

662 L Desmarests is supported by a fellowship from the Region Hauts-de-France, D Juckel is
663 supported by a fellowship from the fondation pour la recherche médicale
664 (ECO202006011744). This work was supported by the Agence Nationale de la Recherche
665 (ANR-22-0E15-0030-01).
666

666

667 **1.8 Acknowledgements**

668 We thank Dr Philip Meuleman (Ghent University) for kindly providing the A6 mAb. The
669 immunofluorescence analyses were performed with the help of the imaging core facility of
670 the BioImaging Center Lille Nord-de-France. We thank Robin Prath and Nicolas Vandenneele
671 for their technical help in the BSL-3 facility
672

672

673 **References**

- 674 1. **Masters PS.** The molecular biology of coronaviruses. *Adv Virus Res* 2006;66:193–292.
- 675 2. **Escors D, Camafeita E, Ortego J, Laude H, Enjuanes L.** Organization of Two Transmissible
676 Gastroenteritis Coronavirus Membrane Protein Topologies within the Virion and Core.
677 *The Journal of Virology* 2001;75:12228–12240.
- 678 3. **Zhang Z, Nomura N, Muramoto Y, Ekimoto T, Uemura T, et al.** Structure of SARS-CoV-2
679 membrane protein essential for virus assembly. *Nat Commun* 2022;13:4399.
- 680 4. **Dolan KA, Dutta M, Kern DM, Kotecha A, Voth GA, et al.** Structure of SARS-CoV-2 M
681 protein in lipid nanodiscs. *eLife* 2022;11:e81702.
- 682 5. **Charley B, Lavenant L, Delmas B.** Glycosylation is Required for Coronavirus TGEV to
683 Induce an Efficient Production of IFNalpha by Blood Mononuclear Cells. *Scand J Immunol*
684 1991;33:435–440.
- 685 6. **de Haan CAM, de Wit M, Kuo L, Montalto-Morrison C, Haagmans BL, et al.** The
686 glycosylation status of the murine hepatitis coronavirus M protein affects the
687 interferogenic capacity of the virus in vitro and its ability to replicate in the liver but not
688 the brain. *Virology* 2003;312:395–406.

- 689 7. **Liang JQ, Fang S, Yuan Q, Huang M, Chen RA, et al.** N-Linked glycosylation of the
690 membrane protein ectodomain regulates infectious bronchitis virus-induced ER stress
691 response, apoptosis and pathogenesis. *Virology* 2019;531:48–56.
- 692 8. **Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, et al.** Genome engineering using the
693 CRISPR-Cas9 system. *Nature Protocols* 2013;8:2281–2308.
- 694 9. **Perrier A, Bonnin A, Desmarests L, Danneels A, Goffard A, et al.** The C-terminal domain
695 of the MERS coronavirus M protein contains a trans-Golgi network localization signal. *J*
696 *Biol Chem* 2019;294:14406–14421.
- 697 10. **Machamer CE, Mentone SA, Rose JK, Farquhar MG.** The E1 glycoprotein of an avian
698 coronavirus is targeted to the cis Golgi complex. *Proc Natl Acad Sci USA* 1990;87:6944–
699 6948.
- 700 11. **Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, et al. (eds).** *Essentials of*
701 *Glycobiology*. 4th ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
702 <http://www.ncbi.nlm.nih.gov/books/NBK579918/> (2022, accessed 7 July 2023).
- 703 12. **Bojar D, Meche L, Meng G, Eng W, Smith DF, et al.** A Useful Guide to Lectin Binding:
704 Machine-Learning Directed Annotation of 57 Unique Lectin Specificities. *ACS Chem Biol*
705 2022;17:2993–3012.
- 706 13. **Togayachi A, Kozono Y, Kuno A, Ohkura T, Sato T, et al.** β 3GnT2 (B3GNT2), a Major
707 Polylactosamine Synthase: Analysis of B3gnt2-Deficient Mice. In: *Methods in*
708 *Enzymology*. Elsevier. pp. 185–204.
- 709 14. **Voß D, Kern A, Traggiai E, Eickmann M, Stadler K, et al.** Characterization of severe acute
710 respiratory syndrome coronavirus membrane protein. *FEBS Letters* 2006;580:968–973.
- 711 15. **Dubuisson J, Hsu HH, Cheung RC, Greenberg HB, Russell DG, et al.** Formation and
712 intracellular localization of hepatitis C virus envelope glycoprotein complexes expressed
713 by recombinant vaccinia and Sindbis viruses. *J Virol* 1994;68:6147–6160.
- 714 16. **Miller AN, Houlihan PR, Matamala E, Cabezas-Bratesco D, Lee GY, et al.** The SARS-CoV-2
715 accessory protein Orf3a is not an ion channel, but does interact with trafficking proteins.
716 *eLife* 2023;12:e84477.
- 717 17. **Shrimal S, Cherepanova NA, Gilmore R.** Cotranslational and posttranslational N-
718 glycosylation of proteins in the endoplasmic reticulum. *Seminars in Cell & Developmental*
719 *Biology* 2015;41:71–78.
- 720 18. **Bigay J, Antony B.** Curvature, Lipid Packing, and Electrostatics of Membrane Organelles:
721 Defining Cellular Territories in Determining Specificity. *Developmental Cell* 2012;23:886–
722 895.
- 723 19. **Shrimal S, Cherepanova NA, Gilmore R.** Cotranslational and posttranslational N-
724 glycosylation of proteins in the endoplasmic reticulum. *Seminars in Cell & Developmental*
725 *Biology* 2015;41:71–78.

- 726 20. **Baker JA, Wong W-C, Eisenhaber B, Warwicker J, Eisenhaber F.** Charged residues next to
727 transmembrane regions revisited: “Positive-inside rule” is complemented by the
728 “negative inside depletion/outside enrichment rule”. *BMC Biol* 2017;15:66.
- 729 21. **Martínez-Gil L, Saurí A, Marti-Renom MA, Mingarro I.** Membrane protein integration
730 into the endoplasmic reticulum: Membrane protein integration. *FEBS Journal*
731 2011;278:3846–3858.
- 732 22. **Goffard A, Callens N, Bartosch B, Wychowski C, Cosset F-L, et al.** Role of N-Linked
733 Glycans in the Functions of Hepatitis C Virus Envelope Glycoproteins. *J Virol*
734 2005;79:8400–8409.
- 735 23. **Carlsson SR, Fukuda M.** The polylactosaminoglycans of human lysosomal membrane
736 glycoproteins lamp-1 and lamp-2. Localization on the peptide backbones. *J Biol Chem*
737 1990;265:20488–20495.
- 738 24. **Nabi IR, Dennis JW.** The extent of polylactosamine glycosylation of MDCK LAMP-2 is
739 determined by its Golgi residence time. *Glycobiology* 1998;8:947–953.
- 740 25. **Fukuda M, Guan JL, Rose JK.** A membrane-anchored form but not the secretory form of
741 human chorionic gonadotropin-alpha chain acquires polylactosaminoglycan. *Journal of*
742 *Biological Chemistry* 1988;263:5314–5318.
- 743 26. **Anderson K, King AMQ, Lerch RA, Wertz GW.** Polylactosaminoglycan modification of the
744 respiratory syncytial virus small hydrophobic (SH) protein: A conserved feature among
745 human and bovine respiratory syncytial viruses. *Virology* 1992;191:417–430.
- 746 27. **Huong TN, Ravi Iyer L, Lui J, Wang DY, Tan BH, et al.** The respiratory syncytial virus SH
747 protein is incorporated into infectious virus particles that form on virus-infected cells.
748 *Virology* 2023;580:28–40.
- 749 28. **Elderfield RA, Koutsakos M, Frise R, Bradley K, Ashcroft J, et al.** NB protein does not
750 affect influenza B virus replication in vitro and is not required for replication in or
751 transmission between ferrets. *Journal of General Virology* 2016;97:593–601.
- 752 29. **Williams MA, Lamb RA.** Polylactosaminoglycan Modification of a Small Integral
753 Membrane Glycoprotein, Influenza B Virus NB. *Molecular and Cellular Biology*
754 1988;8:1186–1196.
- 755 30. **Sudhakar JN, Lu H-H, Chiang H-Y, Suen C-S, Hwang M-J, et al.** Lumenal Galectin-9-Lamp2
756 interaction regulates lysosome and autophagy to prevent pathogenesis in the intestine
757 and pancreas. *Nat Commun* 2020;11:4286.
- 758 31. **Ghosh S, Dellibovi-Ragheb TA, Kerviel A, Pak E, Qiu Q, et al.** β -Coronaviruses Use
759 Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. *Cell* 2020;183:1520-
760 1535.e14.

761 32. **Yamada YK, Yabe M, Ohtsuki T, Taguchi F.** Unique N-linked glycosylation of murine
762 coronavirus MHV-2 membrane protein at the conserved O-linked glycosylation site. *Virus*
763 *Research* 2000;66:149–154.

764
765