

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
كلية علوم الطبيعة و الحياة

Faculté Sciences de la Nature
et de la Vie



جامعة الإخوة منتوري قسنطينة I
Frères Mentouri Constantin I University
Université Frères Mentouri Constantine I

Département de Biochimie et Biologie Cellulaire et
Moléculaire

قسم الكيمياء الحيوية والبيولوجيا
الخلوية والجزيئية

Mémoire en vue de l'obtention du Diplôme de Master en :

Biochimie

Thème :

**Analyse Viro-Informatique des
variants SARS-CoV-2 portée
sur la protéine Spike**

Présenté par :

CHEHBOUB Faten

ZITOUNI Rofaida

Jury :

Président D^r MOKRANI EL Hassen

Encadreur P^r HAMIDECHI Mohamed Abdelhafid

Examinatrice D^r GUENDOUZE Assia

Année Universitaire 2020 – 2021

Remerciements

Nous exprimons, par le biais de cet aboutissement modeste, tous nos profonds respects et remerciements à tous les enseignants, aux corps technique et administratif qui ont participé à notre apprentissage au sein de notre faculté et spécialement ceux de notre département de Biochimie et BCM.

Nos remerciements vont aux respectueux membres du jury qui aimablement accepté d'évaluer notre travail de fin de cycle, sachant leurs lourdes responsabilités surtout en fin d'année universitaire.

Nous exprimons toute cette reconnaissance Aux Docteurs :

MOKRANI El Hassen et GUENDOUIZE Assia.

Notre encadreur, le Professeur HAMIDECHI Mohamed Abdelhafid, est à remercier pour avoir accepté de diriger ce travail et de nous donner les enseignements nécessaires pour la réalisation et l'accomplissement de notre mémoire de fin d'études.

Nous ne pouvons nous permettre d'oublier nos très chers parents, car ils n'ont cessé de nous encourager pour avancer dans nos études et finir par voir un avenir tranquille. Que DIEU vous garde chers

PARENTS !

Table des matières

INTRODUCTION	1
PARTIE 1 : SYNTHÈSE BIBLIOGRAPHIQUE & ÉTAT DE L'ART	
1- STRUCTURE DU CORONAVIRUS	4
1-1/ Taxonomie	4
1-2/ Le Severe Acute Respiratory Syndrome (SARS)	6
2- GÉNOMIQUE ET PROTÉOMIQUE DU SARS-COV-2	7
3- STRUCTURE DE LA PROÉINE SPIKE	14
3-1/ Structure de la sous-unité S1	15
3-1/ Structure de la sous-unité S2	16
4- LES VARIANTS MAJEURS DU SARS-COV-2	19
4-1/ Les VOI	20
4-2/ Les VOC	21
5- LES MÉTHODES DE PRÉDICTION DES STRUCTURES 2D DES PROTÉINES	23
5-1/ Notions de structure 2D	23
PARTIE 2 : PARTIE PRATIQUE	
1- MATÉRIEL ET MÉTHODES	27
1-1/ Matériel	27
1-2/ Méthodes	28
1-2-1/ Prédiction de la structure secondaire	28
1-2-2/ Analyse des domaines transmembranaires hydrophobes	28
1-2-3/ Recherche des mutations	28
1-2-4/ Alignement multiple	29
2- RÉSULTATS & DISCUSSION	30
2-1/ PRÉDICTION 2D	30
2-2/ Prédiction des zones hydrophobes transmembranaires	30
2-3/ Recherche de mutations	33
2-4/ Alignement multiple	34
CONCLUSION	35
BIBLIOGRAPHIE	36

Liste des figures

N° et légende de la figure	Page
Figure 1 : Estimation des nombres de cas cumulatifs Covid dans le monde en date du 23 janvier 2020	1
Figure 2 : Estimation des nombres de cas cumulatifs Covid dans le monde en date du 13 juillet 2020	2
Figure 3 : Estimation des nombres de cas cumulatifs Covid dans le monde en date du 08 juillet 2021	2
Figure 4 : Structure du génome des coronavirus humains (HCoV) - HCoV-229E, HCoV-OC43, HCoV-NL63 et HCoV-HKU1	8
Figure 5 : Structure du génome de SARS-CoV-2 (Nawaz <i>et al.</i> , 2021).	9
Figure 6 : Différentes structures de la protéine S au sein de différents Coronavirus	12
Figure 7 : Structure schématique des deux sous- unités S1 et S2 de la protéine Spike.	15
Figure 8 : Structure schématique d'un trimère de la protéine S montrant les domaines HR1 et HR2 (Heptad Repeat)	18
Figure 9 : Structure schématique montrant la liaison de la protéine S au récepteur ACE2	18
Figure 10 : Processus de liaison et de fusion CoV-2-Cellule hôte assuré par la protéine S	19
Figure 11 : Diagramme de Venn pour la classification physico-chimique des acides aminés	24
Figure 12 : Graphiques représentant les motifs 2D de la protéine S chez la souche originelle (Wuhan) et chez les variants analysés	30
Figure 13 : La protéine S de la souche originelle Wuhan et celle du variant D614G	34

Liste des tableaux

N° et légende de la figure	Page
Tableau 1 : Classification des sept Coronavirus humains et année de leur identification.	5
Tableau 2: Nombre des acides aminés au sein des domaines de la glycoprotéine S de SARS CoV, MERS CoV et le SARS CoV-2	13
Tableau 3 : Mécanismes de dénomination de l'OMS pour le SARS-CoV-2 VOI	21
Tableau 4 : Mécanismes de dénomination de l'OMS pour le SARS-CoV-2 VOC	22
Tableau 5 : Identification des zones transmembranaires hydrophobes chez les variants SARS-CoV-2 et chez le virus référence WuHan	32
Tableau 6 : Principales mutations identifiées entre les variants SARS-CoV-2 contre le virus référence Wuhan	33

INTRODUCTION

Initialement signalé à la mi-décembre 2019 dans la ville chinoise de Wuhan, le nouveau virus du syndrome respiratoire aigu sévère (SRAS-CoV-2) est un bêta-coronavirus à ARN simple brin avec un génome compact de 29 903 nucléotides. Ce virus provoque une maladie grave connue sous le nom de maladie à coronavirus 2019 (COVID-19), qui s'est propagée dans plus de 210 pays en moins de quatre mois, comptant plus de 10 millions de cas confirmés et près de 500 000 décès signalés dans le monde le 28 juin 2020 (OMS, 2021). De son côté, le site Statista¹ affirme un nombre de cas de Covid-19 dans le monde à la date du 17 mai 2021 à 163,7 millions, le nombre de décès dus au Covid-19 dans le monde à la même date à 3,4 millions. En fin, le tableau de bord COVID-19 du *Center for Systems Science and Engineering* (CSSE) de l'Université Johns Hopkins² (JHU), en date du 09 juillet 2021 à 16 :21, rapporte 185,7 millions de cas et 4,013 décès.

L'évolution des nombres de cas n'a pas cessé d'augmenter depuis 2020. Les figures 1, 2 et 3 montrent les variations des nombres de cas de Covid-19 à travers les cinq continents.

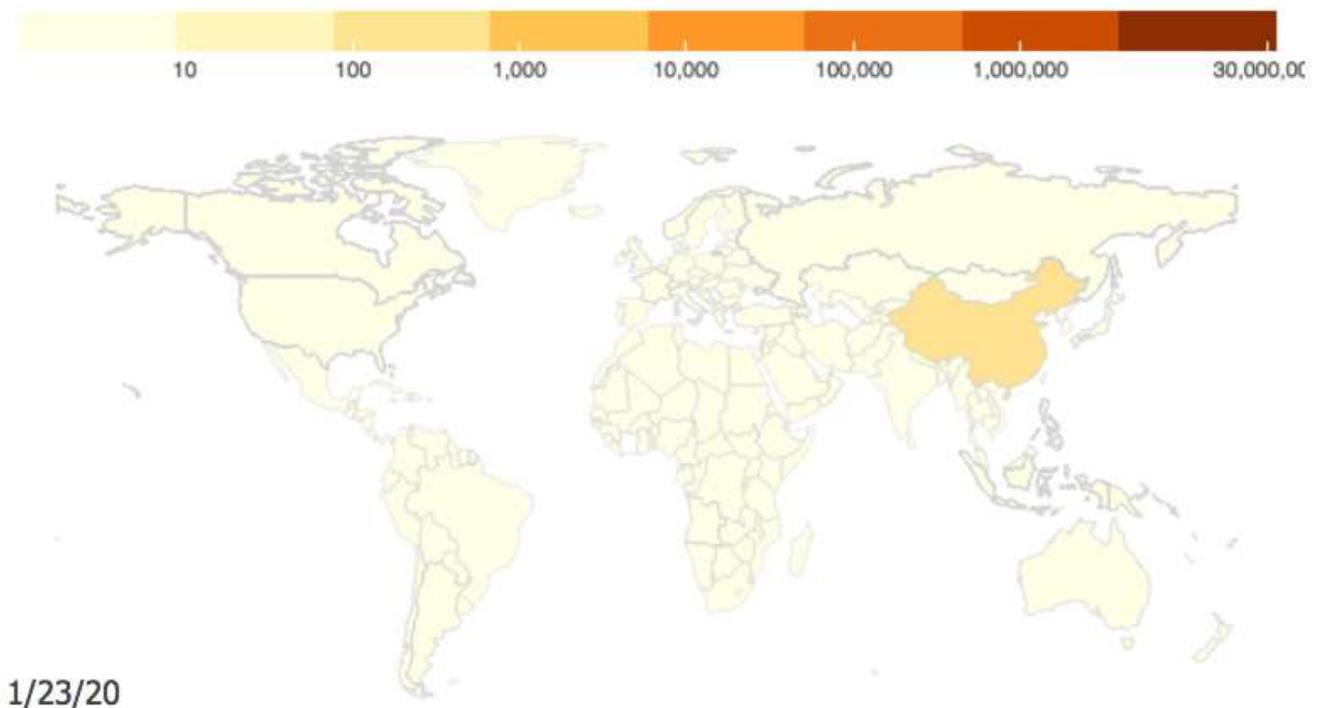


Figure 1 : Estimation des nombres de cas cumulatifs Covid dans le monde en date du 23 janvier 2020

¹ <https://fr.statista.com/themes/6050/le-coronavirus-covid-19/>

² <https://coronavirus.jhu.edu/map.html>

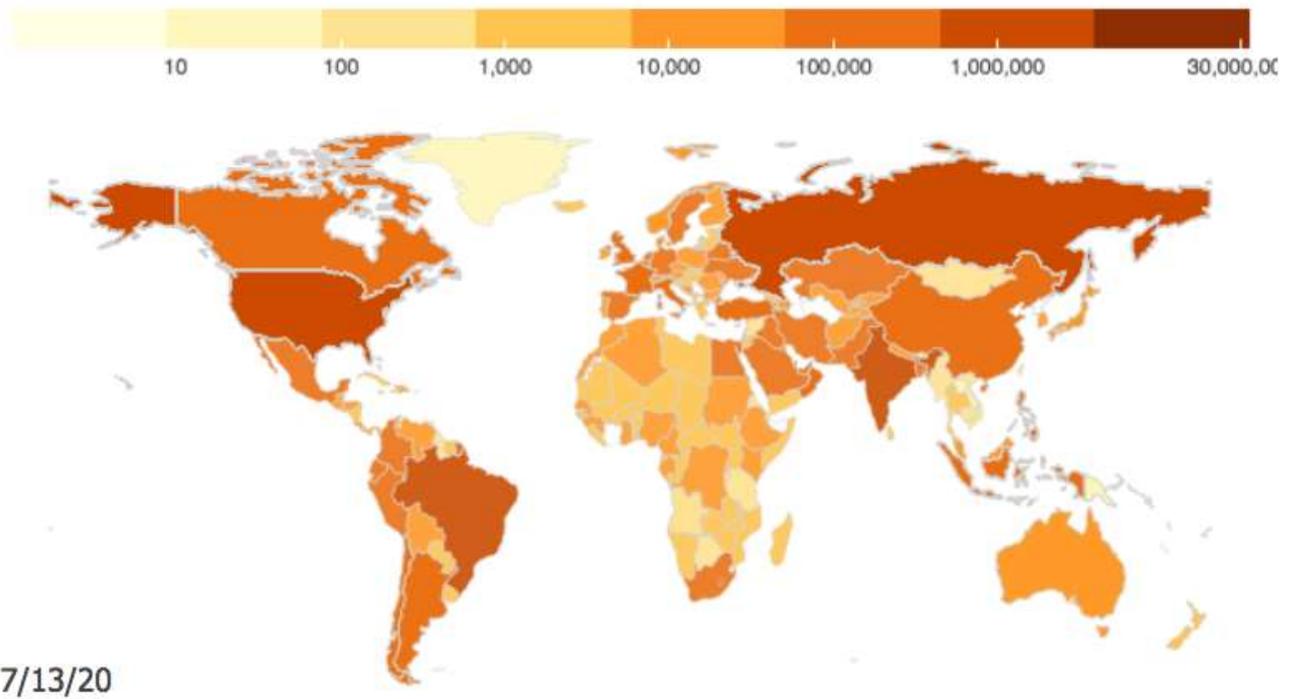


Figure 2 : Estimation des nombres de cas cumulatifs Covid dans le monde en date du 13 juillet 2020

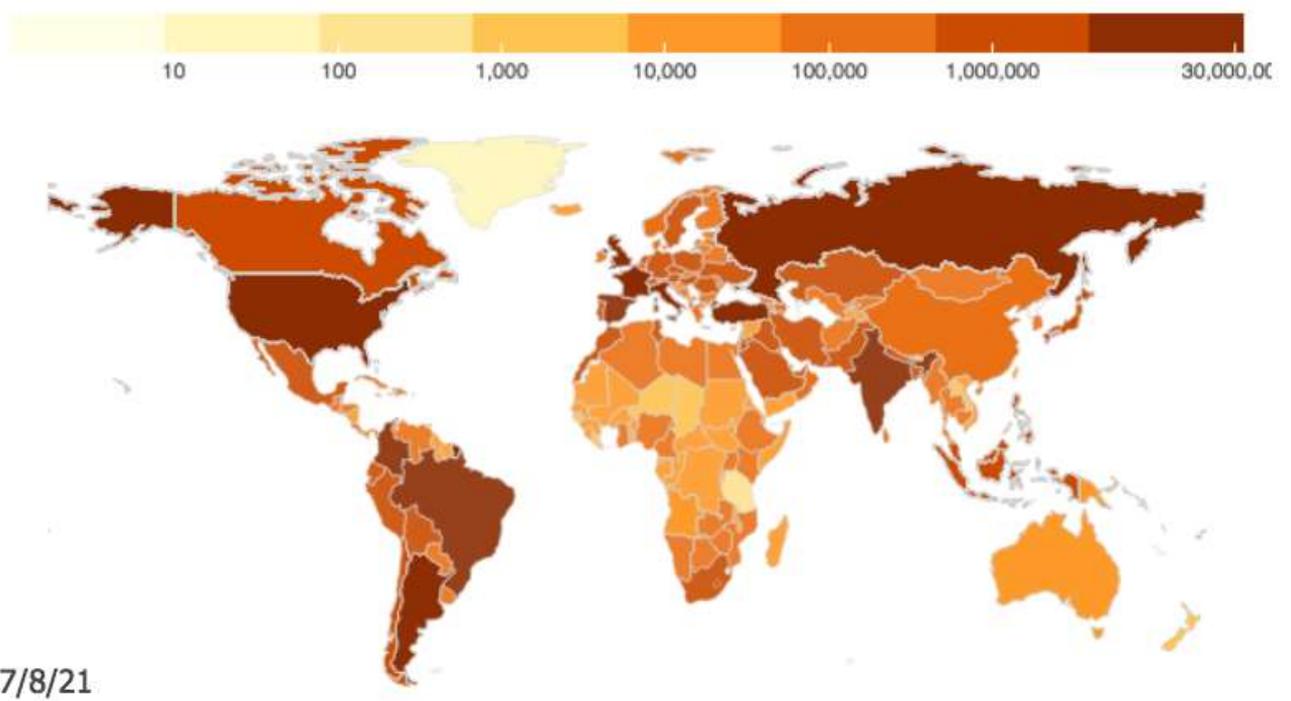


Figure 3 : Estimation des nombres de cas cumulatifs Covid dans le monde en date du 08 juillet 2021

La nouvelle maladie respiratoire COVID-19 a atteint le statut de pandémie mondiale et de grands efforts sont actuellement entrepris pour caractériser moléculairement le SARS-CoV-2. Ces chiffres ont atteints des proportions inquiétantes dans certains pays. Ainsi l'on note, à la même date sus-citée, 606 522 décès aux USA, 530 179 au Brésil et 405 939 en Inde. Dans les pays arabes sont à déplorer 14 643 en Iraq, 9 818 en Jordanie, 7 869 au Liban, et 1 853 en Emirats Arabes. [En Algérie, nous déplorons un nombre de décès de 3 798, et 143 652 cas Covid-19.](#)

La variabilité génomique des isolats du SRAS-CoV-2 dispersés sur toute la planète peut causer des effets étiologiques spécifiques sous-jacents géographiquement.

Dans la présente étude, nous rassemblons les 48 635 génomes complets du SRAS-CoV-2 actuellement disponibles grâce à l'effort de collecte du consortium GISAID et de milliers de laboratoires contributeurs.

L'objectif principal de ce travail est d'analyser et annoter certaines mutations par substitutions apparues dans la glycoprotéine S de surface (Spike) du SRAS-CoV-2 par rapport au génome de référence de Wuhan NC_045512.2, lequel observe une moyenne de 7, 23 mutations par échantillon. Ces substitutions ont été analysées dans les variants les plus répandus dans la planète par région géographique : Royaume-Uni, Inde, Afrique du sud et le Brésil. Au total nous avons analysé cinq protéines Spike avec comme témoin et référence la souche originelle de Wuhan.

PARTIE 1 : SYNTHÈSE BIBLIOGRAPHIQUE & ÉTAT DE L'ART

1- STRUCTURE DU CORONAVIRUS

1-1/ **Taxonomie** : L'histoire des coronavirus humains a commencé en 1965 quand Tyrrell et Bynoe (1966) ont découvert qu'ils pourraient faire passer un virus nommé B814. Il a été trouvé dans la trachée dans des cultures d'organes provenant des voies respiratoires d'un adulte humain présentant un rhume. Au même moment, Hamre et Procknow (1966) ont pu cultiver un virus (229^E) aux propriétés inhabituelles en culture tissulaire à partir des échantillons obtenus d'étudiants en médecine atteints de rhume. Les deux particules, B814 et 229E, étaient éthersensibles et donc vraisemblablement ayant un lipide pour l'infection ; mais ces deux virus n'étaient pas liés à tous les myxo- ou paramyxovirus déjà connus. McIntosh *et al.*, (1967) ont rapporté la récupération de plusieurs souches d'agents sensibles à l'éther du système respiratoire humain en utilisant une technique similaire à celle de Tyrrell et Bynoe. Ces virus ont été appelés «OC» pour désigner que ils ont été cultivés dans des cultures d'organes.

Almeida et Tyrrell (1967) ont effectué une microscopie électronique sur les fluides de cultures d'organes infectés par le B814 et ont trouvé des particules qui ressemblaient au virus de la bronchite infectieuse des poulets. Les particules étaient de taille moyenne (80–150 nm), pléomorphes, revêtues de membrane et recouvertes de saillies de surface largement espacées. L'agent 229E identifié par Hamre et Procknow (1966) et les précédents virus OC identifiés par McIntosh *et al.*, (1967) avaient une morphologie similaire.

À la fin des années 1960, Tyrrell travaillant avec les souches humaines et un certain nombre de virus animaux. Ceux-ci comprenaient le virus de la bronchite infectieuse, le virus de l'hépatite de la souris et le virus de la gastro-entérite transmissible des porcs, tous avaient montré la même morphologie au microscope électronique (McIntosh *K et al.*, 1967 ; Witte *KH et al.*, 1968). Ce nouveau groupe de virus a été nommé coronavirus (corona indiquant l'aspect en forme de couronne avec des projections de surface) et a ensuite été officiellement accepté comme un nouveau genre de virus.

Des recherches utilisant des techniques sérologiques ont abouti à une quantité considérable d'informations concernant l'épidémiologie des coronavirus respiratoires humains. C'était à constater que dans les climats tempérés, les infections à coronavirus survenaient plus souvent en hiver et au printemps comparativement à l'été et à l'automne. Les données également ont révélé que les infections à coronavirus constituent jusqu'à 35% du total des virus respiratoires actifs lors des épidémies. Dans l'ensemble, la proportion d'adultes atteints de rhumes dus aux coronavirus a été estimée à 15%.

Dans les trois décennies qui ont suivi la découverte, les souches humaines OC43 et 229E ont été étudiées, en grande partie parce qu'elles étaient les plus faciles à utiliser. OC43, était étroitement lié au virus de l'hépatite de la souris. Les deux virus (OC43 et 229E) ont démontré une périodicité, avec de grandes épidémies survenant à des intervalles de deux à trois ans (Monto AS, 1974). La souche 229E tend à être épidémique partout aux États-Unis, alors que la souche OC43 était plus prédisposée à des flambées locales. Comme avec de nombreux autres virus respiratoires, la réinfection était courante (Callow KA *et al.*, 1990). L'infection pouvait survenir à tout âge, mais elle était la plus fréquente chez les enfants.

Malgré l'accent mis exclusivement sur souches 229E et OC43, il était clair qu'il y avait d'autres souches de coronavirus également. Comme le montre Bradburne (1970), la souche B814 n'était pas sérologiquement identique à OC43 ou à 229E.

Les coronavirus appartiennent à la famille des coronaviridae qui comprend deux sous familles (les coronavirinae et les torovirinae). les coronavirinae sont divisés en quatre genres appelés Alpha- , Beta- Gamma- et Deltacoronavirus. Le genre Betacoronavirus est lui même subdivisé en quatre clades (A, B, C, D). Les coronavirus humains (HCoV) appartiennent aux Alpha et aux Betacoronavirus (Tableau 1).

Tableau 1 : Classification³ des sept Coronavirus humains et année de leur identification.

Alphacoronavirus	Betacoronavirus
<ul style="list-style-type: none"> ○ HCoV-229E (1966) 	<ul style="list-style-type: none"> ▪ Calde A : <ul style="list-style-type: none"> ○ HCoV-43 (1967) ○ HCoV-HKU1 (2005) ▪ Calde B :

³ (<https://www.sciencedirect.com/science/article/pii/S0515370020302962>)

<ul style="list-style-type: none"> ○ HCoV-NL63 (2004) 	<ul style="list-style-type: none"> ○ SARS-CoV (2003) ○ SARS-CoV-2 (2019) ▪ Clade C : <ul style="list-style-type: none"> ○ MERS-CoV (2012)
--	--

Les analyses phylogénétiques ont montré que les chauves-souris et les rongeurs sont les réservoirs de la majorité des *Alpha-* et *Betacoronavirus*, tandis que les oiseaux sont ceux des *Gamma-* et des *Deltacoronavirus*. Depuis des milliers d'années, les coronavirus ont pu franchir les barrières d'espèces. Il existe un réservoir animal large, notamment chez les mammifères (camélidés, civettes, bovins, etc.) qui peuvent jouer le rôle d'hôtes intermédiaires et d'adaptation à ces virus afin qu'ils infectent ensuite plus efficacement les humains. Ainsi, des coronavirus ont pu émerger comme pathogènes dans l'espèce humaine.

Sept coronavirus sont capables d'infecter l'homme. Quatre sont ubiquitaires et responsables d'infections respiratoires hautes et basses (HCoV), peu sévères en général chez les individus immunocompétents. Deux autres, très pathogènes, ont émergé : le Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) en 2003 et le Middle East Respiratory Syndrome-related coronavirus (MERS-CoV) en 2012.

À la fin de l'année 2019, un nouveau coronavirus, le SARS-CoV-2, responsable de la Covid-19, est apparu en Chine, à Wuhan, puis s'est répandu dans le monde entier. Il s'agit d'un *Betacoronavirus* (clade B) similaire au SARS-CoV (Tableau 1). Toutefois, des analyses phylogénétiques ont montré que le Sars-CoV-2 serait plus proche des SARS-like coronavirus des chauves-souris que des SARS-CoV.

1-2/ **Le Severe Acute Respiratory Syndrome (SARS) :** Ou encore le Syndrome Respiratoire Aigu Sévère : De nombreux coronavirus, découverts pour la première fois dans les volailles domestiques dans les années 30, provoquent des maladies respiratoires, gastro-intestinales, hépatiques et neurologiques chez les animaux. Seuls sept coronavirus sont connus pour provoquer des maladies chez l'homme.

Quatre des sept coronavirus provoquent le plus souvent des symptômes de rhume⁴ : Les virus 229E, OC43, NL63 et HKU1 sont responsables d'environ 15 à 30% des cas de rhume banal. Rarement, des infections graves des voies respiratoires inférieures, dont les bronchiolites, des pneumonies, peuvent survenir principalement chez les nourrissons, les personnes âgées et les sujets immunodéprimés.

Trois des sept coronavirus causent des infections respiratoires beaucoup plus graves et parfois mortelles chez l'homme que d'autres coronavirus et ont provoqué des épidémies majeures de pneumonie mortelle au 21^{ème} siècle :

- Le SARS CoV-2 est un nouveau coronavirus identifié comme la cause de la maladie à coronavirus de 2019 (COVID-19) qui a débuté à Wuhan, en Chine en fin de 2019 et s'est propagé dans le monde entier.
- Le MERS CoV a été identifié en 2012 comme la cause du syndrome respiratoire du Moyen-Orient (MERS : Middle East Respiratory Syndrome).
- Le SARS CoV a été identifié en 2003 comme la cause d'une épidémie de syndrome respiratoire aigu sévère (SARS) qui a débuté en Chine vers la fin de 2002.

Ces coronavirus qui causent des infections respiratoires sévères sont des agents pathogènes zoonotiques, qui débutent chez les animaux infectés et sont transmis à l'homme.

2- GÉNOMIQUE ET PROTÉOMIQUE DU SARS-COV-2

Les principales plateformes (OMS, 2021) employées pour le séquençage du génome du SARS-CoV-2 et l'analyse de ses séquences sont :

- i. Sanger Sequencing : Largement disponible et d'utilisation facile avec un bon rapport coût-efficacité, mais un débit très faible (100 Kb – 2 Mb/run) avec des amplicons de taille inférieure à 1000 pb.

⁴ <https://www.msmanuals.com/fr/professional/maladies-infectieuses/virus-respiratoires/coronavirus-et-syndrome-respiratoire-aigu-s-C3%A9v-C3%A8re-covid-19-mers-et-sras>

- ii. Illumina (iSeq, MiniSeq, MiSeq, NextSeq, HiSeq, NovaSeq) : Un degré d'exactitude et un rendement très élevés. Longueur de lecture maximale de 2 x 300 pb et un débit de séquençage de 1,é – 6000 Gb.
- iii. Ion Torrent : Exécution rapide du séquençage avec une longueur de lecture maximale proche de 400 pb. Un débit de séquençage de 30 Mb – 50 Gb, selon les puces et les dispositifs.
- iv. Oxford Nanopore Technologie (Flongle, MinION, GridION, PromethION) : Données en temps réels avec un taux d'erreur par lecture proche de 5%. Le débit varie de <2 Gb pour les flow cells Flongle et 220 Gb pour les flow cells PromethION.

Le génome des coronavirus est une molécule d'ARN linéaire (Figure 1), non segmentée, directement infectieuse. Sa caractéristique principale est sa taille qui est de 27 à 31000 nucléotides. Il s'agit du plus grand ARN viral connu. L'organisation génomique est conservée parmi toutes les espèces de coronavirus. Les deux premiers tiers du génome, soit environ 20000 nucléotides, sont constitués de deux cadres de lecture ORF (Open Reading Frames) 1a et 1b chevauchant et codant deux précurseurs protéiques d'une taille et d'une complexité sans précédent. Ces précurseurs sont clivés en 15 à 16 fragments qui forment le complexe de réplication. En aval, on trouve les quatre à cinq gènes codant les protéines structurales, dans un ordre précis : HE, S, E, M, N (protéine hémagglutinine estérase [HE], protéine de pointe [S], enveloppe [E], glycoprotéine membranaire [M] et nucléocapside [N]). Le génome des coronavirus comprend aussi des gènes codant des protéines non structurales dont la fonction n'est pas encore connue.

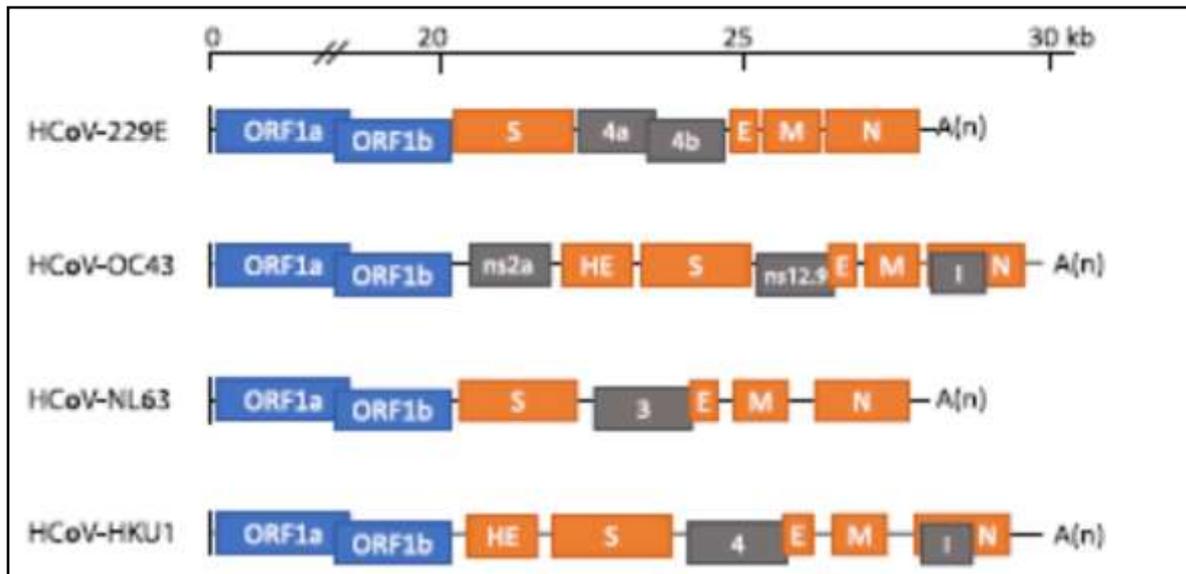


Figure 4 : Structure du génome des coronavirus humains (HCoV) - HCoV-229E, HCoV-OC43, HCoV-NL63 et HCoV-HKU1.

Sur cette figure, sont illustrés les cadres de lecture ouverts (Open Reading Frames : ORF) 1a (ORF1a) et 1b (ORF1b) qui sont représentés par des cases bleu marine qui chevauchent entre-elles. Les gènes codant pour les protéines structurales (S), l'enveloppe (E), la membrane (M), la nucléocapside (N) et l'hémagglutinine-estérase (HE) sont représentés par des cadres oranges. Les gènes codant pour les protéines accessoires sont représentés par des cases gris foncé.

L'ensemble du génome du SRAS-CoV-2 code pour environ 7096 résidus qui composent de nombreuses protéines structurales et non structurales.

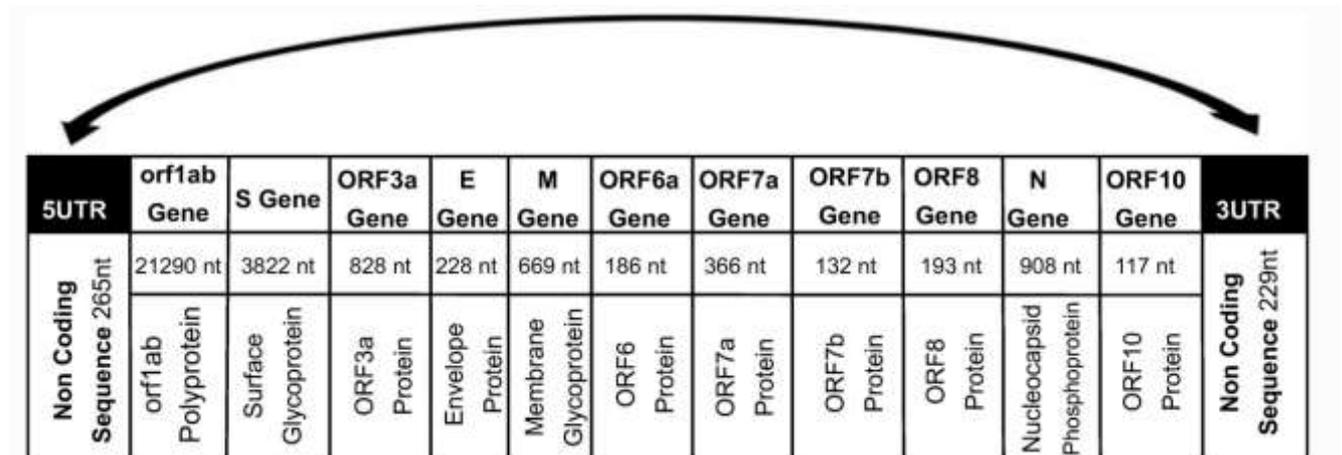


Figure 5 : Structure du génome de SARS-CoV-2 (Nawaz *et al.*, 2021).

La réplication des coronavirus dans les cellules eucaryotes est entièrement intra cytoplasmique, elle fait appel à une stratégie particulière aboutissant à la synthèse discontinue d'ARN subgénomiques de taille décroissante, ayant tous la même extrémité 3'. La taille du génome et la complexité du mode de réplication ont longtemps été un obstacle à l'étude des coronavirus. L'importante plasticité du génome des coronavirus fait de ces virus des agents à fort potentiel évolutif.

Les deux modes d'évolution majeurs des coronavirus sont les mutations et les recombinaisons. L'ARN génomique est répliqué par une ARN polymérase ARN dépendante (RdRp), dépourvue de système de correction d'erreur, et générant de nombreux mutants. Comme pour tous les virus à ARN, la population virale est hétérogène et possède une distribution en quasi-espèces (sous-populations différentes d'un même virus infestant un même hôte). Cette distribution peut être vue comme une stratégie d'optimisation, en étant une structure permettant de disposer d'un réservoir de variants possédant les capacités de faire face aux changements environnementaux. Elle a été décrite pour plusieurs coronavirus non seulement dans le cadre d'infections persistantes, mais aussi d'infections aiguës (Adami C *et al.*, 1995 ; Vabret A *et al.*, 2006).

La très grande taille du génome permet l'émergence des variants présentant de larges délétions, et permet l'utilisation de ces virus comme vecteurs viraux. L'exemple le plus connu

est l'émergence du coronavirus porcin respiratoire ou PRCV, dans les années 1980. Le PRCV est un variant spontané du coronavirus porcin entérique ou TGEV. Il présente une délétion en phase de 672 nucléotides (224 acides aminés) dans le gène codant la protéine S1. Une des conséquences biologiques de cette grande délétion est le changement de tropisme du virus qui, d'entérique pour le TGEV, est devenu respiratoire pour le PRCV (Laude HDB et Eleouet JF, 1998). Soulignons que, dans ce cas, l'émergence d'un variant a été bénéfique pour l'hôte. L'infection respiratoire à PRCV est peu symptomatique et bénigne, sa transmission respiratoire est facile et efficace, et elle permet l'acquisition d'une immunité croisée avec le TGEV, maladie entérique d'évolution grave chez les porcelets.

L'autre mode évolutif des coronavirus est la recombinaison génétique. Ce phénomène est fréquent chez les virus à ARN positif (ARN qui sert tel quel de messenger), et semble favorisé chez les coronavirus, par le mode discontinu de la transcription. La recombinaison correspond à l'échange de matériel génétique, elle peut être homologue si elle a lieu entre deux génomes de coronavirus, ou hétérologue si elle intéresse d'autres gènes viraux ou cellulaires. De nombreuses formes recombinantes ont été décrites *in vitro* et *in vivo* chez les coronavirus. Par exemple, Herrewegh *et al.* (1998) ont montré que les coronavirus félines de type II sont le résultat d'une double recombinaison entre coronavirus félines de type I et coronavirus canins, faisant suite à un franchissement de barrière d'espèces dans le sens chien chat. L'importante capacité évolutive des coronavirus a été mise en lumière par l'émergence du SARS-CoV, dont nous reparlerons plus loin, et a relancé les études chez les autres membres de cette famille.

La glycoprotéine S est une protéine structurelle qui explique la forme de couronne des CoV. La protéine S de longueur 1200 AA (environ 180 KDa) fait partie des protéines de fusion virale de classe I (Lu G *et al.*, 2015 ; Millet JK *et al.*, 2014). Il est exprimé à la surface du virus en tant que protéine trimérique et détermine également le tropisme et la pathogénèse de l'hôte (Li F. *et al.*, 2016 ; Belouzard S. *et al.*, 2012). La glycoprotéine S peut être clivée par les protéases de l'hôte en un ectodomaine S1 N-terminal et une membrane S2 C-terminale ancrée protéine (figure 2) (Wrapp D. *et al.*, 2020). Dans plusieurs CoV, l'amplification plus élevée de la protéine S au niveau de la membrane cellulaire peut également faciliter la fusion membranaire des cellules infectées avec des cellules adjacentes non infectées, ce qui conduit à la formation de cellules géantes ou multi nucléées et propagation ultérieure du virus entre les cellules (Fehr A. R. *et al.*, 2015 ; Qian Z. *et al.*, 2013). Les CoV -S1 la protéine comprend un

domaine de liaison au récepteur (RBD), qui se lie à la cellule hôte via des récepteurs qui incluent le Récepteur ACE2 (Li W *et al.*, 2003). SARS CoV utilise le récepteur ACE2 pour l'entrée (Li W *et al.*, 2003) et utilise la sérine protéase, les cathepsines et TMPRSS2, pour l'amorçage de la protéine S (Glowacka I *et al.*, 2011 ; Matsuyama S *et al.*, 2010). Sur 14 résidus AA dans la sous-unité S1 du SARS CoV, huit résidus sont hautement conservés dans le SARS CoV-2, ce qui suggère que le récepteur ACE2 est également utilisé pour l'entrée cellulaire par ce virus (Wan Y *et al.*, 2020). Les études d'étiquetage l'ont confirmé (Xu X. *et al.*, 2020 ; Gralinski L et Menachery VD, 2020). La S-glycoprotéine du SARS CoV-2 a une structure 3D dans le RBD domaine qui conserve les interactions de van der Waals (Xu X. *et al.*, 2020). Les 394 résidus de glutamine de la RBD du SARS CoV-2 est ligaturé via l'important résidu lysine 31 du récepteur humain, ACE2 (Wan Y. *et al.*, 2020).

Une protéine E est impliquée dans l'assemblage et la libération de particules de type virus (Mortola E et Roy P, 2004 ; Corse E. et Machamer C. E., 2003). La protéine M donne la forme de l'enveloppe virale et se lie à la nucléocapside (Neuman B. W. *et al.*, 2011). Tandis que la protéine N se lie au génome de l'ARN CoV, construisant la nucléocapside pour soutenir le complexe de réplication-transcription (de Haan C. A. et Rottier P. J., 2005). L'interaction entre les protéines M et les protéines N se stabilise la nucléocapside et le noyau interne du virus et, éventuellement, améliorent l'assemblage des particules de virion (Fehr A. R. et Perlman S, 2015 ; Escors D. *et al.*, 2001 ; Narayanan K *et al.*, 2000). Les protéines M, E et S pénètrent ensuite dans le complexe du compartiment intermédiaire réticulum endoplasmique-Golgi et constituent l'enveloppe virale (Narayanan K. *et al.*, 2000). Les protéines accessoires semblent favoriser l'adaptation des CoV à l'hôte humain cellules (Cui J. *et al.*, 2019).

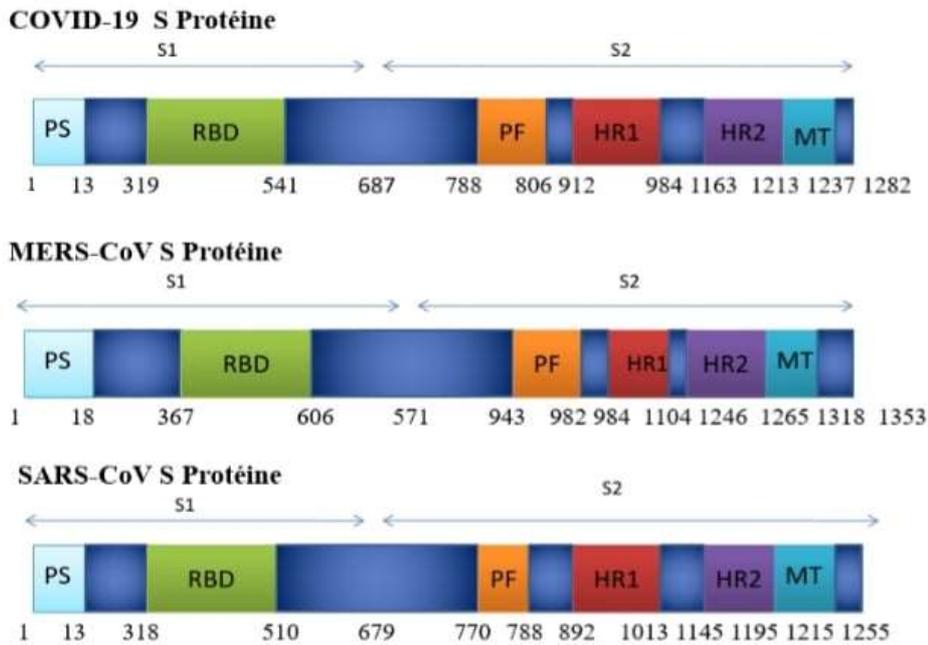


Figure 6 : Différentes structures de la protéine S au sein de différents Coronavirus.

L'analyse génomique de dix génomes du SRAS-CoV-2 isolés de neuf patients a démontré 99,98% d'identité nucléotidique (Lu R. *et al.*, 2020). Un autre rapport a révélé 99,8 à 99,9% de similitude de séquence dans un échantillon de cinq patients infectés (Calisher C *et al.*, 2020). L'analyse phylogénétique démontre que le SRAS-CoV-2 partage 50 à 80,0% d'identité nucléotidique avec le MERS-CoV et SRAS-CoV respectivement (Zhu N. *et al.*, 2020 ; Zhou P. *et al.*, 2020 ; Lu R. *et al.*, 2020). Le SRAS-CoV-2 constitue un clade parmi le sous-genre sarbecovirus (Zhu N. *et al.*, 2020). L'Analyse bioinformatique du génome viral d'un patient COVID-19 partageait une similarité de 89% et de 82 % avec le CoV-ZXC21 de type SARS de chauve-souris et SRAS-CoV humain, respectivement (Chan JF-W *et al.*, 2020). Cependant, la sous-unité externe de Spike RBD de SARS-CoV-2 n'a que 40% identité des acides aminés avec d'autres CoV associés au SARS (Chan JF-W *et al.*, 2020). La protéine S du SRAS-CoV-2 est plus longue (1282 AA) que pour d'autres virus tels que le SARS-CoV (1255 AA) et les CoV de type chauve souris de type SARS (1246 résidus). La S-glycoprotéine du SARS-CoV-2 a été trouvée pour avoir trois insertions courtes à l'extrémité N-terminale, avec quatre variations dans le site de liaison au récepteur dans le RBD par rapport au SARS-CoV (Zhou P *et al.*, 2020). De plus, sa nouvelle séquence ORF8 peut éventuellement coder pour une protéine sécrétée avec une structure α -hélicoïdale. Le tableau

2 résume le nombre des acides aminés de la protéine S de SARS CoV, MERS CoV et SARS CoV-2.

Tableau 2: Nombre des acides aminés au sein des domaines de la glycoprotéine S de SARS CoV, MERS CoV et le SARS CoV-2⁵

	PS	RBD	PF	HR1	HR2	TM
SARS CoV-2	13	293	19	73	51	25
MERS CoV	18	240	40	121	50	24
SARS CoV	13	193	19	122	51	21

La glycoprotéine de pointe est l'une des cibles majeures à explorer en raison de son rôle lors de l'entrée des coronavirus dans les cellules hôtes. Ils ont analysés 320 séquences du génome entier et 320 séquences de protéines de pointe de SARS CoV-2 en utilisant un alignement de séquences multiples. Dans cette étude, 483 variations uniques ont été identifiées parmi les génomes du SARS CoV- 2 y compris 25 mutations non synonymes et une délétion dans la protéine de pointe (S).

Parmi les 26 variations détectées dans la protéine S, 12 variations étaient localisées au domaine N-terminal (NTD) et 6 variations au domaine de liaison au récepteur (RBD) qui pourraient Modifier l'interaction de la protéine S avec l'enzyme de conversion de l'angiotensine du récepteur hôte 2 (ACE2).

En outre, 22 insertions d'acides aminés ont été identifiées dans la protéine de pointe du SARS CoV-2 en comparaison avec celle du SARS CoV-2. Les analyses phylogénétiques de la protéine de pointe ont révélé que le coronavirus de la chauve-souris a une relation évolutive étroite avec le SARS CoV-2 en circulation.

Les données d'analyse de la variation génétique peuvent aider à une meilleure compréhension de la pathogénèse du SARS CoV-2. Sur la base des résultats rapportés, des inhibiteurs potentiels contre la protéine S, peuvent être conçus en considérant ces variations et leur impact sur la structure des protéines.

⁵ (<https://www.futuremedicine.com/doi/full/10.2217/fv1-2020-0129>)

Les variations et les mutations dans la protéine S de SARS CoV-2 peuvent être résumées dans les points suivants⁶ :

- i. Il existe 483 variations uniques dans les différents génomes du SARS CoV-2.
- ii. Dans la seule glycoprotéine SARS CoV-2 S, il existe 40 variations.
- iii. Quatorze mutations synonymes (incorporation du même résidu) et 25 mutations non synonymes (incorporation de résidus différents) ont été observées dans la glycoprotéine S.
- iv. Douze mutations non synonymes sont localisées dans le domaine N-terminal de la glycoprotéine S.
- v. Six mutations non synonymes existent au niveau du domaine de liaison au récepteur (RBD).

3- STRECTURE DE LA PROTEINE SPIKE

D'une taille de 180 à 200 kDa, la protéine Spike se compose d'un domaine N-terminal extracellulaire, d'un domaine transmembranaire ancré dans la membrane virale et d'un court segment C-terminal intracellulaire (Bosch *et al.*, 2003). C'est une protéine transmembranaire de type I qui est présente au sein de l'enveloppe sous forme d'un trimère homogène. Chaque monomère se compose de deux sous-unités fonctionnelles, S1 et S2. La sous-unité S1 contient un domaine RBD (Receptor Binding Domain) qui se fixe au récepteur ACE2 sur la membrane de la cellule pulmonaire hôte. La sous-unité S2 est activée par la liaison de S1 avec ACE2 et contribue à la fusion membranaire. La longueur totale moyenne de la protéine S est de 1273 résidus (figure 3) et se compose d'un peptide signal de 13 résidus situés à l'extrémité N-terminale de la sous-unité S1 (14–685 résidus) et de la sous-unité S2 (686– 1273 résidus). Les deux dernières régions sont respectivement responsables de la liaison au récepteur et de la fusion membranaire (Watanabe Y. *et al.*, 2020).

⁶ <https://www.sciencedirect.com/science/article/abs/pii/S1567134820302203>

Dans la sous-unité S1, il y a un domaine N-terminal (14 à 305 résidus) et un domaine de liaison au récepteur RBD (319 à 541 résidus). La sous-unité S2 comporte le peptide de fusion (788–806 résidus), la séquence de répétition heptapeptidique 1 (HR1) (912–984 résidus), HR2 (1163–1213 résidus), le domaine TM (1213–1237 résidus) et le domaine du cytoplasme (1237–1273 résidus) (Xia S. *et al.*, 2020).

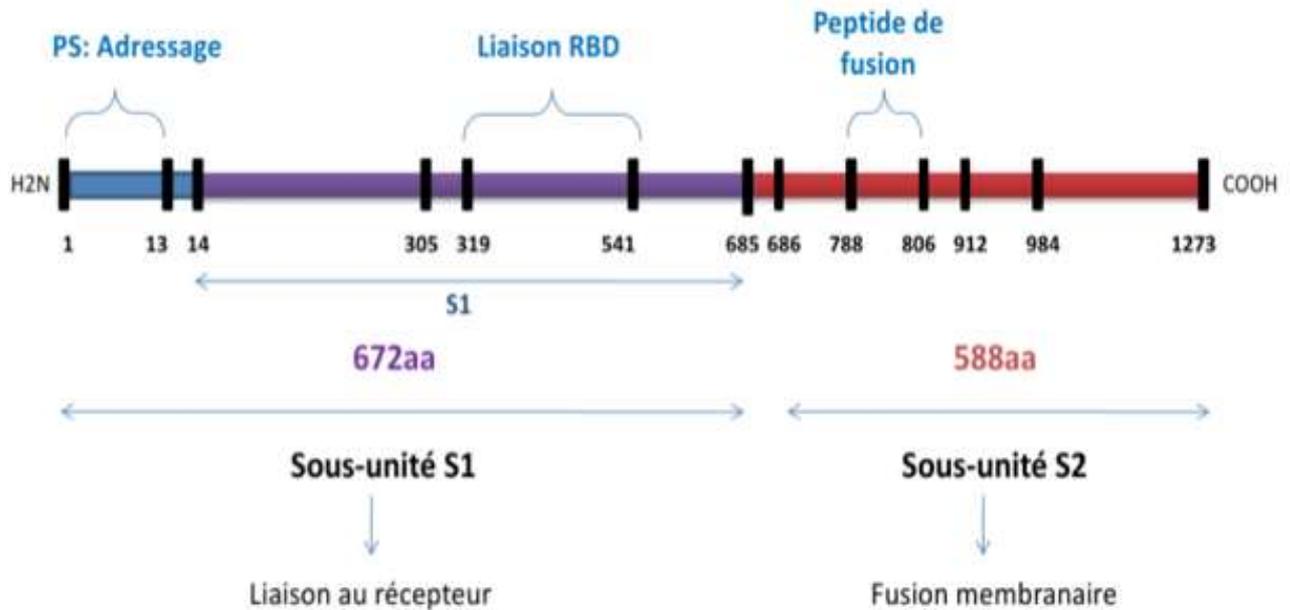


Figure 7 : Structure schématique des deux sous- unités S1 et S2 de la protéine Spike.

A l'état natif, la protéine S existe en tant que précurseur inactif. Lors d'une infection virale, les protéases des cellules cibles activent la protéine S en la coupant en sous-unités S1 et S2 (Bertram S *et al.*, 2013) ; ce qui est nécessaire pour activer le domaine de fusion membranaire après l'entrée virale dans les cellules cibles (Hoffmann M. *et al.*, 2020). Semblable à d'autres coronavirus, la protéine S est clivée en sous-unités S1 et S2 par des protéases cellulaires, et la sérine protéase TMPRSS2 est utilisée comme amorce protéique. Bien que le site de clivage de S du SARS-CoV soit connu, celui de S du SARS-CoV-2 n'a pas encore été rapporté (Hoffmann M *et al.*, 2020) (Du L *et al.*, 2007). (Les polyprotéines de réplique SARS-CoV de résidus d'acide aminés 4382 et 7073 devraient être clivées en 16 sous-unités par deux protéinases virales).

3-1/ Structure de la sous-unité S1 : La liaison des particules virales aux récepteurs cellulaires à la surface de la cellule hôte est l'initiation de l'infection virale ; par conséquent, la

reconnaissance des récepteurs est un déterminant important de l'entrée virale et une cible de conception de médicaments.

Le RBD situé dans la sous-unité S1 se lie au récepteur cellulaire ACE2 dans la région de l'aminopeptidase N. La région S1 contient le NTD et le CTD, et les détails atomiques à l'interface de liaison démontrent des substitutions de résidus clés dans SARS-CoV-2-CTD. En outre, l'interface de liaison la protéine S CTD a plus de résidus qui interagissent directement avec le récepteur ACE2 que SARS-RBD (21 contre 17), et une plus grande surface est enterrée avec la protéine S CTD dans complexe avec ACE2 qu'avec la protéine S RBD. Les mutations des résidus clés jouent un rôle important dans l'amélioration de l'interaction avec ACE2. F486 dans SARS-CoV-2, au lieu de I472 dans SARS RBD, forme de fortes interactions aromatiques-aromatiques avec ACE2 Y83 et E484 dans SARS-CoV-2-CTD, au lieu de P470 dans SARS RBD, forme des interactions ioniques avec K31, qui conduit à une affinité plus élevée pour la liaison au récepteur que la RBD du SRAS-CoV (Wrapp D *et al.*, 2020) (Walls A. C. *et al.*, 2020) (Wang Q. *et al.*, 2020) (Lan J. *et al.*, 2020).

La région RBD est une cible critique pour les anticorps neutralisants (nAbs), et SARS-CoV-2 et SARS-CoV RBD ont une séquence similaire d'environ 73% à 76%. Neuf résidus en contact avec ACE2 dans le CoV RBD sont entièrement conservés, et quatre sont partiellement conservés. L'analyse du RBM (motif de liaison au récepteur, une partie de RBD établissant des contacts directs avec ACE2) du SARS-CoV et du SARS-CoV-2 a révélé que la plupart des résidus essentiels à la liaison de l'ACE2 dans la protéine la protéine S sont conservés dans la Protéine S. Cependant, certaines études ont montré que les anticorps monoclonaux murins (mAbs) et les anticorps polyclonaux contre le SRAS-RBD sont incapables d'interagir avec la protéine SARS-CoV-2 S, révélant des différences d'antigénicité entre le SRAS-CoV et le SRAS-CoV-2 (Wang Q *et al.*, 2020). De même, un anticorps spécifique du SRAS-CoV RBD n'a pas réussi à bloquer l'infection médiée par la protéine S de SL-CoV-SHC014 (Xia S *et al.*, 2019).ce qui suggère que le S1 RBD peut ne pas être une cible médicamenteuse idéale en raison de la caractéristique hautement mutable de large -médicaments anti-CoV à spectre.

3-2/ Structure de la sous-unité S2 : La sous-unité S2, composée successivement des domaines FP, HR1, HR2, TM et d'une fusion de domaine cytoplasmique (CT), est responsable de la fusion et de l'entrée du matériel génétique (ARN) du virus à l'intérieur de la cellule hôte.

FP est un court segment de 15 à 20 résidus conservés de la famille virale. Il est principalement composé de résidus hydrophobes, tels que la glycine (G) ou l'alanine (A), qui s'ancrent à la membrane cible lorsque la protéine S adopte la conformation de la pré-épine. Des recherches antérieures ont montré que la FP joue un rôle essentiel dans la médiation de la fusion membranaire en perturbant et en reliant les bicouches lipidiques de la membrane de la cellule hôte (Millet J. K. et Whittaker G. R., 2018).

HR1 et HR2 sont composés d'un heptapeptide répétitif: HPPHCPC, où H est un résidu hydrophobe ou traditionnellement volumineux, P est un résidu polaire ou hydrophile et C est un autre résidu chargé (Chambers P. *et al.*, 1990). HR1 et HR2 forment le faisceau à six hélices (6-HB), qui est essentiel pour la fusion virale et la fonction d'entrée de la sous-unité S2 (Xia *Set al.*, 2020). HR1 est situé à l'extrémité C-terminale d'un FP hydrophobe, et HR2 est situé à l'extrémité N-terminale du domaine TM (Robson B, 2020). Le domaine TM en aval ancre la protéine S à la membrane virale et la sous-unité S2 se termine par une queue CT (Tang T. *et al.*, 2020).

RBD se lie à ACE2 et S2 change de conformation en insérant FP dans la membrane de la cellule cible, en exposant la bobine enroulée de la préhairpin du domaine HR1 et en déclenchant l'interaction entre le domaine HR2 et le trimère HR1 pour former 6-HB, apportant ainsi l'enveloppe virale et membrane cellulaire à proximité pour la fusion et l'entrée virales (Xia S. *et al.*, 2018). HR1 forme un assemblage homotrimérique dans lequel trois rainures hydrophobes hautement conservées sur la surface qui se lient à HR2 sont exposées. Le domaine HR2 forme à la fois une hélice rigide et une boucle flexible pour interagir avec le domaine HR1. Dans la conformation en épingle à cheveux postfusion des CoV, il existe de nombreuses interactions fortes entre les domaines HR1 et HR2 à l'intérieur de la région hélicoïdale, qui est appelée «région centrale de fusion» (régions HR1core et HR2core, respectivement).

Le ciblage de la répétition de l'heptade (HR) a suscité le plus grand intérêt dans la découverte de médicaments thérapeutiques. La protéine S est une protéine cible importante pour le développement de médicaments spécifiques, tandis que le domaine S1 RBD fait partie d'une région hautement mutable et n'est pas un site cible idéal pour le développement d'inhibiteurs antiviraux à large spectre (Lu G. *et al.*, 2015). En revanche, la région HR de la sous-unité S2 joue un rôle essentiel dans les infections à HCoV et est conservée parmi les

HCoV, tout comme le mode d'interaction entre HR1 et HR2 (Liu S. *et al.*, 2004). Il a été démontré en 2017 qu'un peptide synthétique dérivé de la région de la tige de la protéine d'enveloppe du ZIKV (Zika Virus, Flavivirus) inhibe fortement l'infection par le ZIKV et d'autres flavivirus *in vitro* (Yu Y. *et al.*, 2017) impliquant une efficacité antivirale de peptides dérivés de régions conservées de protéines virales. Les peptides dérivés de la région HR2 des protéines de fusion virale de classe I des virus enveloppés se lient de manière compétitive à HR1 viral et inhibent efficacement l'infection virale (Xia S *et al.*, 2019). Par conséquent, HR1 est une cible prometteuse pour le développement d'inhibiteurs de fusion contre l'infection par le SRAS-CoV-2.

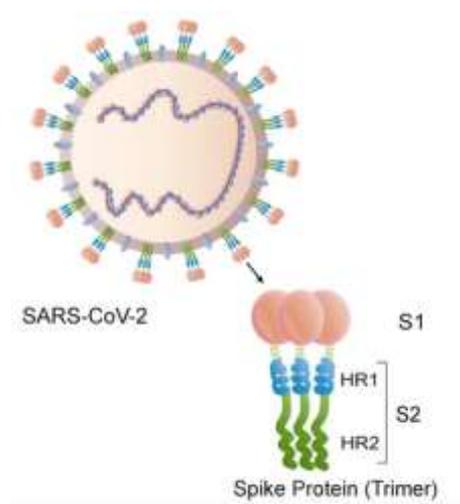
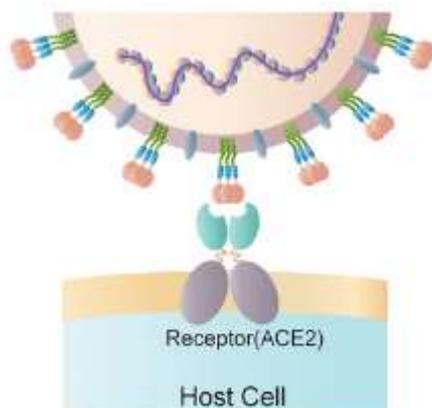


Figure 8⁷ : Structure schématique d'un trimère de la protéine S montrant les domaines HR1 et HR2 (Heptad Repeat).



⁷ <https://www.nature.com/articles/s41401-020-0485-4/figures/1>

Figure 9 : Structure schématique montrant la liaison de la protéine S au récepteur ACE2.

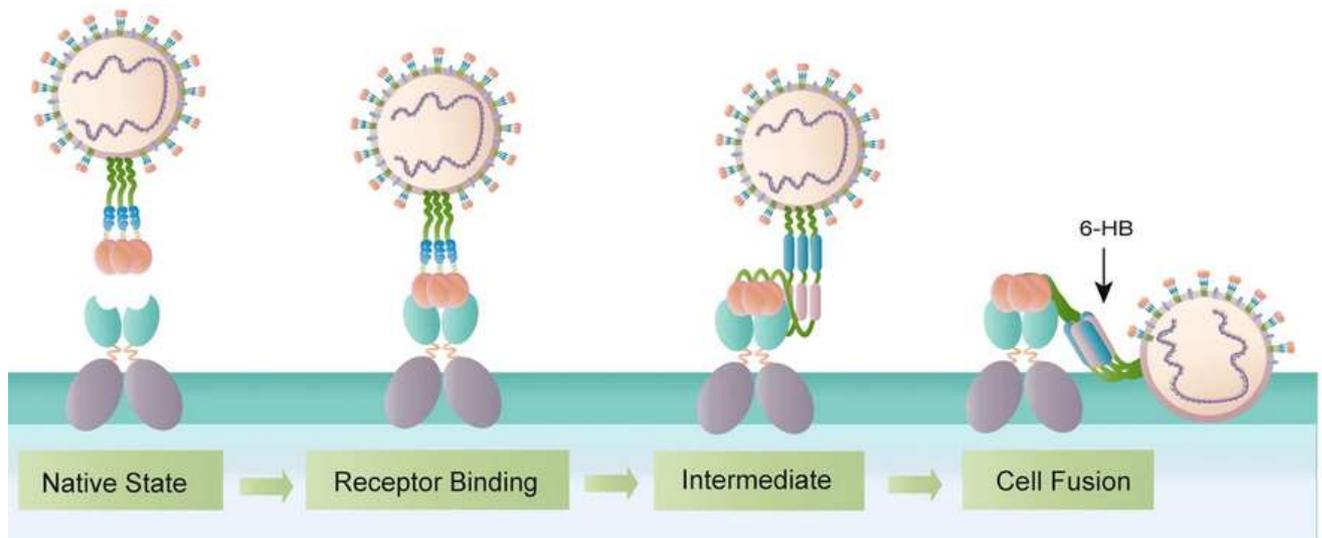


Figure 10 : Processus de liaison et de fusion CoV-2-Cellule hôte assuré par la protéine S.

4- LES VARIANTS MAJEURS DE SARS-COV-2

La notion de variant ou mutant est due à la présence d'une (ou plusieurs) mutation dans le génome du virus et qui sera transmise dans les générations virale suivantes en provoquant une diffusion épidémiologique avérée mais sans incidences sur les propriétés pathologiques du virus.

Toutes les mutations observées officiellement par les organismes de santé publique et par les laboratoires de séquençage (réseau des laboratoires de référence COVID-19 de l'OMS), au sein des génomes des virus CoV-2 sont soit des substitutions soit des délétions. Les mutations par substitutions étant plus fréquentes naturellement. Ces mutations ont concerné certaines régions de la protéine Spike et ont donc créé des mutants du CoV-2. De plus, l'on note que ces mutants ont émergé, pour la première fois, dans certains pays ; la raison pour laquelle ils sont nommés par les noms de ces pays (variant britannique, variant sud-africain, etc.) pour leur première apparition.

L'OMS a proposé des définitions pour les mutants du SARS CoV-2. Des variants d'intérêt (VOI : variant of interest), des variants préoccupants (VOC : variant of concern) et des variants à haut risque (Variant of high consequence). Dès que les variants sont désignés comme VOI et VOC selon les définitions de l'OMS, des mécanismes d'étiquetage seront choisis et appliqués. Lorsque ces définitions sont révisées, le mécanisme d'étiquetage suivra les nouvelles définitions. Les étiquettes seront sélectionnées par l'OMS en utilisant des noms de lettres de l'alphabet grec : alpha, bêta, gamma, etc.

Les étiquettes pour les VOI et les VOC, leurs liens avec les systèmes existants de classification et de nomenclature du SRAS-CoV-2 basés sur la phylogénétique, et leurs principales caractéristiques scientifiques et médicales seront publiés et mis à jour en permanence par l'OMS à l'adresse <https://www.who.int/activities/tracking-SARS-CoV-2-variants>.

Trois systèmes de nomenclature pour nommer et suivre les lignées génétiques du SRAS-CoV-2 sont actuellement utilisés :

- i. La Global Initiative on Sharing All Influenza Data (GISAID),
- ii. Nextstrain et
- iii. Pango.

Chaque système a sa propre approche scientifique pour classer et nommer les lignées virales, et les trois systèmes ont été introduits avant que les VOI et les VOC ne soient reconnus. L'existence de ces différents systèmes de nomenclature peut signifier qu'un même variant porte plusieurs noms, souvent en même temps et sans tenir compte des propriétés des VOI et des VOC.

4-1/ Les VOI : sont des variants avec des marqueurs génétiques spécifiques qui ont été associés à des modifications de la liaison aux récepteurs, à une neutralisation réduite par des anticorps générés contre une infection ou une vaccination antérieure, une efficacité réduite des traitements, un impact diagnostique potentiel ou une augmentation prévue de la transmissibilité ou de la gravité de la maladie. Un variant est dit VOI (à suivre) si, par rapport à un isolat de référence (WuHan par exemple), son génome présente des mutations ayant des implications phénotypiques établies ou présumées, et :

- i. S'il est à l'origine d'une transmission communautaire/de plusieurs cas/d'un groupe de cas de COVID-19, ou qu'il a été détecté dans plusieurs pays ; OU
- ii. S'il est considéré, pour d'autres raisons, comme un variant à suivre par l'OMS, en consultation avec le Groupe de travail de l'OMS sur l'évolution du virus SARS-CoV-2.

Un variant d'intérêt peut nécessiter une ou plusieurs actions de santé publique appropriées, y compris une surveillance accrue des séquences, une caractérisation améliorée en laboratoire ou des enquêtes épidémiologiques pour évaluer la facilité avec laquelle le virus se propage à d'autres, la gravité de la maladie, l'efficacité de la thérapeutique et si les vaccins actuellement autorisés offrent une protection.

Des exemples d'étiquettes VOI, ainsi que leurs désignations scientifiques dans les trois différents systèmes de classification basés sur la phylogénie, sont présentés dans le tableau suivant :

Tableau 3 : Mécanismes de dénomination de l'OMS pour le SARS-CoV-2 VOI

Dénomination OMS	Lignée Pango	Clade/Lignée GISAID	Clade Nextstrain	Premiers cas répertoriés	Date de 1 ^{ère} désignation
Epsilon	B.1.427/B.1.429	GH/452R.V1	21C	USA, Mars 2020	05 mars 2021
Zêta	P.2	GR/484K.V2	20B/S.484K	Brésil, Avril 2020	17 mars 2021
Êta	B.1.525	G/484K.V3	21D	Plusieurs pays, Décembre 2020	17 mars 2021
Thêta	P.3	GR/1092K.V1	21E	Philippines, Janvier 2021	24 mars 2021
Lota	B.1.526	GH/253G.V1	21F	USA, Novembre 2020	24 mars 2021
Kappa	B.1.617.1	G/452R.V3	21B	Inde, Octobre	04 avril 2021

				2020	
Lambda	C.37	GR/452Q.V1	20D	Pérou, Décembre 2020	10 juin 2021

4-2/ Les VOC : sont des variants pour lesquels on observe une augmentation de la transmissibilité, d'une maladie plus grave (par exemple, une augmentation des hospitalisations ou des décès), une réduction significative de la neutralisation par les anticorps générés lors d'une infection ou d'une vaccination précédente, une efficacité réduite des traitements ou des vaccins, ou des échecs de détection diagnostique.

Un VOC est un variant qui répond à la définition du variant à suivre et dont on a montré, au moyen d'une évaluation comparative, qu'il est associé à un ou plusieurs des changements suivants, qui ont une certaine importance pour la santé publique mondiale :

- i. Augmentation de la transmissibilité ou évolution préjudiciable de l'épidémiologie de la COVID-19 ; ou
- ii. Augmentation de la virulence ou modification du tableau clinique ; ou
- iii. Diminution de l'efficacité des mesures de santé publique et sociales ou des outils de diagnostic, des vaccins et des traitements disponibles.

Des exemples d'étiquettes VOC, ainsi que leurs désignations scientifiques dans les trois différents systèmes de classification basés sur la phylogénie, sont présentés dans le tableau suivant :

Tableau 4 : Mécanismes de dénomination de l'OMS pour le SARS-CoV-2 VOC

Dénomination OMS	Lignée Pango	Clade/Lignée GISAID	Clade Nextstrain	Premiers cas répertoriés	Date de 1 ^{ère} désignation
Alpha	B.1.1.7	GRY(anciennement GR/501Y.V1)	20I (V1)	Royaume-Uni, Septembre 2020	18 décembre 2020
Bêta	B.1.351	GH/501Y.V2	20H (V2)	Afrique du sud, Mai 2020	18 décembre 2020
Gamma	P.1	GR/501Y.V3	20J (V3)	Brésil,	11 janvier

				Novembre 2020	2021
Delta	B.1.617.2	G/478K.V1	21A	Inde, Octobre 2020	VOI : 04 Avril 2021 puis VOC : 11 Mai 2021

5- LES MÉTHODES DE PRÉDICTION DES STRUCTURES 2D DES PROTÉINES

5-1/ Notions de structure 2D : Une protéine est une macromolécule d'une succession ordonnée en acides aminés. Cet ordre des résidus définit la structure primaire (donc la séquence, ou formule chimique) dont le premier est toujours situé en extrémité N-terminale et dont le dernier en extrémité C-terminale.

Les résidus qui constituent la structure primaire se succèdent les uns derrière les autres en fonction de leurs propriétés physico-chimiques, en particulier l'hydrophobicité (MARTIN J., 2005) de leurs radicaux, et ne succèdent jamais de manière aléatoire car la fonction de la protéine et son repliement en dépendent.

La structure et la fonction des protéines dépendent des propriétés des acides aminés qui composent ces protéines (Bourne P. E. et Weissig H., 2003). En effet, en fonction de l'hydrophobicité des radicaux amino-acides, on distingue les classes suivantes :

- i. Non polaires aliphatiques : les radicaux de cette classe sont non polaires et hydrophobes. On y distingue : G, A, V, L, I, M et P.
- ii. Aromatiques : Leurs radicaux sont aromatiques et relativement non polaires (hydrophobes). On y distingue F, Y et W. Tous ces résidus participent dans interactions hydrophobes.
- iii. Polaires non chargés : Les radicaux de ces résidus sont plus solubles dans l'eau (hydrophiles) que ceux des acides aminés non polaires, car ils contiennent des groupes fonctionnels qui forment des liaisons hydrogènes avec l'eau. Cette classe regroupe : S, T, C, N et Q.
- iv. Acides : Leur radical est de nature acide et est négativement chargé (COO^-). On y distingue : D et E.
- v. Basiques: Leur radical est de nature basique et est positivement chargé (N^+). On y distingue : K, R et H.

Les différentes classes physico-chimiques des acides aminés peuvent être représentées dans le diagramme de Venn (Figure 4).

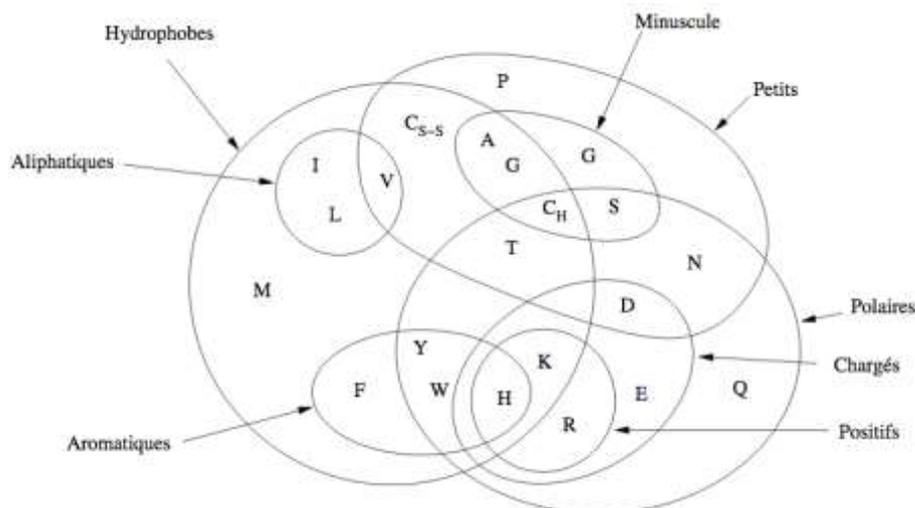


Figure 11 : Diagramme de Venn pour la classification physico-chimique des acides aminés (Taylor W. R., 1986).

La structure secondaire des protéines est principalement définie par trois principaux motifs (éléments), souvent répétés, en fonction des résidus. Les motifs en hélice α et en

feuillet β sont les deux principales structures secondaires régulières, prédites par Pauling et Corey en 1951 (MARTIN J., 2005).

- i. Le motif en hélice α : Hélice droite dont la stabilité est assurée par des liaisons hydrogènes entre le groupe carbonyle (C=O en bout de la chaîne) d'un résidu en position i et le groupe amide (dérivés de la fonction acide dont le *groupe* hydroxyle est remplacé par un *groupe* aminé $-NH_2$) du résidu à la position $i+4$. Les chaînes latérales se dirigent toujours vers l'extérieur de l'hélice.
- ii. Le motif en feuillet β : Les feuillets sont un type plus spacieux de structure secondaire formée de brins. Les brins sont constitués du squelette protéique en «zigzag», généralement pour quatre à dix résidus. Cependant, ils peuvent former des feuillets qui sont caractérisés par un motif de liaisons hydrogènes entre les résidus sur deux brins différents. Le résidu i dans le brin 1 forme une liaison hydrogène avec le résidu j dans le brin 2. La liaison suivante peut être formée de deux manières différentes : soit le résidu $i+2$ dans le brin 1 se lie à $j+2$ dans le brin 2 (appelé feuillet- β parallèle), ou le résidu $i+2$ dans le brin 1 se lie à $j-2$ dans le brin 2 (appelé feuillet- β anti-parallèle). Contrairement à la structure des hélices α , les résidus de liaison hydrogène dans les feuillets β peuvent être éloignés. La plupart des feuillets sont formés de plus de deux brins.
- iii. Les motifs Tours et Boucles (ou l'apériodique, ou Coil et Turn) : C'est une structure non alpha et non bêta. Les tours sont plus courts que les boucles (2 à 4 résidus contre une douzaine voire même une vingtaine de résidus).

Sur le plan bioinformatique, il existe une large palette de programmes informatiques (algorithmes) pour la prédiction des structures secondaires des protéines. L'un des premier programme utilise la méthode de Chou et Fasman (le plus basique); mais d'autres algorithmes plus détaillés sont disponibles sur les bases de données et l'on peut citer :

- i. GOR
- ii. Nearest neighbour methods
- iii. Hidden Markov models
- iv. Neural network
- v. Multiple alignment based self optimization method

Les principes de fonctionnement de tous ces programmes sont détaillés dans la littérature scientifique. La méthode GOR assume que les huit résidus amino-acides à droite et les huit résidus à gauche d'un acide aminé central à la position i de la protéine ($8 + 1 + 8 = 17$ aa) influent la propension de cet acide aminé central à être intégré dans une hélice (H), soit dans un feuillet bêta (E) ou encore dans un coude (C) ou, en fin, dans un tour (T). Cette propension est calculée sous forme d'un score sur la base de tables calculées auparavant à partir de séquences dont les structures 2D ont été expérimentalement établies.

Par exemple, considérons la portion peptidique suivante dont le résidu central est une proline (P) : Arg-Gly-Ser-Ala-Lys-Cys-Trp-Tyr-**Pro**-Ala-Val-Gly-Glu-Lys-Asp-Met-Ile. Les valeurs de chacun des acides aminés constitutifs de cette séquence sont tirées de la table de GOR (Garnier D.J. *et al.*, 1978) :

	R	G	S	A	K	C	W	Y	P	A	V	G	E	K	D	M	I	Σ
H	0	-10	-10	20	60	0	0	-40	-77	60	5	-40	78	0	15	20	-5	+76
E	0	20	20	0	-30	10	-10	35	-18	-20	40	20	-50	10	0	-20	0	+7
C	0	0	-20	0	-25	0	40	0	58	-25	-25	30	0	-20	0	-5	0	+8
T	0	0	-20	-10	-25	55	40	25	36	-40	-30	0	5	-20	0	-15	10	+11

D'après la colonne du total (Σ), le résidu P (Proline) a une propension à être intégré dans une hélice car son score est de +76 et présente la plus forte valeur parmi les quatre états possibles de structures secondaires (H, E, C et T).

Une fois la propension calculée pour ce résidu, on passe au résidu suivant qui, à son tour, devient central avec huit acides aminés à sa gauche et huit autres à sa droite. Il y aura donc $20 \times 4 \times 17$ possibilités. La précision apportée par l'algorithme GOR avoisine les 64%.

PARTIE 2 : PARTIE PRATIQUE

1- MATÉRIEL ET MÉTHODES

La partie expérimentale de notre projet consiste à acquérir les données protéiques et à en prédire les structures secondaires après avoir mis en évidence les mutations de substitutions entre les différents variants SARS-CoV-2.

Cinq variants ont été concernés par cette étude. Il s'agit des variants UK (Royaume uni), Inde, Afrique du sud, nigérian et Brésil ; car ce sont les virus SARS les plus répandus avec une incidence épidémiologique avérée.

1-1/ Matériel : La partie expérimentale a été réalisée sur des séquences de la protéine Spike des variants du SARS-CoV-2 et du virus référence Wu Han, téléchargées à partir de la base de données Uniprot hébergée à l'URL : <https://www.Uniprot.org>.

Des logiciels bioinformatiques pour le traitement et l'analyse des séquences de la protéine Spike du SARS-CoV-2 ont été utilisés :

- i. Pour prédire la structure secondaire, la base de données PRABI (URL : https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html) ; NPS@: Network Protein Sequence Analysis) a été exploitée en utilisant l'algorithme GOR IV de prédiction des structures 2D des protéines.
- ii. Pour l'analyse des domaines transmembranaires hydrophobes (selon l'échelle de kyte et Doolittle) des séquences protéiques Spike, nous avons exploité la base de données PSIPred (URL : <http://bioinf.cs.ucl.ac.uk/psipred/>).
- iii. La recherche des mutations a été réalisée sur la base de données GISAID (URL : <https://www.gisaid.org/>).
- iv. L'alignement multiple a été réalisé en utilisant le programme Clustal Oméga de EMBL (European Molecular Biology Laboratory ; URL : <https://www.ebi.ac.uk/Tools/msa/clustalo/>).

1-2/ Méthodes : Chaque étape d'analyse et d'annotation des protéines Spike des variants et celle de WuHan ont été réalisées *in silico* en ligne car les programmes utilisés sont libres d'utilisation mais non téléchargeables pour la plupart d'entre-eux.

1-2-1/ Prédiction de la structure secondaire : La méthode GOR (Garnier, Osguthorpe et Robson ; Garnier D. J., 1978) utilise un algorithme dont le principe se base sur la théorie de l'information. Les séquences de la protéine Spike sont copiées et collées dans la zone de texte de la base de données du portail PRABI. L'algorithme calcule la structure secondaire et rend le résultat sous la forme d'un graphique avec des barres verticales et colorées synonymes des acides aminés. Les barres bleues indiquent des résidus Hélice, les barres rouges indiquent des résidus Bêta. Le reste des bandes indique une structure irrégulière différente de l'hélice et du feuillet.

1-2-2/ Analyse des domaines transmembranaires hydrophobes : Nous avons exploité la base de données PSIPred afin de prédire les domaines hydrophobes

transmembranaires des protéines Spike. Le programme propose, en plus de la prédiction des domaines transmembranaires, d'autres prédictions telles que la famille de protéines, la fonction de la protéine, etc. La recherche de ces domaines transmembranaires permet de vérifier la diversité structurale voire fonctionnelle des différentes protéines virales Spike. Ainsi, deux structures Spike similaires peuvent avoir des domaines transmembranaires identiques si leurs structures primaires respectives sont homologues.

1-2-3/ Recherche des mutations : Les mutations recherchées sont ponctuelles et sont au nombre de 12 000 mutations cataloguées dans les bases de données (Caël B., 2021). Elles sont soit des substitutions soit des délétions (aucune recombinaison n'a été observée) ; car ce sont ces deux types de mutations les plus observés dans les différents variants du SARS-CoV-2 (OMS, 2021).

Les deux substitutions recherchées au sein des séquences Spike des différents variants sont :

- i. La substitution N501Y : C'est une mutation qui a lieu dans le domaine RBD de la protéine spike. Mécaniquement, la substitution N501Y a amélioré l'affinité de la protéine S pour les récepteurs cellulaires. Elle améliore alors la transmission du SARS-CoV-2. Comme suggéré par son évolution convergente au Brésil et en Afrique du Sud, les résultats indiquent que la substitution N501Y est une mutation adaptative majeure de pointe de préoccupation majeure.
- ii. La substitution D614G : décrite pour la première fois en Mars 2020, largement diffusée pour devenir majoritaire parmi les variants du CoV-2. Des analyses structurales ont révélé que cette substitution augmentait la capacité d'infection du virus et sa réplication dans les cellules épithéliales des voies respiratoires humaines, mais que le virus conservait sa morphologie et ses propriétés de neutralisation par des anticorps *in vitro*.

1-2-4/ Alignement multiple : Les analyses d'homologie des séquences protéiques Spike basées sur l'alignement multiple ont été réalisées pour révéler les substitutions et les délétions à chaque position.

Pour comprendre les relations évolutives entre les variants SARS-CoV-2, l'alignement multiple a été traité par un logiciel Clustal Oméga de la base de données EMBL afin de générer des relations évolutives entre les membres apparentés de diverses séquences Spike analysées.

2- RÉSULTATS & DISCUSSION

2-1/ PREDICTION 2D : La prédiction des structures secondaires des protéines Spike par l'algorithme GOR, dans sa 4^{ème} version, a révélé la présence des principaux motifs qui entrent dans une structure 2D basique à savoir l'hélice, le feuillet et le coude.

Nous constatons beaucoup de similitudes structurales entre les cinq séquences, principalement aux deux extrémités C-terminales et N-terminales. Les segments formés par les bandes rouges aux extrémités C et N-terminales sont identiques chez tous les variants ainsi que la souche WuHan.

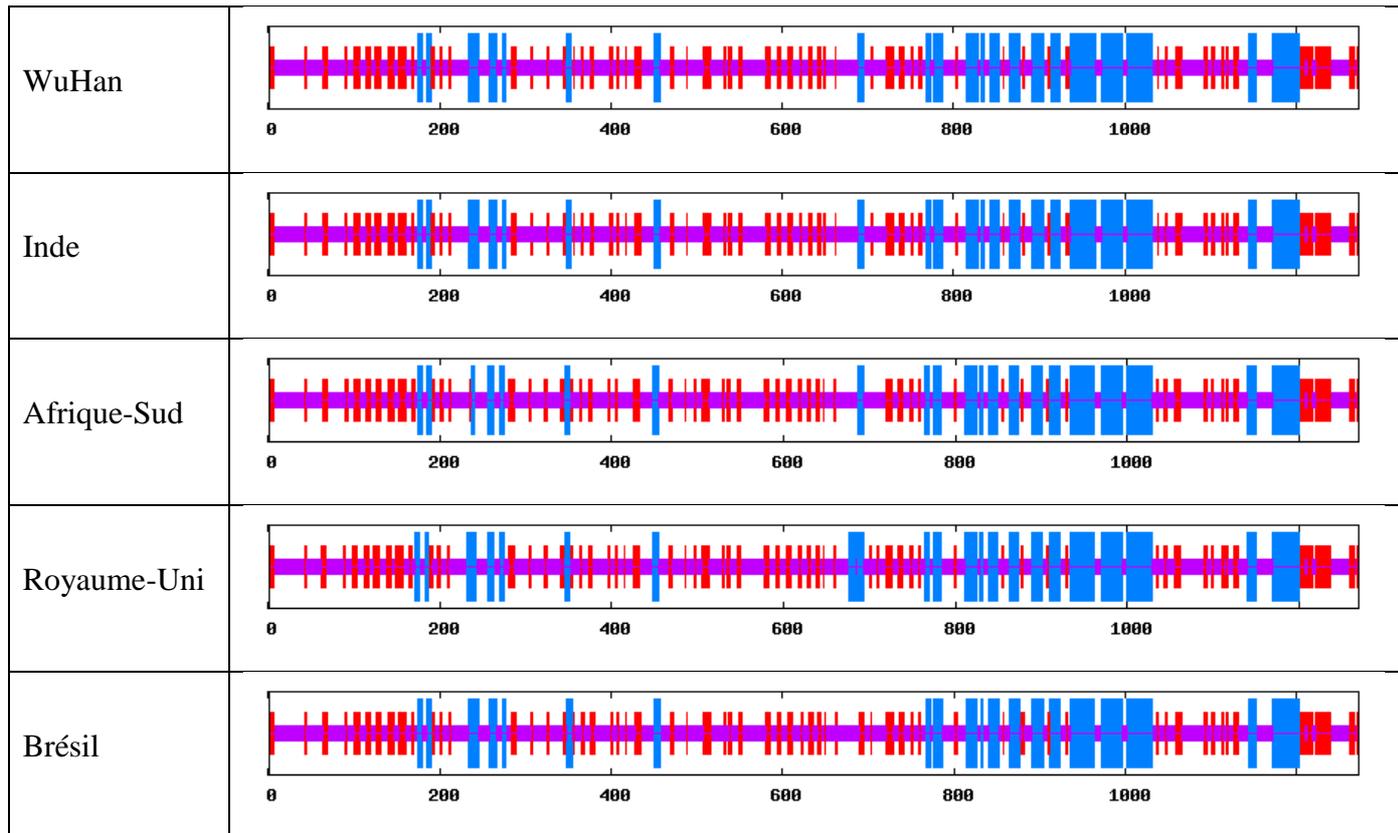


Figure 12 : Graphiques représentant les motifs 2D de la protéine S chez la souche originelle (Wuhan) et chez les variants analysés.

2-2/ Prédiction des zones hydrophobes transmembranaires : Les séquences de la référence Wu Han et celles des variants ont révélé des zones fortement hydrophobes transmembranaires similaires (Tableau 5) en nombres de résidus et dans la nature des propriétés physico-chimiques de ces mêmes résidus.

Ainsi, le variant britannique dispose d'une même zone transmembranaire N-Terminale comprenant 16 résidus dans le cas de Wu Han et 15 dans le cas du variant britannique. De plus, nous pouvons noter une forte similarité entre ces deux sous-séquences. L'alignement local de ces deux peptides a été réalisé avec le programme Emboss-Water de EMBL et a donné une similarité de 100 % :

WuHan	2 PLVDLPIGINITRFQ	16
England	1 PLVDLPIGINITRFQ	15

Ce pendant, nous avons noté l'absence de cette sous-séquence dans les variants : Sud-Africain, Indien et Brésilien.

Au milieu des séquences, vers les résidus en position 871 et plus, nous constatons une sous-séquence transmembranaire chez tous les variants et chez Wu Han, avec une longueur de 16 aa chez les variant Sud-Africain et Brésilien et 18 aa chez Wu Han et les deux variants Britanniques et Indien.

L'alignement multiple de ces cinq sous-séquences a été réalisé avec Clustal Oméga de EMBL. Les résultats suivant ont été obtenus :

WuHan	TSALLAGTITSGWTFG--	16
Inde	TSALLAGTITSGWTFGAG	18
Afrique-Sud	TSALLAGTITSGWTFG--	16
Rouame-Uni	TSALLAGTITSGWTFGAG	18
Brésil	TSALLAGTITSGWTF---	15

Nous pouvons constater que les sous-séquences ont en commun une fenêtre commune de 15 résidus : TSALLAGTITSGWTF. Cette zone est également riche en acides aminés hydrophobes (A, L, I) et aromatiques (F, T et W).

Vers l'extrémité C-terminale, les sous-séquences présentent un motif similaire d'une trentaine d'acides aminés. L'alignement multiple (Clustal Oméga) a donné les résultats suivants :

WuHan	IWLGFIAGLIAIVMVTIMLCCMTSCCCLKG	31
Inde	IWLGFIAGLIAIVMVTIMLCCMTSCCCLKG	31
Afrique-Sud	IWLGFIAGLIAIVMVTIMLCCMTSCCSC---	28
Rouame-Uni	IWLGFIAGLIAIVMVTIMLCCMTSCCCLKG	31
Brésil	IWLGFIAGLIAIVMVTIMLCCMTSCCSC---	28

Dans cet alignement multiple, il est à constater que tous les virus ont en commun le motif suivant de 28 résidus : IWLGFIAGLIAIVMVTIMLCCMTSCCSC.

Tableau 5 : Identification des zones transmembranaires hydrophobes chez les variants SARS-CoV-2 et chez le virus référence WuHan

WuHan	224-239 : 16 aa EPLVDLPIGINIT RFQ	874-889 : 16 aa TSALLAGTITSGWT FG	1216-1246 : 31 aa IWLGFIAGLIAIVMVTIMLCCMTSC CSCLKG
Inde	----- --	874-891 : 18 aa TSALLAGTITSGWT FGAG	1216-1246 : 31 aa IWLGFIAGLIAIVMVTIMLCCMTS CCCLKG
Afrique -Sud	----- --	871-886 : 16 aa TSALLAGTITSGWT FG	1213-1240 : 28 aa IWLGFIAGLIAIVMVTIMLCCMTS CCSC
Royau me-Uni	222-238 : 17 aa PLVDLPIGINITR FQTL	871-888 : 18 aa TSALLAGTITSGWT FGAG	1213-1243 : 31 aa IWLGFIAGLIAIVMVTIMLCCMTS CCCLKG
Brésil	----- --	873-888 : 16 aa TSALLAGTITSGWT FG	1215-1242 : 28 aa IWLGFIAGLIAIVMVTIMLCCMTS CCSC

2-3/ Recherche de mutations : Les résultats des identifications des mutations des variants contre le référentiel Wuhan a donné les résultats exprimés dans le tableau suivant :

Tableau 6 : Principales mutations identifiées entre les variants SARS-CoV-2 contre le virus référence Wuhan

	Substitution N501Y	Substitution D614G	Délétions
WuHan vs Inde	NON	OUI	NON

WuHan vs Afrique-Sud	OUI	OUI	L242 ; A243 ; L244
WuHan vs Royaume-Uni	OUI	OUI	H69 ; V70 ; Y144
WuHan vs Brésil	NON	NON	A694

Les mutations par substitution qui concernent la position 501 sur les protéines Spike ont été observées chez les deux variants Sud-africain et Britannique. En effet, ces deux protéines ont substitué l'acide aminé initial (Wuhan) qui est une Asparagine (N) par un autre acide aminé qui est la Tyrosine (Y). Ces deux résidus ont en commun la propriété d'être tous les deux polaires. La substitution N501Y est la seule mutation dans l'interface entre le RBD de B.1.1.7 et ACE2, ce qui fait craindre que sa reconnaissance par les anticorps neutralisants puisse être affectée (Lin C., 2021).

Seul, le variant Brésilien est caractérisé par une absence de mutation par substitution de l'acide aspartique (D) par une glycine (G). Par contre, les autres variants, à savoir celui de l'Inde, de l'Afrique du sud et du Royaume-Uni ont substitué leur résidu D par le résidu G en position 614 de la protéine Spike.

C'est une mutation faux-sens. Les codons d'ARN de l'acide aspartique (D) et la glycine (G) sont GAU/GAC et GGU/GGC, respectivement. Ainsi, une mutation unique dans le codon d'ARN entraînant la substitution du nucléotide A en G peut entraîner donc la substitution du résidu amino-acide D (polaire) en G (non polaire) dans la séquence de la protéine cible Spike. Les variantes G614 ont des titres infectieux considérablement plus élevés que les variantes D614. Ceci indique que la mutation de la pointe D614G rend le coronavirus variant plus infectieux, avec une réduction de la perte de la sous-unité S1, et peut être transmis plus facilement et rapidement. Les rétrovirus pseudotypés avec la substitution S^{G614} infectaient les cellules exprimant le récepteur ACE2 nettement plus efficacement que ceux avec naturels S^{D614} (Lizhou Z. *et al.*, 2020). Les mêmes auteurs rapportent que la distribution temporelle de cette substitution G614 est passée de 0% en janvier et février 2020 à 26% en mars 2020, puis à 65% en avril 2020 et à 70% en mai 2020 ; ce qui traduit une forte distribution de ce variant. La substitution D614G rend également la protéine S plus flexible (figure 13) ; facilitant ainsi son interaction avec le récepteur ACE2 cellulaire car la protéine S

originelle est sous une forme « fermée », où les *receptor binding domain* (RBD) sont masqués contrairement à la protéine S du variant (Kern, 2021).

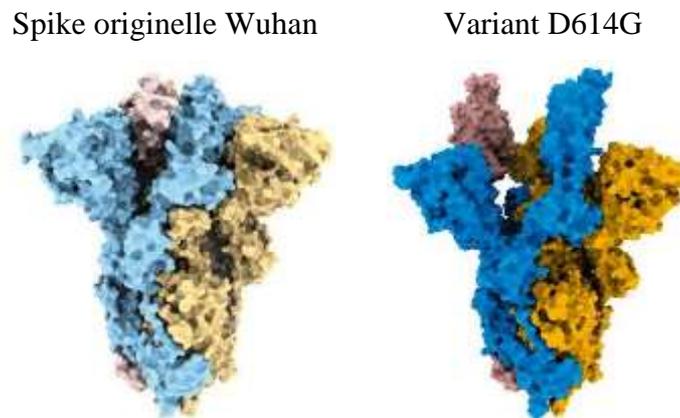


Figure 13 : La protéine S de la souche originelle Wuhan et celle du variant D614G (Kern, 2021).

2-4/ Alignement multiple : Les substitutions N501Y et D614G apparaissent dans l'alignement multiple chez les variants analysés. Ainsi à la position du résidu 501, la protéine des deux variants indien Brésilien est conservée comparativement à la protéine Wuhan originelle. Cependant, nous constatons que les variants Sud Africain et Britannique ont substitué leur résidu N (Asparagine) par un résidu Y (Tyrosine). De même, en position 614, il est à observé que seul le variant de l'Inde a conservé son résidu D (Acide Aspatique) comme la souche Wuhan originelle alors que les variants de l'Inde, de l'Afrique du sud et du royaume uni ont muté par substitution du résidu D en résidu G (Glycine).

	501				550
WuHan	NGVGYQPYRV	VVLSFELLHA	PATVCGPKKS	TNLVKNKCVN	FNFNGLTGTG
India	NGVGYQPYRV	VVLSFELLHA	PATVCGPKKS	TNLVKNKCVN	FNFNGLTGTG
SouthAfrica	YGVGYQPYRV	VVLSFELLHA	PATVCGPKKS	TNLVKNKCVN	FNFNGLTGTG
England	YGVGYQPYRV	VVLSFELLHA	PATVCGPKKS	TNLVKNKCVN	FNFNGLTGTG
Brazil	NGVGYQPYRV	VVLSFELLHA	PATVCGPKKS	TNLVKNKCVN	FNFNGLTGTG

	601				650
WuHan	GTNTSNQVAV	LYQDVNCTEV	PVAIHADQLT	PTWRVYSTGS	NVFQTRAGCL
India	GTNTSNQVAV	LYQGVNCTEV	PVAIHADQLT	PTWRVYSTGS	NVFQTRAGCL
SouthAfrica	GTNTSNQVAV	LYQGVNCTEV	PVAIHADQLT	PTWRVYSTGS	NVFQTRAGCL
England	GTNTSNQVAV	LYQGVNCTEV	PVAIHADQLT	PTWRVYSTGS	NVFQTRAGCL
Brazil	GTNTSNQVAV	LYQDVNCTEV	PVAIHADQLT	PTWRVYSTGS	NVFQTRAGCL

CONCLUSION

Dans les virus à ARN tels le SARS-CoV-2, et contrairement aux virus à ADN, la réplication de l'ARN viral peut comporter de nombreuses mutations donnant naissance à des produits de réplication différents (mutés) par rapport aux produit originels (sauvages).

La protéine de structure Spike est a subi ces mutations donnant naissance, à l'échelle planétaire, des variants qui comportent une protéine Spike différente par rapport à la molécule originelle de Wuhan.

Dans notre travail, grâce à des outils bioinformatiques, nous avons mis le doigt sur les deux substitutions les plus courantes répandues sur la planète.

En effet, les variants apparus ont révélé, d'après notre analyse, que dans certaines positions de la protéine, quelques acides aminés ont été substitués par d'autres.

Nous retenons, dans notre travail, que l'outil bioinformatique et les connaissances biochimiques structurales contribuent énormément dans la mise en évidence des substitutions et l'apparition de nouveaux variants ; et par conséquent, le couple ge de ces outils supporté par la précision du docking moléculaire vont apporter une valeur significative à la recherche en termes de productions pharmacologiques.

Il serait plus souhaitable de s'intéresser également aux autres substitutions telles que la E484K ou encore la E417T (proche de la K417N chez B.1.351) en mettant l'accent sur les interactions de cette protéine avec les récepteurs membranaires et avec le système immunitaire en utilisant le docking moléculaire.

BIBLIOGRAPHIE

- [1] Adami C, Pooley J, Glomb J, Stecker E, Fazal F, Fleming JO, et al. Evolution of mouse hepatitis virus (MHV) during chronic infection: quasispecies nature of the persisting MHV RNA. *Virology* 1995; 209 (2) : 337–46.
- [2] Belouzard S, Millet JK, Licitra BN, Whittaker GR. 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses*. **4** (6) : 1011–1033
- [3] Bertram S, Dijkman R, Habjan M, Heurich A, Gierer S, Glowacka I, et al. 2013. TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium. *J. Virol.* **87** : 6150–60
- [4] Bosch BJ, van der Zee R, de Haan CA, Rottier PJ. 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol.* **77** : 8801–11
- [5] Bradburne A. F. 1970. Antigenic relationships amongst coronaviruses. *Archiv Gesamte Virusforsch.* **31** : 352–364
- [6] Brian D. A. 1995. The Coronavirus Hemagglutinin-esterase Glycoprotein. *In* : S S, ed. *The Coronaviridae*. New-York: Plenum Press.
- [7] Callow K. A., Parry H. F., Sergeant M., Tyrrell D. A. 1990. The time course of the immune response to experimental coronavirus infection of man. *Epidemiol Infect.* **105** : 435– 446
- [8] Chan JF. W., Kok K. H., Zhu Z. et al. 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* **9**. (1) : 221–236

- [9] Corse E., Machamer C. E. 2003. The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction. *Virology*. **312** (1) : 25–34
- [10] Cui J., Li F., Shi Z. L. 2019. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17** (3) : 181–192
- [11] De Haan C. A., Rottier P. J. 2005. Molecular interactions in the assembly of coronaviruses. *Adv. Virus Res.* **64** : 165–230
- [12] Deng X., Van Geelen A., Buckley A. C. *et al.* 2019. Coronavirus endoribonuclease activity in porcine epidemic diarrhea virus suppresses type I and type III interferon responses. *J. Virol.* **93** (8) : 2000–2018
- [13] Du L., Kao R. Y., Zhou Y., He Y., Zhao G., Wong C., *et al.* 2007. Cleavage of spike protein of SARS coronavirus by protease factor Xa is associated with viral infectivity. *Biochem Biophys Res Commun.* **359** : 174–9
- [14] Enjuanes L. S. I., Zuniga S., Moreno J. L. 2007. Coronavirus RNA synthesis: transcription. *In: V T*, ed. *Coronaviruses Molecular and Cellular Biology*. Norkolk, UK: *Caister Academic Press*. 81-107
- [15] Escors D., Ortego J., Laude H., Enjuanes L. 2001. The membrane M protein carboxy terminus binds to transmissible gastroenteritis coronavirus core and contributes to core stability. *J. Virol.* **75** (3) : 1312–1324
- [16] Fehr A. R., Perlman S. 2015. Coronaviruses : an overview of their replication and pathogenesis. *In: Coronaviruses Methods in Molecular Biology*. Maier H, Bickerton E, Britton P (Eds). Springer, NY, USA, 1282, 1–23
- [17] Fung S. Y., Yuen K. S., Ye Z. W., Chan C. P., Jin D. Y. 2020. A tug-of-war between severe acute respiratory syndrome coronavirus 2 and host antiviral defence : lessons from other pathogenic viruses. *Emerg. Microbes Infect.* **9** (1) : 558 – 570

- [18] Gallagher T. M., Buchmeier M. J. 2001. Coronavirus spike proteins in viral entry and pathogenesis. *Virology*. **279** (2) : 371–4
- [19] Gralinski L. E., Menachery V. D. 2020. Return of the coronavirus: 2019-nCoV. *Viruses*. **12** (2) : 135
- [20] Herrewegh A. A., Smeenk I., Horzinek M. C., Rottier P. J., De Groot R. J. 1998. Feline coronavirus type II strains 79-1683 and 79-1146 originate from a double recombination between feline coronavirus type I and canine coronavirus. *J Virol*. **72** (5) : 4508–14
- [21] Hoffmann M., Kleine-Weber H., Schroeder S., Kruger N., Herrler T., Erichsen S., *et al.* 2020. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*. **181** : 271–80
- [22] Hutchinson, E. G., Thornton, J. 1994. M. Protein Sci. *Publ. Protein Soc.* **3** (12) : 2207.
- [23] James M. W., E., Poet, R. 1987. Trends Biochem. *Sci.* **12** : 189
- [24] Kabsch, W., Sander C. 1983. *Biopolymers*. **22** (12) : 2577
- [25] Kienzle T. E., Abraham S., Hogue B. G., Brian D. A. 1990. Structure and expression of the bovine coronavirus hemagglutinin protein. *Adv Exp Med Biol*. **276** : 95–102
- [26] Krempl C., Schultze B., Herrler G. 1995. Analysis of cellular receptors for human coronavirus OC43. *Adv Exp Med Biol*. **380** : 371–4
- [27] Lan J., Ge J., Yu J., Shan S., Zhou H., Fan S., *et al.* 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. **581** : 215–20
- [28] Laude H. D. B., Eleouet J. F. 1998 Le coronavirus porcine PRCV: un virus émergent pas comme les autres. *Virologie*. **2** : 305–16
- [29] Li F. 2016. Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* **3**(1), 237–261

- [30] Li W., Moore M. J., Vasilieva N. *et al.* 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*. **426** (6965) : 450–454
- [31] Li W., Sui J., Huang I. C., Kuhn J. H., Radoshitzky S. R., Marasco W. A., *et al.* 2007. The S proteins of human coronavirus NL63 and severe acute respiratory syndrome coronavirus bind overlapping regions of ACE2. *Virology*. **367** (2) : 367–74
- [32] Lu R., Zhao X., Li J. *et al.* 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. **395** (10224) : 565–574
- [33] Matsuyama S., Nagata N., Shirato K., Kawase M., Takeda M., Taguchi F. 2010. Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. *J. Virol.* **84** (24) : 12658–12664
- [34] McIntosh K., Becker W. B., Chanock R. M. 1967. Growth in suckling-mouse brain of “IBV-like” viruses from patients with upper respiratory tract disease. *Proc Natl Acad Sci USA*. **58** : 2268–2273
- [35] McIntosh K., Kapikian A. Z., Hardison K. A., Hartley J. W., Chanock R. M. 1969. Antigenic relationships among the coronaviruses of man and between human and animal coronaviruses. *J Immunol.* **102** : 1109–1118
- [36] McIntosh K., Kapikian A. Z., Turner H. C., Hartley J. W., Parrott R. H., Chanock R. M. 1970. Seroepidemiologic studies of coronavirus infection in adults and children. *Am J Epidemiol.* **91** : 585–592
- [37] Millet J. K., Whittaker G. R. 2018. Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells. *Virology*. **517** : 3–8
- [38] Monto A. S. 1974. Medical reviews: coronaviruses. *Yale J Biol Med.* **47** : 234–251

- [39] Mortola E, Roy P. 2004. Efficient assembly and release of SARS coronavirus-like particles by a heterologous expression system. *FEBS Lett.* **576** (1–2) : 174–178
- [40] Narayanan K., Maeda A., Maeda J., Makino S. 2000. Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. *J. Virol.* **74** (17) : 8127–8134
- [41] Neuman B. W., Kiss G., Kunding A. H. *et al.* 2011. A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* **174** (1) : 11–22
- [42] Qian Z., Dominguez S. R., Holmes K. V. 2013. Role of the spike glycoprotein of human Middle East respiratory syndrome coronavirus (MERS-CoV) in virus entry and syncytia formation. *PLoS ONE.* **8** (10) : e76469
- [43] Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. J. *Mol. Biol.* 1963, 7 (1), 95
- [44] Richardson, J. S. 1981. *In* Advances in Protein Chemistry. C.B. Anfinsen, J. T. E. and F. M. R., Ed.; Academic Press. Vol. **34** : pp 167–339
- [45] Schwegmann-Wessels C., Herrler G. 2006. Sialic acids as receptor determinants for coronaviruses. *Glycoconj J* ; **23** (1–2) : 51–8
- [46] Tang T., Bidon M., Jaimes J. A., Whittaker G. R., Daniel S. 2020. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antivir Res.* **178** : 104792
- [47] Tyrrell D. A., Almeida J. D., Cunningham C. H. *et al.* 1975. Coronaviridae. *Intervirology.* **5** : 76–82
- [48] Vabret A., Dina J., Mourez T., Gouarin S., Petitjean J., van der Werf S., *et al.* 2006. Inter- and intra-variant genetic heterogeneity of human coronavirus OC43 strains in France. *J Gen Virol* **87** (Pt 11) : 3349–53
- [49] Vlasak R., Luytjes W., Spaan W., Palese P. 1988. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. *Proc Natl Acad Sci.* **85** (12) : 4526–9

- [50] Walls A. C., Park Y. J., Tortorici M. A., Wall A., McGuire A. T., Veerler D. 2020. Structure, Function and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*. **181** : 281–92
- [51] Wan Y., Shang J., Graham R., Baric R. S., Li F. 2020. Receptor recognition by the novel coronavirus from Wuhan : An analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* **94** (7) : e00127–20
- [52] Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, *et al.* 2020 . Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*. **181** : 894–904.e9
- [53] Watanabe Y., Allen J. D., Wrapp D., McLellan J. S., Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*. **369** : 330–3
- [54] Wentworth D. E. 2007. Coronavirus binding and entry. In: V. T, ed. Coronaviruses Molecular and cellular biology. Norfolk, UK: *Caister Academic Press*. pp : 3–31
- [55] Witte K. H., Tajima M., Easterday B. C. Morphologic characteristics and nucleic acid type of transmissible gastroenteritis virus of pigs. *Arch. Gesamte Virusforsch.* **23** (1) : 53-70. DOI : 10.1007/BF01242114.
- [56] Wong S. K., Li W., Moore M. J., Choe H., Farzan M. 2004. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin converting enzyme 2. *J Biol Chem.* **279** (5) : 3197–201
- [57] Wrapp D., Wang N., Corbett K. S. *et al.* 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. **367** (6483) : 1260–1263
- [58] Wu F., Zhao S., Yu B. *et al.* 2020. A new coronavirus associated with human respiratory disease in China. *Nature*. **579** (7798) : 265–269

- [59] Xia S., Yan L., Xu W. 2019. Agrawal AS, Algaissi A, Tseng CK, *et al.* A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Sci Adv.* **5** : eaav 4580
- [60] Xia S., Zhu Y., Liu M., Lan Q., Xu W., Wu Y., *et al.* 2020. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell Mol Immunol.* **17** : 765–7
- [61] Xu X., Chen P., Wang J. *et al.* 2020. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci. China Life Sci.* **63** (3) : 457–460
- [62] Yeager C. L., Ashmun R. A., Williams R. K., Cardellicchio C. B., Shapiro L. H., Look A. T., *et al.* 1992. Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature.* **357** (6377) : 420–2
- [63] Yokomori K., Banner L. R., Lai M. M. 1991. Heterogeneity of gene expression of the hemagglutinin-esterase (HE) protein of murine coronaviruses. *Virology.* **183** (2) : 647–57
- [64] Zhang X. M., Kousoulas K. G., Storz J. 1992. The hemagglutinin/esterase gene of human coronavirus strain OC43: phylogenetic relationships to bovine and murine coronaviruses and influenza C virus. *Virology.* **186** (1) : 318–23
- [65] Zhou P., Yang X. L., Wang X. G. *et al.* 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* **579** (7798) : 270–273
- [66] Zhou P., Yang X. L., Wang X. G. *et al.* 2020. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *BioRxiv.*
- [67] Zhu N., Zhang D., Wang W. *et al.* 2020. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382** (8) : 727–733

- [68] Ziebuhr J. S. E. 2007. The Coronavirus replicase gene: special enzymes for special viruses. *In*: V T, ed. *Coronaviruses Molecular and Cellular Biology*. Norfolk, UK : Caister Academic Press. pp : 33–63

Soutenu le :
Juillet 2021

Présenté par :
CHEHBOUB Faten & ZITOUNI Rofaida

Thème :

Analyse Viro-Informatique des variants SARS-CoV-2 portée sur la protéine Spike

Résumé :

Le but de ce travail est de réaliser une annotation partielle de la protéine de surface du SARS-CoV-2 des variants du Brésil, de l'Inde, du Royaume-Uni et de l'Afrique du sud comparativement au coronavirus originel de Wuhan. L'annotation bioinformatique a porté sur des éléments structuraux tels que la prédiction des domaines hydrophobes transmembranaires des protéines Spike, leurs structures secondaires et la définition des mutations par substitutions en particulier en positions 501 et 614 car les plus répandues dans les cinq continents. Les substitutions ont concerné tous ses variants et certaines délétions ont également été rapportées dans ce travail.

Mots clés : SARS-CoV-2 – Substitutions – Spike - Variants

Jury d'évaluation :

Président	Dr MOKRANI EL Hassen
Encadreur	Pr HAMIDECHI Mohamed Abdelhafid Dr GUENDOUZE Assia
Examinatrice	

