



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



Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة الشهيد حمزة لخضر الوادي

Université Echahid Hamma Lakhdar- EL OUED

كلية العلوم الطبيعية والحياة

Faculté des Sciences de la Nature et de la Vie

قسم البيولوجيا الخلوية والجزيئية

Département de Biologie Cellulaire et Moléculaire

## MEMOIRE DE FIN D'ETUDE

En vue de l'obtention du diplôme de Master Académique en Sciences biologiques

Spécialité : Toxicologie

### THEME

**Certaines analyses bio-informatiques des plus importantes protéines de SARS-CoV-2**

Présenté Par :

Khennoufa Zeghida  
Berdjouh Souhaia  
Saadoun Abir  
Aouine Widad

Devant le jury composé de :

Présidente : Toumi Ikram

MCA

Université d'El Oued.

Examineur : Lanez Elhafnaoui

MCB

Université d'El Oued.

Promoteur : Medjour Abdelhak

MAA

Université d'El Oued.

Année académique 2021/2022

## الملخص:

كوفيد 19 هو مرض يصيب الجهاز التنفسي يسببه فيروس يعرف باسم SARS-CoV-2 كان هذا الفيروس سبب وفاة أكثر من 6 ملايين شخص على مستوى العالم اعتباراً من شهر جوان 2022 حسب تقديرات منظمة الصحة العالمية. مثل هذا الانتشار الواسع للعدوى الناتجة عن هذا الفيروس ومقاومته الشديدة للجهاز المناعي، تنتج في معظم الحالات عن الخصوصية الفريدة للبروتينات التي تميز فيروس SARS-CoV-2. على هذا الأساس ومنذ ظهوره ، إتجه العالم نحو الأبحاث التوصيفية لبروتينات SARS-CoV-2 سواء مخبرياً أو عن طريق إستخدام تقنيات المعلوماتية الحيوية، من أجل تطوير لقاحات وأدوية جديدة لمكافحة الكوفيد 19. عملنا هو جزء من هذا السياق. ففي هذا المنحى، تم إجراء دراسة بإستخدام تقنيات المعلوماتية الحيوية على بروتينات SARS-CoV-2 (البروتين الغشائي، البروتين النووي، البروتين الغمدي الصغير، البروتين سبايك، متعدد البروتين المضاعف 1a متعدد البروتين المضاعف 1ab). لهذا، إستخدمت مع هته البروتينات العديد من الأدوات وبرامج المعلوماتية الحيوية ( BLAST ، Clustal Omega ، COBALT ، Jalview ، Prosite ، SOPMA ، I-TASSER server ، GPMW). من خلال هذه الدراسة، ساهمنا بشكل خاص في تحديد المناطق المحافظة "المسؤولة عن وظيفة هذه البروتينات" (بما في ذلك: (91) MWLSYFIA (98) لبروتين الغشاء، (114) GTG (116) البروتين النووي، CAYCCNIVNVSLVKPSFYVYS (40) (75) للبروتين الغمدي الصغير، (1212) KWPW (1214) للبروتين سبايك، (1118) VGP (1120) متعدد البروتين المضاعف 1a). تمكنا أيضاً من إنشاء هيكل ثنائي الأبعاد وثلاثي الأبعاد وكلها ضرورية لفهم وتفسير الوظائف والتفاعلات وآلية التجميع الذاتي وما إلى ذلك بشكل أفضل لهذه البروتينات وكذلك آلية العدوى وطريقة حياة هذا الفيروس. من ناحية أخرى، يمكن استخدام التوصيف الفيزيائي الكيميائي الذي تم إجراؤه عن طريق برنامج المعلوماتية الحيوية لهذه البروتينات على المساعدة في الاختبارات التجريبية أو تنقية فصل هاته البروتينات.

**الكلمات المفتاحية:** SARS-CoV-2، تحليلات المعلوماتية الحيوية، البروتينات، دراسة in silico، كوفيد

## **Abstract**

Covid-19 is a respiratory disease caused by a virus known as SARS-CoV-2. This virus has been the cause of death of more than 6 million people globally according to WHO estimates as of June 2022. Such contagious invasion and anti-immune resistance are, of which most of cases, originating from the particularity of the proteins characterizing the SARS-CoV-2 virus. Thus, since its appearance, the world has moved towards in vitro and in silico characterization research on the proteins of SARS-CoV-2, in order to develop new vaccines and drugs against Covid 19. Our work is part of this context. In this sense, an in-silico study (bioinformatics analyzes) was carried out on the proteins of SARS-CoV-2 (membrane protein, nucleoprotein, small envelope protein, spike protein, replicase 1a polyprotein and replicase 1ab polyprotein). For this, these proteins were confronted with several tools and bioinformatics programs (BLAST, Clustal Omega, COBALT, Jalview, Prosite, SOPMA, I-TASSER server, GPMaw). Through this study we have contributed, in particular, to the determination of the conservative regions "responsible for the function of these proteins" (including: (91)MWLSYFIA(98) for the membrane protein, (114)GTG(116) for the nucleoprotein, (40)CAYCCNIVNVSLVKPSFYVYS(75) for the small envelope protein, (1212)KWPW(1214) for the spike protein, (1118)VGP(1120) for the replicase polyprotein 1a). We were also able to establish the 2-dimensional and 3-dimensional structure which are all essential to better understand and interpret the functions, interactions, self-assembly mechanism and etc. of these proteins as well as the mechanism of infection and the way of life of this virion. On the other hand, the physicochemical characterization carried out in silico of these proteins, can be used to help in the experimental tests or the separation and purification of the proteins of SARS-CoV-2.

**Keywords:** SARS-CoV-2, Bioinformatics analyses, proteins, in silico study, Covid-19.

## Résumé

Le Covid-19 est une maladie respiratoire causée par un virus connu sous le nom de SARS-CoV-2. Ce virus a été la cause de décès de plus de 6 millions de personnes à l'échelle mondiale d'après les estimations de l'OMS de mois de juin 2022. Une telle invasion contagieuse et résistance anti-immunitaire sont, dont la plupart de cas, originaires de la particularité des protéines caractérisant le virus de SARS-CoV-2. Ainsi, depuis son apparition le monde s'est dirigé vers les recherches de caractérisation in vitro et in silico des protéines de SARS-CoV-2, afin de développer des nouveaux vaccins et médicaments contre le Covid-19. Notre travail s'inscrit dans ce contexte. Dans ce sens on a réalisé une étude in silico (analyses bio-informatique) sur les protéines de SARS-CoV-2 (protéine membranaire, nucléoprotéine, petite protéine d'enveloppe, protéine spike, polyprotéine réplicase 1a et polyprotéine réplicase 1ab). Pour cela, ces protéines ont été affrontées face aux plusieurs outils et programmes bio-informatiques (BLAST, Clustal Omega, COBALT, Jalview, Prosite, SOPMA, I-TASSER server, GPMW). A travers cette étude nous avons contribué, notamment, à la détermination des régions conservatives « responsables à la fonction de ces protéines » (y parmi : (91)MWLSYFIA(98) pour la protéine membranaire, (114)GTG(116) pour la nucléoprotéine, (40)CAYCCNIVNVSLVKPSFYVYS(75) pour la petite protéine d'enveloppe, (1212)KWPW(1214) pour la protéine spike, (1118)VGP(1120) pour la polyprotéine réplicase 1a). Nous avons, également pu établir la structure 2 dimensionnelle et 3 dimensionnelle qui sont tous essentiels pour mieux comprendre et interpréter les fonctions, les interactions, le mécanisme d'autoassemblage et etc. de ces protéines ainsi que le mécanisme de l'infection et le mode de cycle de vie de ce virion. D'autre part, la caractérisation physicochimique réalisé d'une manière in silico de ces protéines, peut être utilisées pour aider dans les essais expérimentaux ou la séparation et la purification des protéines de SARS-CoV-2.

Mots clés : SARS-CoV-2, Analyses bio-informatiques, protéines, étude in silico, Covid-19.

## الاهداء

بسم الله الرحمن الرحيم أهدي هذا العمل لأمي وأبي الكريمين اللذان ضحيا من أجل ان اكمل دراساتي العليا  
أشكركم على كل الدعم الذي قدمتموه لي طوال مسيرتي الدراسية والمساندة الابدية لي . واشكر كل اخوتي  
واخواتي الاعزاء على قلبي لمجهوداتكم معي في هذا العمل ودعمي لأجتاز هذا المشوار الطويل والجميل .  
والشكر الخاص لاستاذي ومشرف هذه المذكرة الدكتور عبد الحق مجور على مجهوداته معنا طوال العام  
وحرصه على اتمام عملنا على أكمل وجه.  
شكرا لصديقتي العزيزة عبير سعدون على دعمي دائما ومساندتي في هذه الرحلة وايضا زميلاتي سهية  
وزغيدة لمجهوداتهم المبذولة في سبيل العلم والعمل.  
وفي الاخير الحمد لله تعالى على وصولي لهذه المرحلة وانهاء كل شئ سعيت لاجله دائما وأبدا .

وداد عوين

باسم الله نبدأ وبه نختم

لم يكن الأمر سهلاً أبداً لكننا فعلناها فالحمد لله حمداً كثيراً طيباً مباركاً فيه والشكر له أن وفقنا لإكمال ثمرة الجهد والإجتهاد لم نكن لنصل الي هذه المرحلة لولا فضل الله علينا .

الى والدتي جنتي وكل حياتي الي من تستحق كل نجاحاتي هي صاحبة عمري والحياة وصلت بتشجيعك ودعواتك دمتي فخرا حفظك الله لنا وبارك في عمرك.

الى والدي تاج الراس سندي وبهجة القلب من ينظر الي نظرة الفخر دائماً أدامك الله لنا وأطال في عمرك.

الى الكتف الذي لا يميل والضلع الثابت اخي سفيان ومعيني طه و سبيل راحتي غيث ، حسام وهند ادامكم الله لي ووفقكم اينما حلت خطاكم .

الى رفيقات الدرب وخير الصحبة : سمية ، بهجة ، أنهار ، نورس ، هدى،ماريا، عبير ، وداد أنتم خير ماجادت بهم الصدف وفقن الله وبارك فيكن.

الي الصحبة الصالحة واخوات الاقامة الجامعية هم حياة للحياة وصانعات الذكرى :حسنا وسمية لحظات الفرح والسعادة معكم لا تنسى ، اللهم لا فراق بيننا لا عن حياتي ولا عن أيامي.

الي أستاذنا الكريم مجور عبد الحق شكراً على الدعم الذي قدمته لنا ونفع الله بك .

الي كل من عرف سهية ومر هيناً ليناً مخلفاً ذكرى طيبة .

سهية برجوح

ها انا اخيرا وليس آخرا اجتاز احد العتبات بالامس كنا فالبكالوريا واليوم نجتاز الماستر . الحمد لله وكفى وصلاة على الحبيب المصطفى واهله ومن وفى اما بعد ثمرة الجهد والنجاح مهدات إلى الوالدين حفظهما الله . كتفي و سندي الذي اتكى عليه "أبي" روضة الحنان والامل "امي" من ابنتكم لكم كل الحب و الشكر والتقدير

ولكل عائلتي الكريمه و زوجي الذي ساندني و جميع اصدقائي بالاخص رفيقتي وداد عوين التي كانت رفيقتي في مشواري هاذا والي زملائي زغيده خنوفه و سهية برجوح ولا ننسى استاذنا الداعم في هذه المذكرة مجور عبد الحق اعانه الله و حفظه

عبير سعدون

نحمد الله عز وجل الذي وفقنا في اتمام هذا البحث العلمي والذي الهمنا الصحة و العافية و العزيمة فالحمد لله  
حمدا كثيرا

اما بعد اهدي إلى من وضع المولى سبحانه وتعالى الجنة تحت قدميها ووقرها في كتابه العزيز...(أمي  
الحبيبة).

الى من كان سندي في الحياة (ابي العزيز ) حفصهما الله

لكل اخوتي (سيف الدين ;احمد وزياد) واخواتي (نوال ;امنة وسماح ) دتم سنادا لي ومحبتي الذين  
كانوا سبب في تشجيع لمسيرتي وبالاخص اخي انور الذي فارقنا الحياة منذ 5 اشهر رحمه الله

الى اعز صديقاتي ( سهية برجوح ) و(منى علال )وزميلاتي (عوين ويداد) و(عبير سعدون) ولاانسى خطيبي  
الذي بذل مجهوداته معي

الى من اعطى واجزل بعطائه الى استاذي ومعلمي المشرف الدكتور (مجور عبد الحق) جزاك الله خيرا

الى كل من عرف ( زغيدة) وكل من وصلهم قلبي ولم يكتبهم قلبي.

ز غيدة خنوفة

## Remerciements

En premier lieu, nous remercions Allah qui nous a aidé pour l'accomplissement de ce travail.

Ensuite, nous remercions :

Monsieur MEDJOUR Abdelhak de nous avoir encadrer durant ce mémoire, d'avoir partagé son expérience et sa culture scientifique avec nous et d'avoir confiance en nos personnes. De travailler sous votre direction a été un honneur et un plaisir pour nous.

Nos sincères remerciements s'adressent également aux membres de jury :

Madame TOUMI Ikram, d'avoir accepté de présider le jury d'évaluation et rectification de notre mémoire.

Monsieur LANEZ Elhafnaoui d'avoir accepté d'examiner, rectifier et évaluer notre travail.

Enfin, nos remerciements s'adressent à tous les enseignants faculté des sciences de la nature et de la vie de l'université Echahid Hamma Lakhdar d'El-Oued qui ont participé du près ou du loin à notre formation durant toutes ces années d'étude. Ainsi que toutes les personnes qui nous ont aidés à arrive à ce stade.

## Liste d'abréviations

**ACE2** : *Angiotensin-converting enzyme 2*

**ADN** : Acide désoxyribonucléique

**ARN** : acide ribonucléique

**ARNm** : Acide ribonucléique messenger

**CoV** : coronavirus

**BLAST**: Basic Local Alignment Search Tool

**CASP**: Critical Assessment of protein Structure Prediction

**COVID-19** Couronna virus disease 1

**E** enveloppe

**ELISA** : Enzyme-linked immunosorbent assay

**FIPV** : Fractional doses of the inactivated poliovirus vaccine

**HCoV** coronavirus humains

**HCoV-OC43** Le coronavirus humain sigle -OC43

**IBV** : *Infectious Bronchitis Virus*

**ICTV** : *The International Committee on Taxonomy of Viruses*

**IDR** : Une intradermoréaction

**IFN** : nterférons

**IgA** : Immunoglobulines A

**IgG** : Immunoglobulines G

**IgM** : Immunoglobulines M

**Kb** : kilobyte

**I-TASSER**: Iterative Threading ASSEmbly Refinement

**M** membrane

**MERS-CoV** Coronavirus Le syndrome respiratoire du Moyen-Orient

**MHV** : Mouse Hepatitis Virus

**N** nucléocapside

**NKG2A** : Natural Killer Cell Receptor

**NSP** : neuroendocrine-specific protein

**OMS** : *L'Organisation mondiale de la santé*

**ORF** : open reading frame

**PCR** : *Polymerase Chain Reaction*

**PEDv** : *Porcine epidemic diarrhea virus*

**Ph** : *Le potentiel hydrogen*

**RBD** : Receptor Binding Domain

**RT-PCR** réaction en chaîne par polymérase

**S** la protéine Spike

**SRAS** syndrome respiratoire aigu severe

**SRAS-CoV** Coronavirus du syndrome respiratoire aigu sévère

**SRAS-CoV-2** Coronavirus 2 du syndrome respiratoire aigu sévère

**TDR** : *Test de diagnostic rapide*

**TMD** : Temporomandibular disorders

**TMPRSS2** Transmembrane protease serine 2

**TROD** tests rapides d'orientation diagnostique

**VLP** : *Virus-like particles*

**Liste des tableaux**

| Tableau | Titre                                                                                         | Page |
|---------|-----------------------------------------------------------------------------------------------|------|
| 01      | Séquences protéiques de SARS-CoV-2 qui ont été utilisées dans pour analyses bio-informatiques | 19   |

Liste des figures

| Figure | Titre                                                                                                                                                                                        | Page |
|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 01     | Le coronavirus SARS-CoV-2 vu au microscope électronique.                                                                                                                                     | 6    |
| 02     | Structure schématisée du SARS-COV-2.                                                                                                                                                         | 8    |
| 03     | Organisation du génome du SARS-CoV-2.                                                                                                                                                        | 9    |
| 04     | Séquence de protéine S.                                                                                                                                                                      | 11   |
| 05     | Séquence Protéine E.                                                                                                                                                                         | 12   |
| 06     | Séquence de protéine N                                                                                                                                                                       | 13   |
| 07     | (A) Image microscopique électronique à transmission d'une seule particule virale SARS-CoV-2                                                                                                  | 14   |
| 08     | Sequence Protéine M                                                                                                                                                                          | 15   |
| 09     | Schéma d'une particule de Coronavirus, la protéine M de la membrane est liée à la fois avec la membrane de l'enveloppe viral et la protéine N qui se lie à son tour avec le génome de l'ARN. | 15   |
| 10     | Séquence de polyprotéine réplicase 1a                                                                                                                                                        | 17   |
| 11     | Séquence de polyprotéine réplicase 1ab                                                                                                                                                       | 18   |
| 12     | Alignement multiple de protéine de membrane de SARS-CoV-2 avec 100 autres séquences similaires des Corona virus.                                                                             | 27   |
| 13     | Résultat d'alignement de 100 séquences protéiques de nucléoprotéines appartenant à des sources différentes pour détermination des régions conservatives.                                     | 28   |
| 14     | L'arbre phylogénétique réalisé par Blast Tree View. La séquence requête est surlignée en jaune.                                                                                              | 29   |
| 15     | Alignement multiple de 100 séquences protéiques de petite protéine d'enveloppe de SARS-CoV par COBALT.                                                                                       | 29   |
| 16     | Résultat d'alignement via BLAST de la séquence de protéine de spécule de SARSCoV-2.                                                                                                          | 30   |

|      |                                                                                                                                                                                                                                                                                                                 |    |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 17   | Résultat d'alignement avec 100 séquences protéiques des différentes sources de polyprotéine 1a le motif conservé est encadré en rouge.                                                                                                                                                                          | 31 |
| 18   | Résultat d'alignement via BLAST de la séquence de polyprotéine réplicase de SARS-CoV-2.                                                                                                                                                                                                                         | 32 |
| 19   | Résultats d'analyse de la séquence protéique de protéine membranaire de SARS-CoV-2 avec Prosite.                                                                                                                                                                                                                | 32 |
| 20   | Résultats d'analyse de la séquence protéique de nucléoprotéine de SARS-CoV-2 avec Prosite.                                                                                                                                                                                                                      | 33 |
| 21   | Résultats d'analyse de la séquence protéique de petite protéine d'enveloppe de SARS-CoV-2 avec Prosite.                                                                                                                                                                                                         | 35 |
| 22 A | Résultats d'analyse de la séquence protéique de la protéine de spicule de SARSCoV-2 avec Prosite.                                                                                                                                                                                                               | 36 |
| 22 B | Résultats d'analyse de la séquence protéique de la protéine de spicule de SARSCoV-2 avec Prosite.                                                                                                                                                                                                               | 37 |
| 22 C | Résultats d'analyse de la séquence protéique de la protéine de spicule de SARS-CoV-2 avec Prosite.                                                                                                                                                                                                              | 38 |
| 23 A | Résultats d'analyse de la séquence protéique de polyprotéine réplicase 1a de SARS-CoV-2 avec Prosite.                                                                                                                                                                                                           | 40 |
| 23 B | Résultats d'analyse de la séquence protéique de polyprotéine réplicase 1a de SARS-CoV-2 avec Prosite.                                                                                                                                                                                                           | 41 |
| 24 A | Résultats d'analyse de la séquence protéique de polyprotéine réplicase 1ab de SARS-CoV-2 avec Prosite.                                                                                                                                                                                                          | 45 |
| 24 B | Résultats d'analyse de la séquence protéique de polyprotéine réplicase 1ab de SARS-CoV-2 avec Prosite.                                                                                                                                                                                                          | 46 |
| 25   | Structure secondaire de la protéine spike de SARS-CoV-2 résultant suit à la <i>l'analyse in silico</i> par I-TASSER server.                                                                                                                                                                                     | 48 |
| 26   | Structures tridimensionnelles de la protéine de spicule de SARS-CoV-2. Les deux structures de haut sont les résultats de I-TASSER server ayant le score TM le plus élevé. Les deux structures d'en bas sont obtenus d'une manière in vitro. Les structures du haut sont comparables avec les structures du bas. | 49 |
| 27   | Graphe d'hydrophobicité de protéine membranaire de virus SARS-CoV-2.                                                                                                                                                                                                                                            | 50 |
| 28   | Graphe d'hydrophobicité de nucléoprotéine de virus SARS-CoV-2.                                                                                                                                                                                                                                                  | 50 |

|           |                                                                           |    |
|-----------|---------------------------------------------------------------------------|----|
| <b>29</b> | Graphe d'hydrophobicité de petite protéine d'enveloppe de virus SARS-CoV  | 51 |
| <b>30</b> | Graphe d'hydrophobicité de glycoprotéine de spicule de virus SARS-CoV-2.  | 51 |
| <b>31</b> | Graphe d'hydrophobicité de polyprotéine réplicase 1a de virus SARS-CoV-2. | 52 |
| <b>32</b> | Graphe d'hydrophobicité de polyprotéine réplicase 1ab de virus SARS-CoV   | 52 |

|                                     |    |
|-------------------------------------|----|
| Remerciements                       |    |
| Dédicaces                           |    |
| Résumé                              |    |
| Liste d'abréviation                 |    |
| Liste des figures                   |    |
| Liste des tableaux                  |    |
| Introduction                        | 1  |
| Partie I : Synthèse bibliographique |    |
| 1. Covid 19                         | 3  |
| 1.1. Historique                     | 3  |
| 1.2. Définition                     | 3  |
| 1.3. Symptômes du Covid             | 3  |
| 1.4. Diagnostic                     | 4  |
| 1.4.1. Diagnostic Clinique          | 4  |
| 1.4.2. Diagnostic biologique        | 4  |
| 1.5. Traitements de la COVID-19     | 5  |
| 1.6. Vaccins contre COVID-19        | 5  |
| 2. virus SARS.COVS2                 | 6  |
| 2.1. Définition                     | 6  |
| 2.2. Classification                 | 6  |
| 2.3. Structure du SARS-CoV-2        | 7  |
| 2.3.1. Génome                       | 8  |
| 2.3.2. Protéines                    | 9  |
| 2.3.2.1. Protéine spike (S)         | 9  |
| 2.3.2.2 Protéine E                  | 11 |

|                                                                                           |    |
|-------------------------------------------------------------------------------------------|----|
| 2.3.2.3. Protéine N                                                                       | 12 |
| 2.3.2.4. Protéine M                                                                       | 14 |
| 2.3.2.5. Relation entre les protéines S,E,N,M                                             | 15 |
| 2.3.2.6. Polyprotéines réplases 1ab, 1a                                                   | 16 |
| 2.3.2.7. Complexe réplase-transcriptase (RTC)                                             | 16 |
| Partie II : Matériel et méthodes                                                          |    |
| 2. Matériel et méthodes                                                                   | 19 |
| 2.1 Matériel                                                                              | 19 |
| 2.1.1. Séquences protéiques                                                               | 19 |
| 2.1.2. Applications bio-informatiques utilisées                                           | 25 |
| 2.2. Méthodes                                                                             | 25 |
| 2.2.1. Alignement en utilisant BLAST (Basic Local Alignment Search Tool)                  | 25 |
| 2.2.2. Recherche des domaines, motifs et sites fonctionnels en utilisant Prosite          | 26 |
| 2.2.3. I-TASSER server pour prédiction de structure bidimensionnelle et tridimensionnelle | 26 |
| 2.2.4. Prédiction des certaines caractéristiques physico-chimiques par le logiciel GPMW   | 26 |
| 3. Résultats et discussions                                                               | 27 |
| 3.1. Alignement avec BLAST (recherche de similarités et régions conservatives).           | 27 |
| 3.1.1. Protéine membranaire                                                               | 27 |
| 3.1.2. Nucléoprotéine                                                                     | 28 |
| 3.1.3. Petite protéine d'enveloppe                                                        | 29 |
| 3.1.4. Glycoprotéine de spicule (Spike)                                                   | 29 |
| 3.1.5. Polyprotéine réplase 1a                                                            | 30 |

|                                                                                       |    |
|---------------------------------------------------------------------------------------|----|
| 3.1.6. Polyprotéine réplicase 1ab                                                     | 31 |
| 3.2. Analyse des protéines via Prosite                                                | 32 |
| 3.2.1. Protéine membranaire                                                           | 32 |
| 3.2.2. Nucléoprotéine                                                                 | 33 |
| 3.2.3. Petite protéine d'enveloppe                                                    | 34 |
| 3.2.4. Glycoprotéine de spicule (Spike)                                               | 34 |
| 3.2.5. Polyprotéine réplicase 1a                                                      | 39 |
| 3.2.6. Polyprotéine réplicase 1ab                                                     | 42 |
| 3.3. Analyses des protéines via I-TASSER server                                       | 47 |
| 3.3.1. Glycoprotéine de spicule (spike glycoprotein)                                  | 47 |
| 3.4. Prédiction des certaines caractéristiques physico-chimiques par le logiciel GPMW | 50 |
| 3.4.1. Protéine membranaire                                                           | 50 |
| 3.4.2. Nucléoprotéine                                                                 | 50 |
| 3.4.3. Petite protéine d'enveloppe                                                    | 51 |
| 3.4.4. Glycoprotéine de spicule                                                       | 51 |
| 3.4.5. Polyprotéine réplicase 1a                                                      | 52 |
| 3.4.6. Polyprotéine réplicase 1ab                                                     | 52 |
| Conclusion générale                                                                   | 53 |
| Références bibliographiques                                                           | 55 |
| Annexe                                                                                |    |

# *Introduction*

## Introduction

Les virus constituent l'entité biologique la plus abondante dans les différents écosystèmes. Ce sont, on le sait, des parasites obligatoires capables d'infecter potentiellement toutes les cellules vivantes (de la cellule humaine à la bactérie) (PERSONNIC *et al.*, 2006). Un virus peut être simplement défini comme un parasite intracellulaire obligatoire. Chaque particule virale, ou virion, est constituée d'un seul acide nucléique, ARN ou ADN, codant pour le génome viral entouré d'une enveloppe protéique, et est capable de se répliquer uniquement dans les cellules vivantes de bactéries, d'animaux ou de plantes.

Les virus sont très contagieux et se répandent rapidement et peuvent dans certaines mesures aboutir à des épidémies ou même des pandémies, souvent se manifestant par des maladies bénignes ou des symptômes très graves.

Cependant, plus récemment, trois CoV ont causé de graves menaces pour la santé humaine mondiale : les épidémies de coronavirus du syndrome respiratoire aigu sévère (SRAS-CoV de 2002 à 2003) et de coronavirus du syndrome respiratoire du Moyen-Orient (MERS-CoV de 2012 à aujourd'hui) et le coronavirus du syndrome respiratoire aigu 2 (SRAS-CoV-2 de 2019 à aujourd'hui) (HUI, 2022).

Le SARS-CoV-2 (severe respiratory syndrome coronavirus-2) s'est propagé à l'échelle mondiale depuis l'épidémie initiale à Wuhan, en Chine, en novembre 2019 et a été déclaré agent causal de la pandémie de COVID-19 en mars 2020. Jusqu'en 20 juin 2022, dans le monde, il y avait 536 590 224 cas confirmés de COVID-19, dont 6 316 655 décès qui ont été signalés à l'OMS (OMS, 2022).

Le génome de l'ARN du SARS-CoV-2 compte 29 811 nucléotides, codant pour 29 protéines, ces protéines bien qu'elles soient structurelles (telles que : protéine spike, protéine membranaire, petite protéine d'enveloppe ou nucléoprotéines) ou ayant un rôle enzymatique ou agissant en tant que cofacteurs (telles que les Nsps...) peuvent toutes subir des mutations, ces mutations peuvent conduire à l'apparition des nouveaux variants de SARS-CoV-2. Ainsi, plusieurs variants préoccupants et variants d'intérêt ont depuis été découvertes avec des mutations dans le protéome et qui ont entraîné une virulence et une transmissibilité différentes de la souche initiale. De nombreuses mutations enregistrées, en particulier dans ceux des variants préoccupants et variants d'intérêt, ont été décrites comme se produisant dans la glycoprotéine de pointe (spike) (BEAUDOIN *et al.*, 2022).

Les analyses bio-informatiques (analyses in silico) des protéines de SARS-CoV-2 ont beaucoup contribué d'avancer rapidement les recherches sur : l'identification des nouvelles mutations au niveau des gènes qui codent pour les protéines de SARS-CoV-2, l'identification des nouveaux variants de SARS-CoV-2, la détermination des régions conservatives responsables du fonctionnement des protéines, la détermination des sites de clivage protéolytique responsable de la maturation des certaines protéines et responsable également du bon mécanisme de fusion de l'enveloppe virale de SARS-CoV-2 et la membrane cellulaire de pneumocyte, l'établissement des arbres phylogénétiques et recherche d'ancêtres de coronavirus ce qui aide à la bonne classification taxonomique des virus CoV actuels et les futures sous genre et espèces et variants, l'établissement de la voie correcte de mécanisme de l'interaction avec la cellule hôte et de la multiplication virale, ainsi que l'autoassemblage des nouveaux virions, la conception des vaccins actuels et des nouveaux vaccins, l'invention des médicaments anti-SARS-CoV-2 et etc.

Notre étude est conçue pour contribuer aux analyses in silico des protéines de SARS-CoV-2 et qui prend, sans doute, sa place dans le contexte déjà décrit par le paragraphe précédent, comme il va servir, également, de plateforme pour des futures recherches dans ce domaine.

# *Synthèses bibliographiques*

## 1. Covid 19

### 1.1. Historique

Les coronavirus sont connus dans le monde des vétérinaires depuis les années 1930. Lors de l'identification des premiers coronavirus humains HCoV-OC43 et - 229E dans les années 1960, une vingtaine de coronavirus infectant des espèces animales aviaires (poulet) et mammifères (chien, chat, porc, bovin, etc.) étaient déjà décrits. Les coronavirus doivent leur nom à leur aspect en microscopie électronique avec des spicules formant une couronne autour de la particule virale (KIN *et al.*, 2016). Le syndrome respiratoire aigu sévère (SRAS) a été reconnu pour la première fois comme menace mondiale en mars 2003. Les premiers cas connus de SRAS sont survenus dans la province du Guangdong, en Chine, en novembre 2002. L'agent pathogène de coronavirus du SRAS (SRAS-CoV) retrouvé chez l'homme serait un virus animal. En juillet 2003, la propagation au niveau mondial du SRAS-CoV avait entraîné 8098 cas de SRAS dans 26 pays, avec 774 décès (BITAR, *et al.*, 2003). Par la suite, l'OMS a notifié 2066 cas confirmés en laboratoire d'infection par le MERS-CoV, avec au moins 720 décès entre 2012 et 2017. Bien que retrouvées dans 27 pays, plus de 80% des maladies étaient rapportées d'Arabie saoudite (TRAORE, 2021). En décembre 2019, Wuhan, dans la province du Hubei, en Chine, est devenue le centre d'une épidémie de pneumonie de cause inconnue, qui a suscité une intense attention non seulement en Chine mais au niveau international. Le 7 janvier 2020, des scientifiques chinois isolaient un nouveau coronavirus 2019 (2019-nCoV) ce qui permettait le développement rapide du test de diagnostic PCR en temps réel (RT-PCR) (PARODI, 2021).

### 1.2. Définition

La maladie à coronavirus (COVID-19) est une maladie infectieuse émergente due à un groupe de virus à ARN appartenant à la famille des *Coronaviridae*. Le virus doit son nom à l'apparence de ses particules virales qui portent des excroissances évoquant une couronne. Il se transmet par contact direct avec des gouttelettes respiratoires (en toussant et/ou en éternuant) ou en touchant des surfaces contaminées provoquant des infections digestives et respiratoires (MERYAM, *et al.*, 2021).

- “Co” pour “corona”;
- “vi” pour “virus”;
- “D” pour “disease” (« maladie » en anglais) ;
- “19” pour l'année de son apparition : 2019.

### 1.3. Symptômes du Covid

Les symptômes principaux de la maladie sont la fièvre, la fatigue et une toux sèche. Certains patients ont aussi présenté des douleurs, une congestion et un écoulement nasal, des

maux de gorge et une diarrhée (JOSSE, 2020). Ces symptômes sont généralement bénins. Mais environ une personne sur six présente des symptômes plus sévères, notamment la dyspnée. La pneumonie est la complication la plus fréquente du Covid-19. Il existe aussi des cas asymptomatiques, c'est-à-dire que les patients n'ont aucun symptôme apparent malgré la détection du virus. Des symptômes très variables pour le SRAS-CoV et le MERS-CoV (EL OUAZZANI, 2021)

## **1.4. Diagnostic**

### **1.4.1. Diagnostic clinique**

Il dépend de l'évolution de la maladie justifiant une prise en charge. L'incubation dans la majorité des cas est de quatre à cinq jours, presque toujours comprise entre deux et onze jours, avec une durée maximale estimée à quatorze jours. (ZHOU, 2020)

La COVID-19 est maintenant classée en quatre niveaux basés sur la gravité des symptômes : légère, modérée, sévère et critique (WANG, 2020).

- Les cas légers ne présentent que des symptômes bénins sans caractéristique radiographique. (HAMIDI *et al.*, 2021) ;
- Les cas modérés présentent de la fièvre, des symptômes respiratoires et des caractéristiques radiographiques ;
- Les cas sont dits sévères s'ils répondent à l'un des trois critères suivants : Dyspnée, fréquence respiratoire supérieur à 30 fois / min, saturation en oxygène inférieure à 93% dans l'air ambiant et PaO<sub>2</sub> /FiO<sub>2</sub> moins de 300 mm Hg ;
- Les cas sont dits critiques s'ils répondent à l'un des trois critères suivants : Insuffisance respiratoire, choc septique et défaillance d'organes multiples.

### **1.4.2. Diagnostic biologique**

Depuis l'avènement de la COVID-19 à travers le monde, en plus des signes cliniques, les marqueurs biologiques et de l'imagerie ont contribué énormément au diagnostic de la maladie. Le principe de ces tests repose soit sur la détection de l'antigène ou de l'anticorps ou certains éléments caractéristiques de la tomodensitométrie. La confirmation de la COVID-19 est faite par l'identification de l'ARN du SARS-CoV-2 dans des échantillons biologiques dont l'un des piliers est la détection du génome viral dans les voies aériennes supérieures (oro/nasopharyngées) pour le diagnostic précoce de l'infection. Elle se fait concrètement en analysant la présence du virus au sein d'un écouvillon nasopharyngé prélevé chez un patient suspect (RABAAN *et al.*, 2021).

La réaction de polymérisation en chaîne utilisant la transcriptase inverse (RT-PCR) sur prélèvement rhino-pharyngé est la méthode de référence pour le diagnostic de l'infection à

SARS-CoV-2, mais les tests antigéniques présentent l'avantage d'une mise en œuvre plus simple et plus rapide (KLAUS *et al.*, 2021).

Les tests de détection des anticorps sériques (ELISA) sont utilisés dans la phase ultérieure et après la guérison.

En l'absence de test de référence pour la confirmation ou dans les cas des signes évocateurs avec un résultat de RT-PCR négatif, la sérologie peut être effectuée avec le dosage des IgM et des IgG. La sérologie est également un outil pertinent pour les études épidémiologiques. Néanmoins, il est important de rappeler que le taux d'anticorps anti-SARS-CoV-2 décroît avec le temps et peut impacter les résultats d'études basés sur la séroprévalence (JAMAI, 2020).

### **1.5. Traitements de la COVID-19**

Il n'existe actuellement aucun traitement antiviral spécifique homologué pour les infections par le SARS-CoV-2, et l'objectif principal en milieu clinique reste la réduction des signes cliniques et la fourniture de soins de soutien. Il faudra probablement plusieurs années pour développer et évaluer les médicaments spécifiques pour traiter le nouveau coronavirus. Compte tenu de l'urgence de la demande clinique, de nombreux médicaments sont approuvés pour être utilisés dans le cadre d'essais cliniques contre l'infection par le SARS-CoV-2, tels que le lopinavir/ritonavir, l'arbidol, l'interféron-alpha, le favipiravir, le phosphate de chloroquine, le darunavir/cobicistat, l'oseltamivir et la méthylprednisolone (TRAORE, 2021).

L'hydroxychloroquine en association avec l'azithromycine ont montré aussi leurs efficacités chez les patients atteints de COVID-19 en Chine et en France (LIU, 2020).

### **1.6. Vaccins contre COVID-19**

La vaccination permet de se protéger et de protéger les autres. Couplé avec les mesures barrières, le vaccin contribuera à maîtriser l'impact de l'épidémie de la Covid-19 sur le long terme. Les premiers objectifs du programme de vaccination seront de réduire la morbidité et la mortalité attribuables à la maladie (hospitalisations, admissions en soins intensifs et décès) et de maintenir les activités essentielles du pays, particulièrement celles du système de santé pendant l'épidémie (OMS, 2021).

En août 2021, selon l'Organisation mondiale de la santé (OMS), il y aurait 110 vaccins contre le coronavirus SARS-CoV-2 autorisés ou en phase d'étude clinique, ainsi que 184 vaccins potentiels à l'étude. Plusieurs vaccins étudiés au cours des essais cliniques de phase III ont montré une efficacité allant jusqu'à 95 %. 21 vaccins sont approuvés par au moins une autorité nationale pour administration au public (OMS, 2021)

- Deux vaccins à ARN : Pfizer-BioNTech et Moderna ;

- Cinq vaccins à vecteur viral : Spoutnik V, Spoutnik Light, Oxford–AstraZeneca, Convidecia et Janssen ;
- Cinq vaccins de sous-unité protéique : EpiVacCorona, ZF2001, Abdala, SOBERANA 02 et MVC-COV1901 ;
- Neuf vaccins à virus inactivé : BBIBP-Corv, WIBP-CorV, CoronaVac, Covaxin, CoviVac, Covidful, KCONVAC, COVIran Barekat et QazCovid-in (MULANGU, 2019).

## 2. Virus de SARS-COV-2

### 2.1. Définition

Le SARS-Coronavirus-2 (SARS-CoV-2) est l'agent pathogène qui provoque la maladie nommée COVID-19 (Figure 01) (ARS, 2020). Le SARS-CoV-2 est un virus enveloppé à capside hélicoïdale dont le génome est constitué d'ARN simple brin (monocaténaire) (HAS, 2020).



**Figure 01 :** Le coronavirus SARS-CoV-2 vu au microscope électronique (HAS, 2020).

### 2.2. Classification

Les coronavirus sont des virus à ARN monocaténaire de sens positif (groupe IV de la classification Baltimore) correspondant à la sous-famille Orthocoronavirinae de la taxonomie de l'ICTV, dans la famille Coronaviridae, et de l'ordre Nidovirales. Selon les caractéristiques de leurs séquences protéiques, les CoV sont classés en 4 genres (alpha-CoV, beta-CoV, gamma-

CoV et delta-CoV), qui tous contiennent des virus pathogènes pour les mammifères (ZI-WEI *et al.*, 2020) :

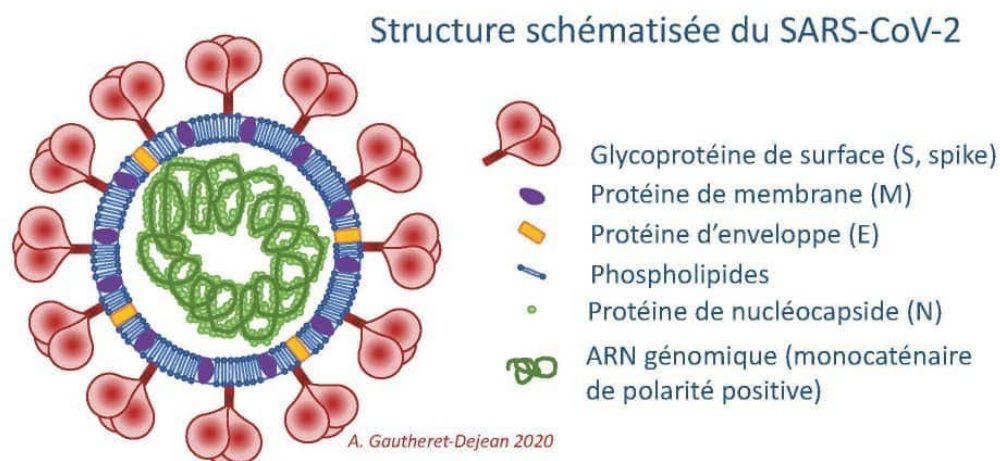
- *Alphacoronavirus*, qui inclut le virus de la diarrhée épidémique porcine (PEDv), le virus de la gastro-entérite transmissible porcine (TGEV), le coronavirus du syndrome de la diarrhée aiguë porcine (SADS-CoV), le coronavirus canin, le coronavirus entérique félin, le virus de la péritonite infectieuse féline (FIPV) (MERTENS *et al.*, 2020) ;
- *Betacoronavirus*, dont le virus respiratoire-respiratoire du SRAS (SARS-CoV), le SARS-CoV-2, le coronavirus du syndrome respiratoire du Moyen-Orient (MERS-CoV), virus de l'hépatite murine (MHV), coronavirus bovins, virus de la sialodacryoadénite du rat, virus de la sialodacryoadénite porcine, hémagglutinoxe porcine, virus de l'hémagglutinoxe porcine coronavirus équin. Dans ce genre *Betacoronavirus*, le SARS-CoV et le SARS-CoV-2 appartiennent tous les deux au sous-genre *Sarbecovirus* au sein duquel trois clades distincts ont été identifiés : Clade 1: souches "chauve-souris" de Bulgarie et Kenya, Clade 2: SARS-CoV-2 et souches "chauve-souris" de Chine orientale, Clade3: SARS-CoV et souches "chauves-souris" de Chine du sud-ouest ;
- *Gammacoronavirus*, surtout trouvé chez des oiseaux migrateurs, causant notamment des bronchites ; un *Gammacoronavirus* a été isolé d'un béluga en captivité ;
- *Deltacoronavirus*, connus depuis peu, qui semblent surtout infecter les oiseaux, mais aussi trouvé chez les porcs (DELAPLACE, 2021).

### 2.3. Structure du SARS-CoV-2

Le génome ARN des coronavirus, qui, avec une longueur médiane de 29 kb, est le plus long de tous les virus à ARN, est composé de six à dix cadres de lecture ouverts (ORF) qui sont responsables du codage à la fois de la réplicase et des protéines structurales du virus. Chacun des composants du génome viral est conditionné dans une nucléocapside hélicoïdale entourée d'une bicouche lipidique. L'enveloppe virale des coronavirus est généralement composée de trois protéines qui comprennent la protéine membranaire (M), la protéine d'enveloppe (E) et la protéine de pointe (S). (WAHIDA *et al.*, 2021).

Le génome du SARS-CoV-2 code pour d'importantes protéines qui serviront à établir sa structure ou sa réplication dans la cellule de l'hôte. De façon générale, on peut résumer le cycle de vie du virus SARS-CoV-2 de cette façon. Tout d'abord la protéine de surface du virus,

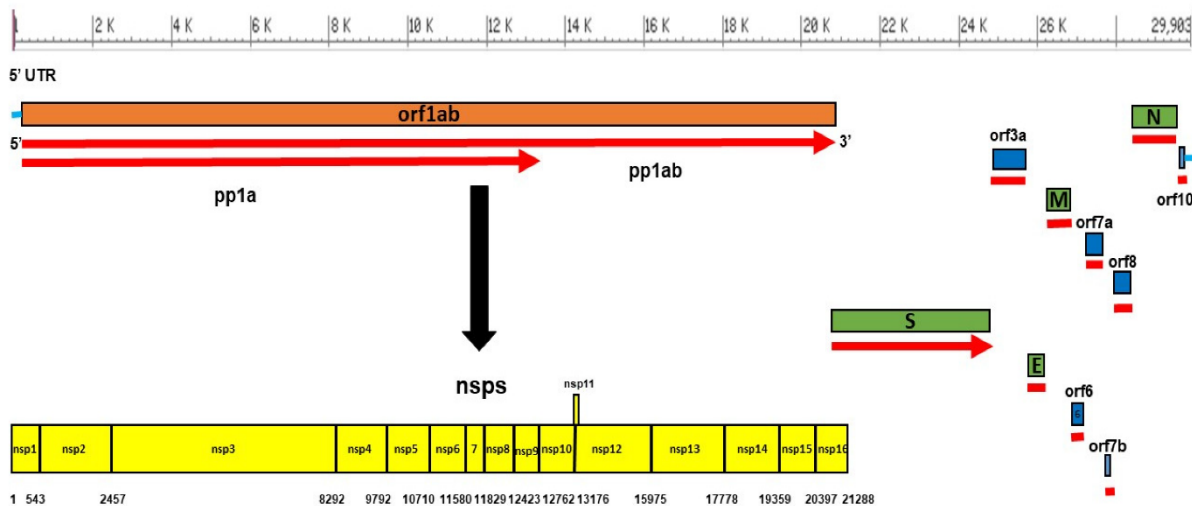
la protéine S va lier le récepteur ACE2 à la surface de la membrane cellulaire des pneumocytes. Lors de cette liaison, un changement de conformation de la protéine S va favoriser la fusion entre le SARS-CoV-2 et la membrane cellulaire afin d'atteindre le système endosomal de la cellule. Le virus SARS-CoV-2 pourra alors libérer son ARN afin de pouvoir débiter la traduction des protéines. Le génome du bêta coronavirus SARS-CoV-2 est assez particulier puisqu'il contient plusieurs cadres de lecture ouvert (open reading frame), en l'occurrence ORF1a et ORF1b qui vont permettre la synthèse de deux grandes polyprotéines 1a et 1ab qui seront réduites en plusieurs fragments (polypeptides) par des protéases virales. Ces polypeptides permettront la synthèse de l'ARN viral dans l'hôte (MIHOUBI *et al.*, 2021). L'ARN sera ensuite transcrit en ARNm et traduit en protéines qui permettront l'assemblage d'un virion dans le système réticulo-endoplasmique de la cellule hôte, la production de vésicules contenant des virions et l'exocytose afin de libérer le virus (WAHIDA *et al.*, 2021)



**Figure 02 :** Structure schématisée du SARS-COV-2 (MIHOUBI *et al.*, 2021).

### 2.3.1. Génome

La figure 03 montre l'organisation du génome ainsi que la taille du génome qui peut varier de 26 à 32 kilobases (kbp) et contient de 6 à 11 cadres de lecture ouverts (ORFs) qui encodent 9680 acides aminés. Le premier ORF représente environ 67 % (près des deux tiers) du génome et code 16 protéines non structurales (nsp) (SINGH, 2020). Ces NSP sont nécessaires à la réplication virale (DE WILDE *et al.*, 2017) tels que la protéase de type 3-chymotrypsine, la protéase de type papaine et l'ARN polymérase dépendante de l'ARN, alors que les protéines de structure du virus (protéine spike (S), hémagglutinine-estérase (HE) protéines de membrane (M), protéine de capsid (N) sont codées par le tiers restant du génome (Bonny *et al.*, 2020)



**Figure 03 :** Organisation du génome du SARS-CoV-2 (HELMY *et al.*, 2020).

## 2.3.2. Protéines

### 2.3.2.1. Protéine spike (S)

La protéine de coronavirus S (figure 04) est une grande protéine transmembranaire virale multifonctionnelle de classe I (une seule hélice transmembranaire). La taille de cette abondante protéine varie de 1160 acides aminés (chez la volaille l'IBV) à 1400 acides aminés (F-CoV, coronavirus félin). Elle paraît dans un trimère sur la surface du virion, donnant au virion un aspect corona ou en forme de couronne. (KHAMMOUM *et al.*, 2021).

Fonctionnellement, Elle est nécessaire pour l'entrée des particules de virion infectieux dans la cellule par interaction avec divers récepteurs cellulaires hôtes. En outre, elle agit comme un facteur critique pour le tropisme tissulaire et la détermination de gamme d'hôtes. (LORIN, 2020)

Notamment, la protéine S est l'une des protéines immunodominantes vitales des CoV capable d'induire des réponses immunitaires de l'hôte.

Les éctodomains dans toutes les protéines S des CoV ont des organisations de domaines similaires, divisées en deux sous-unités, S1 et S2. Le premier, S1, contribue à la liaison au récepteur de l'hôte, tandis que le second, S2, interagit dans la fusion. Le premier (S1) est divisé en deux sous-domaines, à savoir le N-terminal domaine (NTD) et domaine C-terminal (CTD). Ces deux sous-domaines agissent comme des domaines de liaison aux récepteurs, interagissant efficacement avec divers récepteurs hôtes. Le S1 CTD contient le motif de liaison au récepteur (RBM). Dans chaque protéine de pointe de coronavirus, le trimère S1 se situe au sommet de la trimère tige S2. (DHAMA *et al.*, 2020).

Le clivage protéolytique est une modification post-traductionnelle irréversible qui entraîne des changements de conformation créant l'apparition ou la disparition de propriétés biologiques (BONNIN, 2018). Les protéases sont des enzymes capables de cliver les liaisons peptidiques des protéines. Elles peuvent intervenir dans de nombreux processus biologiques, comme la maturation des protéines ou la digestion des aliments et sont présentes au niveau des endosomes intracellulaires ou à la surface cellulaire. On distingue les exoprotéases, qui coupent le premier ou le dernier acide aminé d'une chaîne peptidique, et les endoprotéases, qui sont capables de couper dans la chaîne peptidique. Il est possible de regrouper les protéases en fonction de leur nature ou des acides aminés de leur site actif, comme pour les protéases à cystéine ou les protéases à sérine (DOUCET, 2006). Comme leur nom l'indique, les protéases à cystéine possèdent une cystéine dans leur site actif tandis que les protéases à sérine possèdent trois acides aminés dans leur site catalytique, dont la sérine. De nombreuses protéases clivent en fonction des acides aminés entourant le site de coupure : elles reconnaissent un motif précis de la chaîne polypeptidique la plupart des coronavirus ont besoin de protéases cellulaires pour activer la fusion de leur enveloppe virale avec les membranes des cellules cibles (STITRA *et al.*, 2008). Les coronavirus peuvent utiliser des protéases endosomales, comme les cathepsines (L et B) ou la furine ou des protéases de la surface cellulaire comme TMPRSS2 (Millet *et al.*, 2015). Les mécanismes d'entrée du SARS-CoV ont été beaucoup étudiés. Très rapidement, il a été montré que l'entrée était inhibée par des agents lysomotropiques mais que leur action était indirecte et liée à l'inhibition de la cathepsine L, une protéase endosomale active à pH acide (SIMMONS *et al.*, 2004 ; MATSUYAMA *et al.*, 2005). Il a ensuite été montré que l'entrée virale par la voie endosomale pouvait être court-circuitée lorsque les virus liés à la surface des cellules étaient exposés à des protéases exogènes telles que la trypsine ou l'élastase. Pour le SARS-CoV, la voie d'entrée directe par action de la trypsine est 100 à 1000 fois plus efficace que l'entrée virale par voie endosomale par action de la cathepsine. D'autre part, en modèle de fusion cellule-cellule, la fusion des cellules exprimant la protéine S du SARS-CoV à leur surface peut être induite par un traitement par la trypsine. L'ensemble de ces données a donc rapidement indiqué que la fusion du SARS-CoV était stimulée par un clivage protéolytique. Ces mécanismes ont d'abord été étudiés en utilisant la trypsine, comme protéase modèle. Ainsi il a été montré que la fusion virale a lieu après un clivage séquentiel au niveau de 2 sites de la protéine S : il y a un premier clivage au niveau de l'arginine 667 à la jonction entre S1 et S2 qui semble faciliter un second clivage au niveau de l'arginine 797 située dans une région appelée S2', responsable de l'activation de la fusion. En effet, ce deuxième clivage a lieu au niveau N-terminal d'une région très conservée parmi les coronavirus qui présente les

caractéristiques d'un peptide de fusion (ESCHENBRENNER, 2019). La protéine S du SARS-CoV peut également être clivée par d'autres protéases comme l'élastase qui permet un clivage de quelques acides aminés en amont du site S2' du SARS-CoV, suggérant que la position du clivage en S2' permettant l'activation de la fusion présente une certaine flexibilité. La trypsine utilisée comme outil n'est pas très pertinente du point de vue physiologique. Ainsi des travaux ont déterminé quelles protéases pulmonaires de type sérine-protéase pourraient être impliquées dans l'infection du SARS-CoV.

|      |            |            |            |            |            |            |            |            |
|------|------------|------------|------------|------------|------------|------------|------------|------------|
| 1    | MFVFLVLLPL | VSSQCVNLT  | RTQLPPAYTN | SFTRGVVYPD | KVFRSSVLHS | TQDLFLPFFS | NVTWFHAIHV | SGTNGTKRFD |
| 81   | NPVLPFNDGV | YFASTEKSN  | IRGWIFGTTL | DSKTQSLLIV | NNATNVVIKV | CEFQFCNDPF | LGVYYHKNK  | SWMESEFRVY |
| 161  | SSANNCTFEY | VSQPFLMDLE | GKQGNFKNLR | EFVFNIDGY  | FKIYSKHTPI | NLVRDLPPGF | SALEPLVDLP | IGINITRFQT |
| 241  | LLALHRSYLT | PGDSSSGWTA | GAAAYVGYL  | QPRTFLLKYN | ENGTITDAVD | CALDPLSETK | CTLKSFTVEK | GIYQTSNFRV |
| 321  | QPTESIVRFP | NITNLCPFGE | VFNATRFASV | YAWNKRKRN  | CVADYSVLN  | SASFSTFKCY | GVSPTKLNDL | CFTNVYADSF |
| 401  | VIRGDEVQRV | APGQTGKIAD | YNYKLDDFT  | GCVIWNSNN  | LDKVVGGNYN | YLYRLFRRSN | LKPFERDIST | EIYQAGSTPC |
| 481  | NGVEGFNCYF | PLQSYGFQPT | NGVGYPYRV  | VVLSFELLHA | PATVCGPKKS | TNLVKNKCVN | FNFNGLTGTG | VLTESNKKFL |
| 561  | PFQQFGRDIA | DTTDAVRDPQ | TLEILDITPC | SFGGVSVITP | GTNTSNQVAV | LYQDVNCTEV | PVAIHADQLT | PTWRVYSTGS |
| 641  | NVFQTRAGCL | IGAETHVNSY | ECDIPIGAGI | CASYQTQTN  | PRRARSVASQ | SIIAYTMSLG | AENSVAYSNN | SIAIPTNFTI |
| 721  | SVTTEILPVS | MTKTSVDCTM | YICGDSTEC  | NLLQYGSFC  | TQLNRALTGI | AVEQDKNTQE | VFAQVKQIYK | TPPIKDFGGF |
| 801  | NFSQILPDPS | KPSKRSFIED | LLFNKVTLD  | AGFIKQYGD  | LGDIARDLI  | CAQKFNGLTV | LPPLLTDEMI | AQYTSALLAG |
| 881  | TITSGWTFGA | GAALQIPFAM | QMAYRFNGIG | VTQNVLYENQ | KLIANQFNFA | IGKIQDSLSS | TASALGKLQD | VVNQNAQALN |
| 961  | TLVKQLSSNF | GAISSVLNDI | LSRLDKVEAE | VQIDRLITGR | LQSLQTYVTQ | QLIRAAEIRA | SANLAATKMS | ECVLGQSKRV |
| 1041 | DFCGKGYHLM | SFPQSAPHGV | VFLHVTYVPA | QEKNFTTAPA | ICHDGKAHFP | REGVVFVNGT | HWFTVQRNFY | EPQIITDNT  |
| 1121 | FVSGNCDVVI | GIVNNTVYDP | LQPELDSFKE | ELDKYFKNHT | SPDVLGDIS  | GINASVVNIQ | KEIDRLNEVA | KNLNSLIDL  |
| 1201 | QELGKYEQYI | KWPWYIWLGF | IAGLIAIVMV | TIMLCCMTSC | CSCLKGCCSC | GSCCKFDEDD | SEPVKGVKL  | HYT        |

**Figure 04 :** Séquence de protéine S (Source : Uniprot, code d'accès : P0DTC2-1).

### 2.3.2.2 Protéine E

La protéine du coronavirus E (figure 05) dont la taille varie de 8,4 à 12 KDa, est la plus énigmatique et la plus petite des principales protéines. Elle joue un rôle multifonctionnel dans la pathogenèse, l'assemblage et la libération du virus (BOUZIDI, 2021).

C'est un petit polypeptide membranaire intégral qui agit comme une viroporine (ion canal). L'inactivation ou l'absence de cette protéine est liée à la virulence altérée des coronavirus due à des changements de morphologie et de tropisme (ANZI, 2021).

La protéine E se compose de trois domaines, à savoir une région terminale aminée hydrophile (7-12 acides aminés), un grand domaine transmembranaire hydrophobe (25 acides aminés) et un domaine C-terminal efficace (plus de 30 acides aminés dont les 4 derniers forment le motif PBM signe de mutation et d'altération). La protéine E du SARS-CoV-2 révèle une constitution d'acides aminés similaire sans aucune substitution. (DHAMA *et al.*, 2020).

La protéine E est conservée parmi les souches de coronavirus. Il s'agit d'une protéine membranaire intégrale impliquée dans plusieurs aspects du cycle de vie du virus, tels que l'assemblage, le bourgeonnement, la formation de l'enveloppe et la pathogenèse. Au cours du cycle de réplication, la protéine E est abondamment exprimée à l'intérieur de la cellule infectée,

mais seule une petite partie est incorporée dans l'enveloppe du virus. La majorité de la protéine participe à l'assemblage viral et au bourgeonnement. Il peut agir comme une viroporine en oligomérisant après insertion dans les membranes hôtes pour créer un pore hydrophile qui permet le transport des ions (RIM, 2021).

La protéine E de SARS-CoV forme un canal perméable au  $\text{Ca}^{2+}$  dans le compartiment intermédiaire de l'appareil de Golgi du réticulum endoplasmique. L'activité du canal ionique de la protéine E modifie l'homéostasie du  $\text{Ca}^{2+}$  dans les cellules, stimulant l'activation de l'inflammasome NLRP3, ce qui conduit à la surproduction d'IL-1beta. Le SRAS-CoV surstimule la voie inflammatoire NF-kappaB et interagit avec la protéine cellulaire synténine, déclenchant l'activation de p38 MARK. Ces cascades de signalisation entraînent une inflammation et une immunopathologie exacerbées (LE COUPANEC, 2018).

1 MYSFVSEETG TLIVNSVLLF LAFVVFLLVT LAILTALRLC AYCCNIVNVS LVKPSFYVYS RVKNLNSSRV  
71 PDLLV

**Figure 05 :** Séquence Protéine E (Source : Uniprot, code d'accension : P0DTC4-1)

### 2.3.2.3. Protéine N

Beaucoup d'études sont axées sur la caractérisation de cette protéine car elle est cruciale dans la composition du virus. La protéine N du coronavirus est (figure 06) polyvalente formée de 422 acide aminé avec un poids moléculaire de 46 KDa. Parmi plusieurs fonctions, elle joue un rôle dans la formation complexe avec le génome viral, facilite l'interaction des protéines M nécessaire lors de l'assemblage du virion et améliore l'efficacité de la transcription du virus. Elle contient trois domaines distincts et hautement conservés, à savoir, un NTD, Un domaine de liaison à l'ARN (LKR) et un CTD (KHAMMOUM *et al.*, 2021).

Le NTD se lie avec la fin 3' du génome viral, peut-être via des interactions électrostatiques, et est fortement divergé à la fois en longueur et en séquence.

Le LKR chargé est riche en sérine et en arginine et est également connu sous le nom de domaine SR (sérine et arginine). Le LKR est capable d'interagir directement avec l'interaction ARN in vitro et est responsable de la signalisation cellulaire. Il module également la réponse antivirale de l'hôte en agissant comme antagoniste d'interféron (IFN) et interférence ARN. (KHAMMOUM *et al.*, 2021).

Le CTD est situé en position C-terminale et permet l'auto-assemblage de la protéine. Comparant à celle du SRAS-CoV, la protéine N du SRAS-CoV-2 possèdent cinq mutations d'acides aminés, dont deux sont intrinsèquement région dispersée (IDR; positions 25 et 26), une

dans chaque NTD (position 103), LKR (position 217) et CTD (position 334) (figure 07) (DHAMA *et al.*, 2020).

La protéine N est constituée de deux domaines structurés séparés par un long lieu flexible, donnant lieu à un degré élevé de liberté conformationnelle. Les domaines séparés permettent à la protéine N d'effectuer de nombreuses fonctions dans le cycle de vie viral telles que la réplication / transcription de l'ARN, l'assemblage du virion et l'interférence du système immunitaire. Fait intéressant, la nucléocapside et les protéines de pointe sont les principaux immunogènes chez les patients COVID-19.

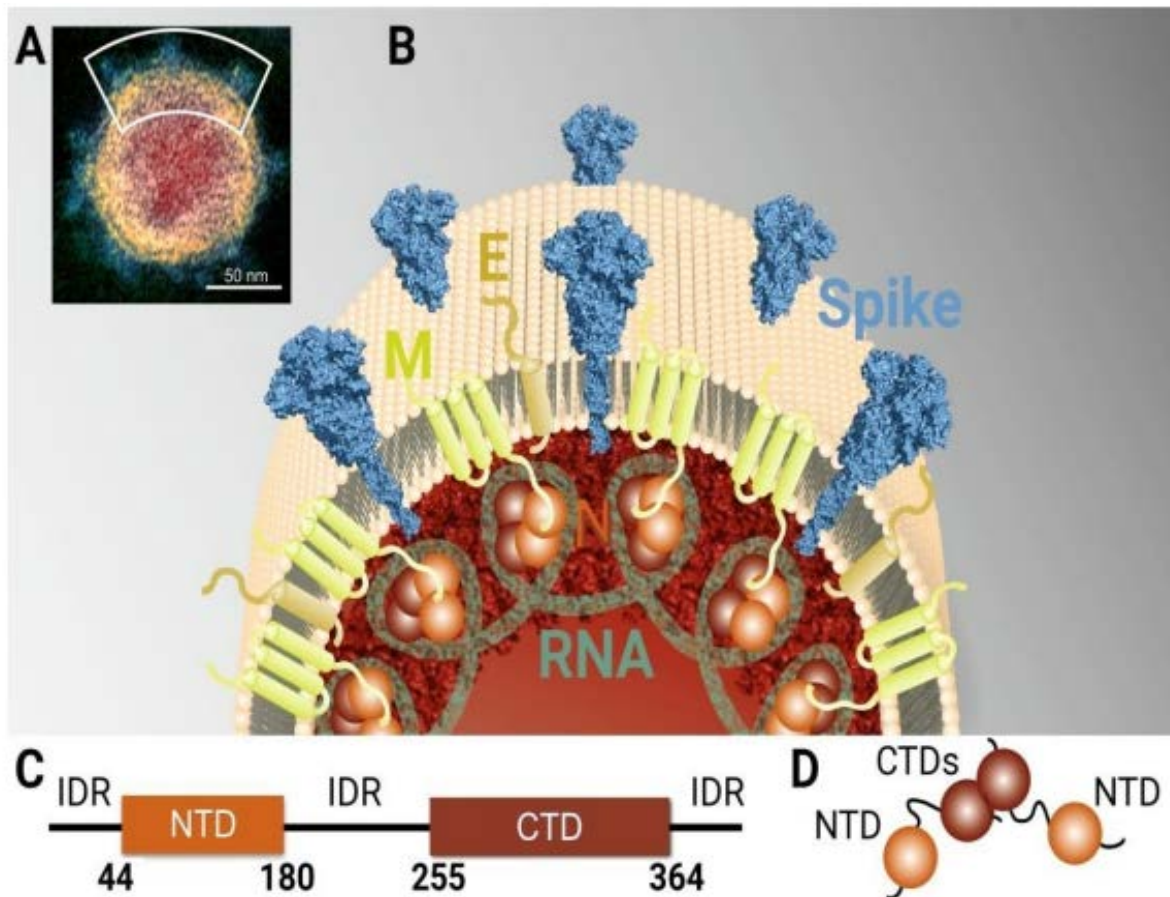
La protéine N comprend plusieurs régions intéressantes. En voici quelques-unes :

- 1-235= site liaison de l'ARN viral ;
- 62= site de sumoylation ;
- 184-196= motif riche en SR (sérine et arginine) et site de liaison d'une autre protéine N au niveau des acides aminés ;
- 211- 290 (région centrale) ;
- 209= site d'auto-liaison ;
- 248-365= site puissant de liaison à l'ARN viral ;
- 283-422= permet son association en dimère, trimère, tétramère et hexamère ;
- 343-402= séquence essentielle mais aussi suffisante pour l'oligomérisation ;
- 363-382= séquence responsable de l'association avec l'acide nucléique (ARN).

|     |            |             |            |            |            |            |            |
|-----|------------|-------------|------------|------------|------------|------------|------------|
| 1   | MSDNGPQNQR | NAPRITFGGP  | SDSTGSNQNG | ERSGARSKQR | RPQGLPNNTA | SWFTALTQHG | KEDLKFPGRQ |
| 71  | GVPINTNSSP | DDQIGYYRRA  | TRRIRGGDGK | MKDLSRWYF  | YYLGTGPEAG | LPYGANKDGI | IWVATEGALN |
| 141 | TPKDHIGTRN | PANNAIVLQ   | LPQGTTLPG  | FYAEGSRGGS | QASSRSSRS  | RNSSRNSTPG | SSRGTSARM  |
| 211 | AGNGGDAALA | LLLLDRLNQL  | ESKMSGKGQQ | QQGQTVTKKS | AAEASKKPRQ | KRTATKAYNV | TQAFGRRGPE |
| 281 | QTQGNFGDQE | LIRQGTDYKH  | WPQIAQFAPS | ASAFFGMSRI | GMEVTPSGTW | LTYTGAIKLD | DKDPNFKDQV |
| 351 | ILLNKHIDAY | KTFPPTPEPKK | DKKKKADETQ | ALPQRQKKQQ | TVTLLPAADL | DDFSKQLQQS | MSSADSTQA  |

**Figure 06 :** Séquence de protéine N (Source : Uniprot, code d'accèsion : P0DTC9-1)

Les efforts pour élucider la structure de la protéine N visent à comprendre comment les domaines individuels interagissent avec l'ARN et d'autres protéines. Cependant, la protéine N est très sensible à la protéolyse en raison de la région de liaison intrinsèquement désordonnée et, par conséquent, la structure de la protéine de nucléocapside pleine longueur reste insaisissable. La protéolyse est une stratégie clé dans la prolifération virale et engendre souvent un changement dans la fonction des protéines (COULIBALY *et al.*, 2019).



**Figure 07 :** (A) Image microscopique électronique à transmission d'une seule particule virale SARS-CoV-2. (B) Modèle 2D agrandi de la membrane virale montrant les quatre protéines structurales : glycoprotéine spike S, protéine de membrane M, protéine d'enveloppe E et nucléoprotéine N, avec membrane virale et le génome de l'ARN. (C) Organisation de domaine de la protéine N pleine longueur montrant les régions structurales sous forme de boîtes (NTD et CTD) et les régions intrinsèquement désordonnées (IDR) sous forme de ligne. (D) Modèle schématique du dimère de protéine N pleine longueur formé à travers les domaines CTD (le N-NTD est montré en brun et le N-CTD en brun foncé) (COULIBALY *et al.*, 2019).

#### 2.3.2.4. Protéine M

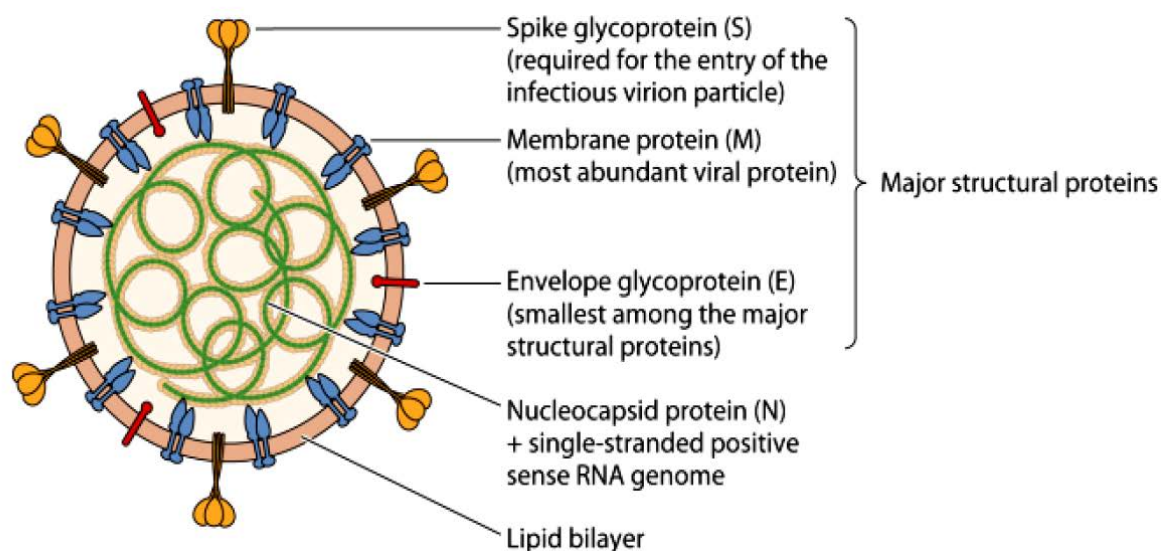
La protéine M (figure 08) est la protéine virale la plus abondante présente dans la particule de virion, donnant une forme définie à l'enveloppe virale. Il se lie à la nucléocapside et agit comme un organisateur de l'assemblage des coronavirus (figure 09). Les protéines du coronavirus M sont très diverses en teneur aminoacide mais conservent une similitude structurelle globale au sein de différents genres. La protéine a trois domaines transmembranaires (JUCKEL *et al.*, 2020), flanqués d'une courte terminaison amine à l'extérieur du virion et une longue terminaison carboxyle à l'intérieur du virion. Dans l'ensemble, l'échafaudage viral est maintenu par l'interaction M-M. Il est à noter que la protéine

M du SRAS-CoV-2 n'a pas de substitution d'acides aminés par rapport à celle du SRAS-CoV. (DHAMA *et al.*, 2020).

La protéine M est également considérée comme l'organisateur central de l'assemblage du CoV, interagissant avec toutes les autres principales protéines structurales coronavirales. Les interactions homotypiques entre les protéines M sont le principal moteur de la formation de l'enveloppe du virion mais, à elles seules, ne suffisent pas à la formation du virion (DHAMA *et al.*, 2020).

|     |       |       |       |       |       |        |       |       |       |       |       |        |
|-----|-------|-------|-------|-------|-------|--------|-------|-------|-------|-------|-------|--------|
| 1   | MADSN | GTITV | EELKK | LLEQW | NLVIG | FLFLT  | WICLL | QFAYA | NNRFL | YIIK  | LIFLW | LLWPV  |
| 61  | TLACF | VLA   | AAV   | YRINW | ITGGI | AIAMAC | LVGL  | MWLSY | FIASF | RLFAR | TRSMW | SFNPET |
| 121 | NVPLH | GTILT | RPLLE | SELVI | GAVIL | RGHLR  | IAGHH | LGRCD | IKDLP | KEITV | ATSRT | LSYYK  |
| 181 | LGASQ | RVAGD | SGFAA | YSRYR | IGNYK | LNTDH  | SSSDN | IAL   | VQ    |       |       |        |

**Figure 08 :** Séquence de protéine M (Source : Uniprot, code d'accès : P0DTC5-1).



**Figure 09 :** Schéma d'une particule de Coronavirus, la protéine M de la membrane est liée à la fois avec la membrane de l'enveloppe virale et la protéine N qui se lie à son tour avec le génome de l'ARN. (KHAMMOUM *et al.*, 2021).

#### 2.3.2.5. Relation entre les protéines S,E,N,M

Une protéine E est impliquée dans l'assemblage et la libération de particules de type virus. La protéine M donne la forme de l'enveloppe virale et se lie à la nucléocapside. Tandis que la protéine N se lie au génome de l'ARN CoV, construisant la nucléocapside pour soutenir le complexe de réplication-transcription. L'interaction entre les protéines M et les protéines N stabilise la nucléocapside et, éventuellement, améliore l'assemblage des particules de virion et favorise l'achèvement de l'assemblage viral. Les protéines M, E et S pénètrent ensuite dans le

complexe du compartiment intermédiaire réticulum endoplasmique-Golgi et constituent l'enveloppe virale (ROSSIGNOL *et al.*, 2021).

Les protéines accessoires semblent favoriser l'adaptation des CoV avec les cellules l'hôte humaines.

L'interaction de S avec M est nécessaire pour la rétention de S dans le complexe ER-compartiment Golgien intermédiaire et son incorporation dans de nouveaux virions, mais indispensable pour le processus d'assemblage. Ensemble, M et E constituent l'enveloppe virale et leur interaction est suffisante pour la production et la libération de VLP (ES-SAAD, 2019).

#### **2.3.2.6. Polyprotéines réplicases 1ab, 1a**

Outre les protéines structurales susmentionnées, il existe plusieurs protéines non structurales, à savoir NSP1 à NSP 10 et NSP12 à NSP16, codées par des gènes situés dans la région 5' du génome de l'ARN viral. Le gène de la réplicase du SARS-CoV code pour 16 NSPs avec de multiples fonctions enzymatiques. Celles-ci sont connues ou devraient inclure des types d'enzymes qui sont des composants communs de la machinerie de réplication des virus à ARN à brin positif : une activité ARN polymérase dépendante de l'ARN (RdRp, nsp12). Les activités enzymatiques et les domaines fonctionnels des CoVs nsps devraient être conservés entre les différents genres de CoVs, suggérant leur importance dans la réplication virale (RAMDANI, 2019).

#### **2.3.2.7. Complexe réplicase-transcriptase (RTC).**

Pour créer un environnement propice à la synthèse d'ARN et responsable de la réplication de l'ARN et la transcription des ARN sous-génomiques. Les NSPs contiennent également d'autres domaines enzymatiques. De plus, le gène de la réplicase code pour des protéines qui ont des domaines indicatifs de l'activité exoribonucléase 3-5 (homologue ExoN, nsp14), activité endoribonucléase (homologue de XendoU, nsp15), activité adénosine diphosphate ribose 1'-phosphatase (ADRP, nsp3) et ribose à activité 2 -O- méthyl transférase (2 -O- MT, nsp16). Le produit protéique exprimé par *E. coli* pour nsp9 correspond aux résidus 4118–4230 de la polyprotéine réplicase ORF1a (la protéine nsp9 mature putative) (KRISHNA *et al.*, 2008). Le génome contient un grand gène de réplicase englobant des protéines non structurales (Nsps), suivi de gènes structuraux et accessoires. En raison du décalage du cadre ribosomique, le gène de la réplicase code pour deux cadres de lecture ouverts (ORF), rep1a et rep1b, qui sont traduits en deux grandes polyprotéines, pp1a et pp1ab (figure 10 et 11) (BARTLAM *et al.*, 2005).

```

1 MESLVPGFNE KTHVQLSLPV LQVRDVLVRG FGDSVEEVL S EARQHLKDG CGLVEVEKGV LPQLEQPYVF IKRSDARTAP
81 HGHVMVELVA ELEGIQYGRS GETLGLVLPH VGEIPVAYRK VLLRKNGNKG AGGHSYGADL KSFIDLDELG TDPYEDFQEN
161 WNTKHSSSVT RELMRELNGG AYTRYVDNNF CGPDGYPLEC IKDLLARAGK ASCTLSEQLD FIDTKRGVYC CREHEHEIAW
241 YTERSEKSYE LQTPFEIKLA KKFDTFNGEC PNFVFPLNSI IKTIQPRVEK KKLDFGFMGRI RSVYPVASPN ECNQMCSTL
321 MKCDHCGETS WQTGDFVKAT CEFCEGTENLT KEGATTCGYL PQNAVVKIYC PACHNSEVGP EHSLEAYHNE SGLKTILRKG
401 GRTIAFGGCV FSYVGCHNKC AYWVPRASAN IGCNHTGVVG EGSEGLNDNL LEILQKEKVN INIVGDFKLN EEIAILASF
481 SASTSAFVET VKGLDYKAFK QIVESCNGFK VTKGKAKKGA WNIGEQKSIL SPLYAFASEA ARVRSIFSR TLETAQNSVR
561 VLQKAAITIL DGISQYSRLR IDAMMFTSDL ATNNLVVMAY ITGGVVQLTS QWLTNIFGTV YEKLKPVLDW LEEKFKEGVE
641 FLRDGWEIVK FISTCACEIV GGQIVTCAKE IKESVQTFKK LVNKFALCA DSIIIGGAKL KALNLGETFV THSKGLYRKC
721 VKSREETGLL MPLKAPKEII FLEGETLPT E VLTEEVVLKT GDLQPLEQPT SEAVEAPLVG TPVCINGLML LEIKDTEKYC
801 ALAPNMMVTN NTFTLKGGA TKVTFGDDTV IEVQGYKSVN ITFELDERID KVLNEKCSAY TVELGTEVNE FACVVADAVI
881 KTLQPVSELL TPLGIDLDEW SMATYYLFDE SGFEKLASHM YCSFYPPDED EEEGDCEEEE FEPSTQYEGY TEDDYQGKPL
961 EFGTSAALP PEEGEEDWL DDDSQQTQVQ QDGEFLNNDI YTSITVEVQP QLEMELTPVV QTIEVNSFSG QLEMLTILRKG
1041 KNADIVEEAK KVKPTVVVNA ANVYLKHGGG VAGALNKATN NAMQVESDDY IATNGPLKVG GSCVLSGHNL AKHCLHVVG
1121 NVNKGEDIQL LKSAYENFNQ HEVLLAPLLS AGIFGADPIH SLRVCVDTVR TNVYLAVFDK NLVDKLVSSF LEMKSEKQVE
1201 QKIAEIPKEE VKPFITESKP SVEQRKQDDK KIKACVEEV T TLEETKFLT ENLLYIDIN GNLPDSDATL VSDIDITFLK
1281 KDAPIYIGDV VQEGVLTA V IPTKKAGGT E IMAKALRKV PTDNYITTY P QGGLNGYTV E VNGLTSLIKWA DNNCKLAPL
1361 SNEKQEILGT VSWNLREMLA HAEETRKLM P VCVETKAIVS TIQRKYKGIK IQEGVVDYGA RFYFYTSKTT VASLINTLND
1441 LNETLVMTPL GYVTHGLNLE EAARYMRSK VPATVSVSSP DAVTAYNGYL TSSSKTPEEH FIETISLAGS YKDSYSQS
1521 TQLGIEFLKR GDKSVYTSN PTFHLDGEV ITFDNLKTL SLREVRTIKV FTTVDNINLH TQVVDMSMTY GQQFGPTYLD
1601 GADVTIKPH VCDNIKFADD LNQLTGYYKM ASRELKVTFF PDLNGDVVAI DYKHYTPSKF KGAKLLHKPI VWHVNNATNK
1681 LTLQQLIEKF NPPALQDAYY RARAGEAANF CALILAYCNK TVGELGDVRE TMSYLFQHAN LDSCKRVLNV VCKTCGQQQT
1761 TLKGVEAVMY MGTLSYEQFK KGVQIPCTCG KQATKYLVQQ ESPFVMSAP PAQYELKHGT FTCASEYTN YQCGHYKHIT
1841 SKETLYCIDG ALLTKSEYK GPITDVFYKE NSYTTTIKPV TYKLDGVVCT EIDPKLDNYY KKDNSYFTEQ PIDLVPNPQY
1921 PNASFDNFKL VCDNIKFADD LNQLTGYYKM ASRELKVTFF PDLNGDVVAI DYKHYTPSKF KGAKLLHKPI VWHVNNATNK
2001 ATYKPNTWCI RCLWSTKPVE TSNSFDVLKS EDAQGMNDLA CEDLKPVSEE VVENPTIQKD VLECNCVTTE VVGDIILKPA
2081 NNSLKITEEV GHTDLMAAYV DNSSLTIKKP NELSRVLGLK TLATHGLAAV NSVPWDTIAN YAKPFLNKVV STTTNIVTRC
2161 LNRVCTNYMP YFFTLQLQC TFTRSTNSRI KASMPPTIAK NTVKSVGKFC LEASFNYLKS PNFSKLINII IWFLLLSVCL
2241 GSLFIYSTAAL GVLMSNLGMP SYCTGYREGY LNSTNVTIAT YCTGSIPCS CLSGLDSLTP QPSLETIQIT ISSFKWDLTA
2321 FGLVAEWFLA YILFTRFFYV LGLAAIMQLF FSYFAVHFIS NSWLMWLIIN LVQMAPISAM VRMYIFFASF YVWKSIVHV
2401 VDGCSNSTCM MCYKRNRATR VECTTIVNGV RRSFYVYANG GKGFCKLHNW NCVNCDTFCA GSTFISDEVA RDLSLQFKRP
2481 INPTDQSSYI VDSVTVKNGS IHLYFDKAGQ KTYERHLSH FVNLDNLRAN NTKGSLPINV IVFDGKSKCE ESSAKSASVY
2561 YSGLMCQPII LLDQALVSDV GDSAEVAVKM FDAYVNTFSS TFNVPMKELK TLVATAEAE AKNVSLDNVL STFISAAQRG
2641 FVDSVDVETK VVECLKLSHQ SDIEVTGDSC NNYMLTYNKV ENMTPRDLGA CIDCSARHIN AQVAKSHNIA LIWNVKDFMS
2721 LSEQLRKQIR SAAKKNLPLF KLTCATTRQV VNVVTTKIAL KGGKIVNNWL KQLIKVTLVF LFVAAIFYLI TPVHVMSKHT
2801 DFSSEIIGYK AIDGGVTRDI ASTDTCFANK HADFDWFSQ RGGSYTNDKA CPLIAAVITR EVGFVVPGLP GTILRTTNGD
2881 FLHFLPRVFS AVGNICYTPS KLIEYDFAT SACVLAECT IFKDASGKPV PYCYDTNVLE GSVAYESLRP DTRYVLMDS
2961 IIQFPNTYLE GSVRVVTTFD SEYCRHGTCE RSEAGVCVST SGRWLVNDY YRSLPGVFCG VDAVNLLTNM FTPLIQPIGA
3041 LDISASIVAG GIVAIVVTCL AYYFMRERRA FGEYSHVAF NTLFLMSFT VLCLTPVYSF LPGAYSVIYL YLTFYLTNDV
3121 SFLAHIQWMV MFTPLVPFWI TIAYIICIST KHFYWFFSNY LKRRVVFNGV SFSTFEEAAL CTFLLNKEMY LKLRSVDLLP
3201 LTQYNRYLAL YNKYKYFSGA MDTTSYREAA CCHLAKALND FSNSGSDVLY QPPQTSITS VLQSGFRKMA FPSGKVEGCM
3281 VQVTCGTTTL NGLWLDVVY CPRHVICTSE DMLNPYEDL LIRKSNHNF VQAGNVQLRV IGHSQMNCVL KLVDTANPK
3361 TPKYKFVRIQ PGQTFSLVAC YNGSPSGVYQ CAMRPNFTIK GSFLNGSCGS VGFNIDYDCV SFCYMHMEL PTGVHAGTDL
3441 EGNFYGPFDV RQTAQAAGTD TTITVNVLAW LYAAVINGDR WFLNRFTTTL NDFNLVAMKY NYEPLTQDHV DILGPLSAQT
3521 GIAVLDMCAS LKELLQNGMN GRTILGSALL EDEFTPFDDV RQCSGVTFQS AVKRTIKGTH HWLLLTILTS LLVLVQSTQW
3601 SLFFFLYENA FLPPFAMGIIA MSAFAMMFVK HKHAFCLFL LPSLATVAYF NMVYPASWV MRIMTWLDMV DTSLSGFKLK
3681 DCMYASAVV LLILMTARTV YDDGARRVWT LMNVLTLYK VYGNALDQA ISMWALISV TSNYSGVTT VMFLARGIVF
3761 MCVEYCPIFF ITGNTLQCIM LVYCFGLGYFC TCYFGLFCLL NRYFRLTLGV YDYLVTQEF RYMNSQGLLP PKNSIDAFKL
3841 NIKLLGVGK PCIKVATVQS KMSDVKCTSV VLLSVLQQLR VESSKLWAQ CVQLHNDILL AKDTEAFEK MVSLLSVLLS
3921 MQGAVDINKL CEEMLDNRAT LQAIASEFSS LPSYAAFATA QEAYEQAVAN GDSEVVLLKL KKSLLNAKSE FDRDAAMQRK
4001 LEKMADQAMT QMYKQARSED KRAKVTSAMQ TMLFTMLRKL DNDALNNIIN NARDGCVPLN IIPLTTAACL MVVIPDYNNTY
4081 KNTCDGTTFT YASALWEIQ VVDADSKIVQ LSEISMDNSP NLAWPLIVTA LRANSVAVKL NNELSPVALR QMSCAAGTTQ
4161 TACTDDNALA YYNTTKGGRF VLALLSDLQD LKWARFPKSD GTGTIYTELE PPCRFVTDTP KGPVKVLYF IKGLNNLNRG
4241 MVLGSLAATV RLQAGNATEV PANSTVLSFC AFAVDAKAY KDYLASGGQP ITNCVKMLCT HTGTGQAITV TPEANMDQES
4321 FGGASCCLYC RCHIDHPNPK GFCDLKGKYV QIPTTCANDP VGFTLKNTVC TVCGMWKGYG CSCDQLREPM LQSADAQSFL
4401 NGFAV

```

**Figure 10 :** Séquence de polyprotéine réplicase 1a (Source : Uniprot, code d'accès : P0DTD1-1).

|      |             |               |              |             |             |             |             |             |               |
|------|-------------|---------------|--------------|-------------|-------------|-------------|-------------|-------------|---------------|
| 1    | MESLVPGFNE  | KTHVQLSLPV    | LQVRDVLVRG   | FGDSVEEVL   | EARQHLKDG   | CGLVEVEKGV  | LPQLEQPYVF  | IKRSDARTAP  | HGHVMVELVA    |
| 91   | ELEGIQYGRS  | GETLGVLPVPH   | VGEIPVAYRK   | VLLRKNNGK   | AGGHSYGADL  | KSFDLGDELG  | TDPYEDFQEN  | WNTKHSSGVT  | RELMRELNGG    |
| 181  | AYTRYVDNNF  | CGPDGYPLEC    | IKDLLARAGK   | ASCTLSEQLD  | FIDTKRGVYC  | CREHEHEIAW  | YTERSEKSYE  | LQTPFEIKLA  | KKFDTFNCEC    |
| 271  | PNFVFPLNSI  | IKTIQPRVEK    | KKLDGFMGRI   | RSVYPVASPN  | ECNQMCCLSTL | MKCDHCGETS  | WQTGDFVKAT  | CEFCGTENLT  | KEGATTCGYL    |
| 361  | PQNAVVKIYC  | PACHNSEVGP    | EHSLAEYHNE   | SGLKTIILRK  | GRTIAFGGCV  | FSYVGCHNKC  | AYWVPRASAN  | IGCNHTGVVG  | EGSEGLNDNL    |
| 451  | LEILQKEKVN  | INIVGDFKLN    | EEIAIILASF   | SASTSAFVET  | VKGLDYKAFK  | QIVESCNGFK  | VTGKGAKKGA  | WNIGEKSIL   | SPLYAFASEA    |
| 541  | ARVVRSIFSR  | TLETAQNSVR    | VLQKAAITIL   | DGISQYSRLR  | IDAMMFTSDL  | ATNNLVVMAY  | ITGGVVQLTS  | QWLTNIFGT   | YEKLKPVLDW    |
| 631  | LEEKFKEGVE  | FLRDGWEIVK    | FISTCACEIV   | GGQIVTCAKE  | IKESVQTFFK  | LVNKFALCAL  | DSIIIGGAKL  | KALNLGETFV  | THSKGLYRKC    |
| 721  | VKSREETGLL  | MPLKAPKEII    | FLEGETLPTE   | VLTEEVVLKT  | GDLQPLEQPT  | SEAVEAPLVG  | TPVCINGLML  | LEIKDTEKYC  | ALAPNMMVTN    |
| 811  | NTFTLKGKAP  | TKVTFGDDTV    | IEVQGYKSVN   | ITFELDERID  | KVLNEKCSAY  | TVELGTEVNE  | FACVVAADAVI | KTLQPVSELL  | TPLGIDLDEW    |
| 901  | SMATYYLFDE  | SGEFKLASHM    | YCSFYPPDED   | EEEGDCEEEE  | FEPSTQYEYG  | TEDDYQGKPL  | EFGATSAAALQ | PEEEQEEDWL  | DDDSQQTGQ     |
| 991  | QDGSEDNQTT  | TIQTIVEVQP    | QLEMELTPVV   | QTEIEVNSFSG | YKLTDNVYI   | MKNADIVEEAK | KVKPTVVVNA  | ANVYLKHGGG  | VAGALNKATN    |
| 1081 | NAMQVESDDY  | IATNGPLKVG    | GSCVLSGHN    | AKHCLHVVGP  | NVNKGEDIQL  | LKSAYENFNQ  | HEVLLAPLLS  | AGIFGADPIH  | SLRVCVDTVR    |
| 1171 | TNVYLAVF    | FDK NLYDKLVSS | F LEMKSEKQVE | QKIAEIPKEE  | VKPFITESKP  | SVEQRKQDDK  | KIKACVEEV   | TTLTEETFLT  | ENLLLYIDIN    |
| 1261 | GNLHPDSATL  | VSDIDITFLK    | KDAPYIVGDV   | VQEGVLTAVV  | IPTKKAGGTT  | EMLAKALRKV  | PTDNYITTP   | GGGLNGYT    | EAKTVLKKCK    |
| 1351 | SANGLIYASII | NNSLKEIIEV    | TSMNLREMLA   | LDQETKFLMP  | YKESVQTFK   | LIQKRYKGIK  | IQEGVVDYGA  | RFYFYTSKTT  | VASTINLND     |
| 1441 | LNETLVTMPL  | GYVTHGLNLE    | EAARYMRSK    | VPATVSVSSP  | DAVTAYNGYL  | TSSSKTPEEH  | FIETISLAGS  | YKDWSSYSGS  | TQLGIEFLKR    |
| 1531 | GDKSVYYSN   | PTTFHLDGEV    | ITFDNLKTL    | SLREVRTIKV  | FTTVDNINLH  | TQVVDMSMTY  | GQQFGPTYLD  | GADVTKKPH   | NSHEGKTFYV    |
| 1621 | LPNDDTLRVE  | AFEYYHTTDP    | SFLGRYMSAL   | NHTKKWKYPQ  | VNGLTSIKWA  | DNNCYLATAL  | LTLQQIELKF  | NPPALQDAY   | RARAGEAANF    |
| 1711 | CALILAYLKP  | VSNDLKEIIEV   | TMSYLFQHAN   | DNSSTLTKKP  | NELSDVLTGL  | TLATHGLAAV  | NSVPWDITAN  | YAKGFPLNKV  | STTTINVTTRC   |
| 1801 | ESPFVMSAP   | PAQYELKHGT    | FTCASEYTG    | YQCGHYKHIT  | SKETLYCIDG  | ALLTKSSEYK  | GPITDVFYKE  | NSYTTTIKPV  | TYKLDGVVCT    |
| 1891 | EIDPKLDNYY  | KKDNSYFTEQ    | PIDLVPNQPY   | PNASFDFNFK  | VCDNIKFAAD  | LNQLTGYKKP  | ASRELKVTFF  | PDLNGDVVAI  | DYKHYTPSFK    |
| 1981 | KGAKLLHKPI  | VWHVNATNK     | ATYKPNWCI    | RCLWSTKPVE  | TSNSFDVLKS  | EDAQGMNDLA  | CEDLKPVSEE  | VVENPTIQKD  | VLECNAVKTTE   |
| 2071 | VVGDIILKPA  | NNSLKEIIEV    | GHTDLMAAYV   | DNSSTLTKKP  | RDLSLQFKRP  | INPTDQSSYI  | VDSVTVKNGS  | IHLTYFDKAGV | KTYERHSLSH    |
| 2161 | LNRVCTNYMP  | YFFTLQLC      | TFTRSTNSRI   | KASMPPTIAK  | NTVKSVMKFC  | LEASFNYLKS  | PNFSKLINII  | IWFLLSVCL   | GSLIYSTAAL    |
| 2251 | GVLMNSLGM   | P SYCTGYREGY  | LNSTNVTIAT   | YCTGSIPCSV  | CLSGLDLSDT  | YPSLETIQT   | ISSFKWDLTA  | FGLVAEFLA   | YILFTRFFYV    |
| 2341 | LGLAAIMQLF  | FSYFAVHFIS    | NSWLMWLIIN   | LVQMAPISAM  | VRMYIFFASF  | YVWKSYYVHV  | VDGCSNSTCM  | MCYKRNRA    | TR VECTTIVNGV |
| 2431 | RRSFYVYANG  | GKGRFCKLHNW   | NCVNCDTFCA   | GSTFISDEVA  | RDLSLQFKRP  | INPTDQSSYI  | VDSVTVKNGS  | IHLTYFDKAGV | KTYERHSLSH    |
| 2521 | FVNLDNLN    | NTKGLPINV     | IVFDGKSKCE   | ESSAKSASVY  | YSQLMCQPIL  | LLDQALVSDV  | GDSAEVAVKM  | FDAYVNTFSS  | TFNVPMKLLK    |
| 2611 | TLVATAEAE   | L AKNVSLDNVL  | STFISAAARQ   | FVDSVETKD   | VVECLKLSHQ  | SDIEVTGDSC  | NNYMLTYNKV  | ENMTPRDLGA  | CIDCSARHIN    |
| 2701 | AQVAKSHNIA  | LIWNVKDFMS    | LSEQLRKQIR   | SAAKNNLPF   | KLTCATTQV   | VNVVTTKIAL  | KGGKIVNNWL  | KQLIKVTLVF  | LFVAAIFYLI    |
| 2791 | TPVHVMSKHT  | DFSSSEIIGYK   | AIDGGVTRDI   | ASTDTCFANK  | HADFTWFSQ   | RGGSYTNDKA  | CPLIAAVITR  | EVGFVVPGLP  | GTILRTTNGD    |
| 2881 | FLHFLPRVFS  | AVGNICYTSP    | KLIEYTD FAT  | SACVLAAECT  | IFKDASGKPV  | PYCYDTNVLE  | GSVAYESLRP  | DTRYVLM DGS | IIQFPNTYLE    |
| 2971 | GSVRVVTTFD  | SEYCRHGTCE    | RSEAGVCVST   | SGRWVLNNDY  | YRSLPGVFCG  | VDVAVNLLTM  | FTPLIQPIGA  | LDISASIVAG  | GIVAVVVTCL    |
| 3061 | AYYFMRFRRA  | FGEYSHVAVF    | NTLLFLMSFT   | VLCLTPVYSF  | LPGVYSVIYL  | YLTFLYLTNDV | SFLAHIQWMM  | MFTPLVPFVI  | TIAYIICIST    |
| 3151 | KHFFYVYANG  | GKGRFCKLHNW   | NCVNCDTFCA   | GSTFISDEVA  | RDLSLQFKRP  | INPTDQSSYI  | VDSVTVKNGS  | IHLTYFDKAGV | KTYERHSLSH    |
| 3241 | FSNSGSDVLY  | QPPQTSITSA    | VLQSGFRKMA   | FPSGKVEGCM  | VQVTCGTTTL  | NGLWLDVVY   | CPRHVICTSE  | DMLNPNYEDL  | LIRKSNHNFL    |
| 3331 | VQAGNVQLRV  | IGHSMQNCVL    | KLKVD TANPK  | TPKYKFVRIQ  | PGQTFSVLAC  | YNGSPSGVYQ  | CAMRPNFTIK  | GSFLNGSCGS  | VGFNIIDYDCV   |
| 3421 | SFCYMHMHL   | PTGVHAGTDL    | EGNFYGP FVD  | RQTAQAAGTD  | TTITVNVLAW  | LYAAVINGDR  | WFLNRFTTTL  | NDFNLVAMKY  | NYEPLTQDHW    |
| 3511 | DILGLYSAMT  | QGVAVLDMCAS   | LKELLQNGMN   | GRTILGSALL  | EDEFPTQVQV  | RQCSGVTFQS  | AVKRTIKGTH  | HWLLLTITLS  | LLVLVQSTQW    |
| 3601 | SLFFFLYENA  | FLPFAMGIIA    | MSAFAMMFVK   | HKHAFCLCLF  | LPSLATVAYF  | NMVMYPASWV  | MRIMTWLDMV  | DTSLSGFKLK  | DCVMYASAVV    |
| 3691 | LLILMTARTV  | YDDGARRVMT    | LMNVLT LVYK  | VYYGNALDQA  | ISMWALIIISV | TSNYSGVVTT  | VMFLARGIVF  | MCVEYCPIFF  | ITGNTLQCIM    |
| 3781 | LVYCFGLGYF  | TCYFGLFCLL    | NRYFRLTLGV   | YDYLVTSTQEF | RYMNSQGLLP  | PKNSIDAFKL  | NIKLLGVGGK  | PCIKVATVQS  | KMSDVKCTSV    |
| 3871 | VLLSVLQQLR  | VESSSKLWAQ    | CVQLHNDILL   | AKDTEAFEFK  | MVSLLSVLLS  | MQGAVDINKL  | CEEMLDNRAT  | LQAIASEFSS  | LPSYAAFATA    |
| 3961 | QEAYEQAVAN  | GDSEVVLKKL    | KKSLNVAKSE   | FDRDAAMQRK  | LEKMAQAMT   | QMYKQARSED  | KRAKVTSAMQ  | TMLFTMLRKL  | DNDALNNIIN    |
| 4051 | NARDGCVPLN  | IIPLTAAAKL    | MVVIPDYNTY   | KNTCDGTTFT  | YASALWEIQQ  | VVDADSKIVQ  | LSEISMDNSP  | NLAWPLIVTA  | LRANSVVKLQ    |
| 4141 | NNELSPVALR  | QMSCAAGTTQ    | TACTDDNALA   | YYNTTKGGRF  | VLALLSDLQD  | LKWARFPKSD  | GTGTIYTELE  | PPCRFVTDTP  | KGPKVKYLYF    |
| 4231 | IKGLNNLNRG  | MVLGSLAATV    | RLQAGNATEV   | PANSTVLSFC  | AFAYDAAKAY  | KDYLASGGQP  | ITNCVKMLCT  | HTGTGQAITV  | TPEANMDQES    |
| 4321 | FGGASCCLYC  | RCHIDHPNPK    | GFCDLKGKYY   | QIPTTCANDP  | VGFTLKNTVC  | TVCGMWKGYG  | CSCDQLREPM  | LQSADAQSFL  | NRVCGVSAAR    |
| 4411 | LTPCGTGTST  | DVVYRAFDIY    | NDKVAGFAKF   | LKTNCCRFQE  | KDEDDNLIDS  | YFVVKRHTFS  | NYQHEETIYN  | LLKDCAVAK   | HDFKFRIDG     |
| 4501 | DMVPHISRQR  | LTKYTMADLV    | YALRHFD EGN  | CDTLKEILVT  | YNCCDDDYFN  | KKDWDYDFVEN | PDILRVYANL  | GERVVRQALL  | TVQFCDAMRN    |
| 4591 | AGIVGVLTLD  | NQDLNGNWDY    | FGDFIQTTTP   | SGVPVVDSSY  | SLLMPILTLT  | RALTAEHV    | TDLTTPYIKW  | DLLKYDFTEE  | RLKLFDRYFK    |
| 4681 | YWDQTYHPNC  | VNCLDRCIL     | HCANFNVLFS   | TVFPPTSFGP  | LVRKIFVDGV  | PFVSTGYHF   | RELGVVHNQD  | VNLHSSRLSF  | KELLVYAADP    |
| 4771 | AMHAASGNLL  | LDKRTTCFSV    | AALTNNVAFQ   | TVKPGNFNKD  | FYDFAVSKGF  | FKEGSSVELK  | HFFFAQDGNA  | AISDYDYRY   | NLPTMCDIRQ    |
| 4861 | LLFVVEVVVK  | YFDCYDGGCI    | NANQVIVNNL   | DKSAGFPFNK  | WGKARLYYDS  | MSYEDQDALF  | AYTKRNVPT   | ITQMNLKYAI  | SAKNRARTVA    |
| 4951 | GVSICSTMTN  | RQFHQKLLKS    | IAATRGATTV   | IGTSKFYGGW  | HNMLKTYVSD  | VENPHLMGWD  | YPKCDRAMPN  | MLRIMASLRL  | ARKHTTCCSL    |
| 5041 | SHRFYRLANE  | CAQVLSMVM     | CGGSLYVVKP   | GTSSGDATTA  | YANSVFNICQ  | AVTANVNALL  | STDGNKIADK  | YVRNLQHRLY  | ECLYRNDRVD    |
| 5131 | TDFVNEFYAY  | LRKHFSMMIL    | SDDAVVCFNS   | TYASQGLVAS  | IKNFKSVLYY  | QNNVFMSEAK  | CWTETDLTKG  | PHFCSQHTM   | LVKQGGDYVY    |
| 5221 | LPYPDPSPRIL | GAGCFVDDIV    | KTDGTLMIER   | FVSLAIDAYP  | LTKHPNQEYA  | DVFHLYLYQI  | RKLHDELTHG  | MLDMYSVMLT  | NDNTSRYPW     |
| 5311 | EFYEAMYTPH  | TVLQAVGACV    | LCNSQTSLRC   | GACIRRPFLC  | CKCCYDHVIS  | TSHKLVL SVN | PVVCNAPGCD  | VTDTVQLYLG  | GMSYYCKSHK    |

Figure 11 : Séquence de polyprotéine réplicase 1ab (Source : Uniprot, code d'accèsion : P0DTC1-1).

# *Matériel et méthodes*

## 2. Matériel et méthodes

### 2.1. Matériel

#### 2.1.1. Séquences protéiques

Les séquences protéiques sous format FASTA ont été pris à partir de la banque des données protéiques Uniprot. Ces sont les séquences des principales protéines appartenant au virus SARS-CoV-2 et qui se trouvent sur Uniprot, elles sont regroupées dans le tableau 01.

**Tableau 01** : Séquences protéiques de SARS-CoV-2 qui ont été utilisées pour analyses bio-informatiques

| N° | Numéro d'accèsion et longueur | Protéine                                                            | Séquence                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|----|-------------------------------|---------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 01 | P0DTC5<br>222 Aa              | Membrane protein<br>(Protéine membranaire)                          | MADSNGTITV EELKKLLEQW NLVIGFLFLT WICLLQFAYA<br>NRNRFYIIK LIFLWLLWPV TLACFVLA AV YRINWITGGI<br>AIAMACLVGL MWLSYFIASF RLFARTRSMW SFNPETNILL<br>NVPLHGTILT RPLLESELVI GAVILRGHLR IAGHHLGRCD<br>IKDLPKEITV ATSRTLSTYYK LGASQRVAGD SGFAAYSRYR<br>IGNYKLNTDH SSSSDNIAL VQ                                                                                                                                                                                                                                |
| 02 | P0DTC9<br>419 Aa              | Nucleoprotein<br>(Nucléoprotéine)                                   | MSDNGPQNQR NAPRITFGGP SDSTGSNQNG ERSRARSQR<br>RPQGLPNNTA SWFTALTQHG KEDLKFPGRQ GVPINTNSSP<br>DDQIGYYRRA TRRIRGGDGK MKDLSPRWYF YYLGTGPEAG<br>LPYGANKDGI IWVATEGALN TPKDHIGTRN PANNAIVLQ<br>LPQGTTLPKG FYAEGSRGGS QASSRSSSR RNSSRNSTPG<br>SSRGTSARM AGNGGDAALA LLLDRLNQL ESKMSGKGQ<br>QQGQTVTKKS AAEASKKPRQ KRTATKAYNV TQAFGRRGPE<br>QTQGNFGDQE LIRQGTDYKH WPQIAQFAPS ASAFFGMSRI<br>GMEVTPSGTW LTYTGAIKLD DKDPNFKDQV ILLNKHIDAY<br>KTFPTEPKK DKKKKADETQ ALPQRQKKQ TVTLLPAADL<br>DDFSKQLQQS MSSADSTQA |
| 03 | P0DTC4<br>75 Aa               | Envelope small<br>membrane protein<br>(Petite protéine d'enveloppe) | MYSFVSEETG TLIVNSVLLF LAFVVFLLVT LAILTALRLC<br>AYCCNIVNVS LVKPSFYVYS RVKNLNSSRV PDLLV                                                                                                                                                                                                                                                                                                                                                                                                              |
| 04 | P0DTC2<br>1273 Aa             | Spike glycoprotein<br>(Glycoprotéine de spicule)                    | MFVFLVLLPL VSSQCVNLTT RTQLPPAYTN SFTRGVVYPD<br>KVFRSSVLHS TQDLFLPFFS NVTWFHAIHV SGTNGTKRFD<br>NPVLPFNDGV YFASTEKSNIRGWIFGTTL DSKTQSLLIV<br>NNATNVVIKV CEFQFCNDPF LGVYHKNK SWMESEFRVY<br>SSANNCTFEY VSQPFMDLE GKQGNFKNLR EFVFKNIDGY<br>FKIYSKHTPI NLVRDLPQGF SALEPLVDLP IGINITRFQT<br>LLALHRSYLT PGDSSSGWTA GAAAYVGYL QPRTFLLKYN<br>ENGTITDAVD CALDPLSETK CTLKSFTVEK GIYQTSNFRV<br>QPTESIVRFP NITNLCPFGE VFNATRFASV YAWNRKRISN                                                                      |

|    |                   |                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|----|-------------------|---------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                   |                                                               | CVADYSVLVN SASFSTFKCY GVSPTKLN DL CFTNVYADSF<br>VIRGDEVRI APGQTGKIAD YNYKLPDDFT GCVIAWNSNN<br>LDSKVGGNYN YLYRLFRRSN LKPFERDIST EIYQAGSTPC<br>NGVEGFNCYF PLQSYGFQPT NGVGYQPYRV VVLSFELLHA<br>PATVCGPKKS TNLVKNKCVN FNFNGLTGTG VLTESNKKFL<br>PFQQFGRDIA DTTDAVRDPQ TLEILDITPC SFGGVSVITP<br>GTNTSNQVAV LYQDVNCTEV PVAIHADQLT PTWRVYSTGS<br>NVFQTRAGCL IGAHEVNNSY ECDIPIGAGI CASYQTQTNS<br>PRRARVASQ SIIAYTMSLG AENSVAYSNN SIAIPTNFTI<br>SVTTEILPVS MTKTSVDCTM YICGDSTECN NLLQYGSFC<br>TQLNRALTGI AVEQDKNTQE VFAQVKQIYK TPPIKDFGGF<br>NFSQILPDPS KPSKRSFIED LLFNKVTLAD AGFIKQYGDC<br>LGDIAARDLI CAQKFNGLTV LPPLLTDEMI AQYTSALLAG<br>TITSGWTFGA GAALQIPFAM QMAYRFNGIG VTQNVLYENQ<br>KLIANQFNSA IGKIQDSLSS TASALGKLQD VVNQNAQALN<br>TLVKQLSSNF GAISSVLNDI LSRLDKVEAE VQIDRLITGR<br>LQSLQTYVTQ QLIRAAEIRA SANLAATKMS ECVLGQSKRV<br>DFCGKGYHLM SFPQSAPHGV VFLHVTYVPA QEKNFTTAPA<br>ICHDGKAHP REGVFVSNGT HWFVTQRNFY EPQIITDNT<br>FVSGNCDVVI GIVNNTVYDP LQPELDSFKE ELDKYFKNHT<br>SPDVDLGDIS GINASVVNIQ KEIDRLNEVA KNLNESLIDL<br>QELGKYEQYI KWPWYIWLGF IAGLIAIVMV TIMLCCMTSC<br>CSCCLKGCCSC GSCCKFDEDD SEPVKLGKVL HYT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| 05 | P0DTC1<br>4405 Aa | Replicase<br>polyprotein 1a<br>(Polyprotéine<br>réplicase 1a) | MESLVPGFNE KTHVQLSLPV LQVRDVLVRG FGDSVEEVLS<br>EARQHLKDG CGLVEVEKGV LPQLEQPYVF IKRSDARTAP<br>HGHVMVELVA ELEGIQYGRS GETLGLVPH VGEIPVAYRK<br>VLLRKNGNKG AGGHSYGADL KSFDLGDDELG TDPYEDFQEN<br>WNTKHSSGVT RELMRELNGG AYTRYVDNNF CGPDGYPLEC<br>IKDLLARAGK ASCTLSEQLD FIDTKRGVYC CREHEHEIAW<br>YTERSEKSYE LQTPFEIKLA KKFDTFNGEC PNFFVPLNSI<br>IKTIQPRVEK KKLDGFMGRI RSVYPVASPN ECNQMCSTL<br>MKCDHCGETS WQTGDFVKAT CEFCGTENLT KEGATTGCL<br>PQNAVVKIYC PACHNSEVGP EHSLAEYHNE SGLKTLIRKG<br>GRTIAFGGCV FSYVGCHNKC AYWVPRASAN IGCNHTGVVG<br>EGSEGLNDNL LEILQKEKVN INIVGDFKLN EEIAIILASF<br>SASTSAFVET VKGLDYKAFK QIVESCNGFK VTKGKAKKGA<br>WNIGEQKSIL SPLYAFASEA ARVRSIFSR TLETAQNSVR<br>VLQKAAITIL DGISQYSLRL IDAMMFTSDL ATNNLVMMAY<br>ITGGVVQLTS QWLTNIFGTV YEKLKPVLDW LEEKFKEGVE<br>FLRDGWEIVK FISTCACEIV GGQIVTCAKE IKESVQTFFK<br>LVNKFLALCA DSIIIGGAKL KALNLGETFV THSKGLYRKC<br>VKSREETGLL MPLKAPKEII FLEGETLPTE VLTEEVVLKT<br>GDLQPLEQPT SEAVEAPLVG TPVCLNGML LEIKDTEKYC<br>ALAPNMMVTN NTFTLKGGAP TKVTFGDDTV IEVQGYKSVN<br>ITFELDERID KVLNEKCSAY TVELGTEVNE FACVADAVI<br>KTLQPVSELL TPLGIDLDEW SMATYYLFDE SGFEKLASHM<br>YCSFYPPDED EEEDGCEEEE FEPSTQYEGY TEDDYQGKPL<br>EFGATSAALQ PEEQEEDWL DDDSQQTVGQ QDGSQEDNQT<br>TIQTIVEVQP QLEMELTPVV QTIEVNSFSG YLKLTDNVYI<br>KNADIVEEAK KVKPTVVVNA ANVYLKHGGG VAGALNKATN<br>NAMQVESDDY IATNGPLKVG GSCVLSGHN AKHCLHVVGP<br>NVNKGEDIQL LKSAYENFNQ HEVLLAPLLS AGIFGADPIH<br>SLRVCVDTVR TNVYLAVFDK NLYDKLVSSF LEMKSEKQVE<br>QKIAEIPKEE VKPFITESKP SVEQRKQDDK KIKACVEEVT<br>TTLEETKFLT ENLLLYIDIN GNLHPDSATL VSDIDITFLK<br>KDAPYIVGDV VQEGVLTAVV IPTKKAGGTT EMLAKALRKV |

|  |  |  |             |            |             |             |
|--|--|--|-------------|------------|-------------|-------------|
|  |  |  | PTDNYITTY   | GQGLNGYTV  | EAKTVLKKCK  | SAFYILPSII  |
|  |  |  | SNEKQEILGT  | VSWNLREMLA | HAEETRKLMP  | VCVETKAIVS  |
|  |  |  | TIQRKYKGIK  | IQEGVVDYGA | RFYFYTSKTT  | VASLINTLND  |
|  |  |  | LNETLVTMPL  | GYVTHGLNLE | EAARYMRSK   | VPATVSVSSP  |
|  |  |  | DAVTAYNGYL  | TSSSKTPEEH | FIETISLAGS  | YKDWYSYGQS  |
|  |  |  | TQLGIEFLKR  | GDKSVYYTSN | PTTFHLDGEV  | ITFDNLKTL   |
|  |  |  | SLREVRTIKV  | FTTVDNINLH | TQVVDMSMTY  | GQQFGPTYLD  |
|  |  |  | GADVTKIKPH  | NSHEGKTFYV | LPNDDTLRVE  | AFEYHYTTDP  |
|  |  |  | SFLGRYMSAL  | NHTKKWKYPQ | VNGLTSIKWA  | DNNCYLATAL  |
|  |  |  | LTLQQIELKF  | NPPALQDAYY | RARAGEAANF  | CALILAYCNK  |
|  |  |  | TVGELGDVRE  | TMSYLFQHAN | LDSCRVLNV   | VCKTCGQQQT  |
|  |  |  | TLKGVEAVMY  | MGTLSYEQFK | KGVQIPCTCG  | KQATKYLQQ   |
|  |  |  | ESPFVMSAP   | PAQYELKHGT | FTCASEYTN   | YQCGHYKHIT  |
|  |  |  | SKETLYCIDG  | ALLTKSSEYK | GPITDVFYKE  | NSYTTTIKPV  |
|  |  |  | TYKLDGVVCT  | EIDPKLDNYY | KKDNSYFTEQ  | PIDLVPNQPY  |
|  |  |  | PNASFDNFKF  | VCDNIKFADD | LNQLTGYKKP  | ASRELKVTF   |
|  |  |  | PDLNGDVVAI  | DYKHYPSTFK | KGAKLLHKPI  | VWHVNNATNK  |
|  |  |  | ATYKPNWCI   | RCLWSTKPVE | TSNSFDVLKS  | EDAQGMNLA   |
|  |  |  | CEDLKPVSEE  | VVENPTIQKD | VLECNVKTTE  | VVGDIILKPA  |
|  |  |  | NNSLKITEEV  | GHTDLMAAYV | DNSSLTIKKP  | NELSRVLGLK  |
|  |  |  | TLATHGLAAV  | NSVPWDTIAN | YAKPFLNKVV  | STTTNIVTRC  |
|  |  |  | LNRVCTNYMP  | YFFTLLLQLC | TFTRSTNSRI  | KASMPPTIAK  |
|  |  |  | NTVKSVMGFC  | LEASFNYLKS | PNFSKLINII  | IWFLLLSVCL  |
|  |  |  | GSLIYSTAAL  | GVLMSNLGMP | SYCTGYREGY  | LNSTNVTIAT  |
|  |  |  | YCTGSIPCSV  | CLSGLDSLDT | YPSLETIQIT  | ISSFKWDLTA  |
|  |  |  | FGLVAEWFLA  | YILFTRFFYV | LGLAAIMQLF  | FSYFAVHFIS  |
|  |  |  | NSWLMWLIIN  | LVQMAPISAM | VRMYIFFASF  | YRWKSYVHV   |
|  |  |  | VDGCNSSTCM  | MCYKRNRRAT | VECTTIVNGV  | RRSFVYVANG  |
|  |  |  | GKGFCCKLHNW | NCVNCDTFCA | GSTFISDEVA  | RDLSLVFKRP  |
|  |  |  | INPTDQSSYI  | VDSVTVKNGS | IHLVFDKAGQ  | KTYERHSLSH  |
|  |  |  | FVNLNLRAN   | NTKGSPLINV | IVFDGKSKCE  | ESSAKSASVY  |
|  |  |  | YSQLMCQPIL  | LLDQALVSDV | GDSAFAVAVKM | FDAYVNTFSS  |
|  |  |  | TFNVPMKELK  | TLVATAEAE  | AKNVSLDNVL  | STFISAARQG  |
|  |  |  | FVDSVETKD   | VVECLKLSHQ | SDIEVTGDSC  | NNYMLTYNKV  |
|  |  |  | ENMTPRDLGA  | CIDCSARHIN | AQVAKSHNIA  | LIWNVKDFMS  |
|  |  |  | LSEQLRKQIR  | SAAKKNNLPF | KLTCATTRQV  | VNVVTTKIAL  |
|  |  |  | KGGKIVNNWL  | KQLIKVTLVF | LFVAAIFYLI  | TPVHVMSKHT  |
|  |  |  | DFSSEIIGYK  | AIDGGVTRDI | ASTDTCFANK  | HADFDTWFSQ  |
|  |  |  | RGGSYTNDKA  | CPLIAAVITR | EVGFVVPGLP  | GTILRTTNGD  |
|  |  |  | FLHFLPRVFS  | AVGNICYTPS | KLIEYTDFA   | SACVLAAECT  |
|  |  |  | IFKDASGKPV  | PYCYDTNVLE | GSVAYESLRP  | DTRYVLMDGS  |
|  |  |  | IIQFPNTYLE  | GSVRVVTTFD | SEYCRHGTCE  | RSEAGVCVST  |
|  |  |  | SGRWVLNNDY  | YRSLPGVFCG | VDAVNLLTNM  | FTPLIQPIGA  |
|  |  |  | LDISASIVAG  | GIVAIIVTCL | AYYFMRFRRA  | FGEYSHVAVF  |
|  |  |  | NTLLFLMSFT  | VLCLTPVYSF | LPGVYSVIYL  | YLTIFYLTNDV |
|  |  |  | SFLAHIQWMV  | MFTPLVPFWI | TIAYIICIST  | KHFYWFSSNY  |
|  |  |  | LKRRVVFNGV  | SFSTFEEAAL | CTFLLNKEMY  | LKLRSVDLLP  |
|  |  |  | LTQYNRYLAL  | YNKYKYFSGA | MDTTSYREAA  | CCHLAKALND  |
|  |  |  | FSNSGSDVLY  | QPPQTSITSA | VLQSGFRKMA  | FPSGKVEGCM  |
|  |  |  | VQVTCGTTTL  | NGLWLDVVY  | CPRHVICTSE  | DMLNPNYEDL  |
|  |  |  | LIRKSNHNF   | VQAGNVQLRV | IGHSMQNCVL  | KLKVDATANPK |
|  |  |  | TPKYKFVRIQ  | PGQTFSVLAC | YNGSPSGVYQ  | CAMRPNFTIK  |
|  |  |  | GSFLNGSCGS  | VGFNIDYDCV | SFCYMHMEL   | PTGVHAGTDL  |
|  |  |  | EGNFYGPVVD  | RQTAQAAGTD | TTITVNVLAW  | LYAAVINGDR  |
|  |  |  | WFLNRFTTTL  | NDFNLVAMKY | NYEPLTQDHV  | DILGPLSAQT  |
|  |  |  | GIAVLDMCAS  | LKELLQNGMN | GRTILGSALL  | EDEFTPFDDV  |
|  |  |  | RQCSGVTFQS  | AVKRTIKGTH | HWLLLTILTS  | LLVLVQSTQW  |

|    |                   |                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|----|-------------------|-----------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                   |                                                                 | SLFFFLYENA FLPFAMGIIA MSAFAMMFVK HKHAFCLCLFL<br>LPSLATVAYF NMVYMPASWV MRIMTWLDMV DTSLSGFKLK<br>DCVMYASAVV LLILMTARTV YDDGARRVWT LMNVLTLVYK<br>VYYGNALDQA ISMWALIISV TSNYSGVVT VMFLARGIVF<br>MCVEYCPIFF ITGNTLQCIM LVYCFLGYFC TCYFGLFCLL<br>NRYFRLTLGV YDYLVTQEF RYMNSQGLLP PKNSIDAFKL<br>NIKLLGVGGK PCIKVATVQS KMSDVKCTSV VLLSVLQQLR<br>VESSSKLWAQ CVQLHNDILL AKDTTEAFEK MVSLLSVLLS<br>MQGAVDINKL CEEMLDNRAT LQAIASEFSS LPSYAAFATA<br>QEAYEQAVAN GDSEVVLKKL KKSLNVAKSE FDRDAAMQRK<br>LEKMADQAMT QMYKQARSED KRAKVTSAMQ TMLFTMLRKL<br>DNDALNNIIN NARDGCVPLN IIPLTAAKL MVVIPDYNTY<br>KNTCDGTTFT YASALWEIQQ VVDADSKIVQ LSEISMDNSP<br>NLAWPLIVTA LRANSVVKLQ NNELSPVALR QMSCAAGTTQ<br>TACTDDNALA YYNTTKGGRF VLALLSDLQD LKWARFPKSD<br>GTGTIYTELE PPCRFVTDTP KGPKVKYLYF IKGLNNLNRG<br>MVLGSLAATV RLQAGNATEV PANSTVLSFC AFAVDAAKAY<br>KDYLASGGQP ITNCVKMLCT HTGTGQAITV TPEANMDQES<br>FGGASCCLYC RCHIDHPNPK GFCDLKGKYV QIPTTCANDP<br>VGFTLKNTVC TVCGMWKGYG CSCDQLREPM LQSADAQSFL<br>NGFAV                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| 06 | P0DTD1<br>7096 Aa | Replicase<br>polyprotein 1ab<br>(Polyprotéine<br>réplicase 1ab) | MESLVPGFNE KTHVQLSLPV LQVRDVLVRG FGDSVEEVLS<br>EARQHCLKDGT CGLVEVEKGV LPQLEQPYVF IKRSDARTAP<br>HGHVMVELVA ELEGIQYGRS GETLGVLVPH VGEIPVAYRK<br>VLLRKNGNKG AGGHSYGADL KSFDLGDELG TDPYEDFQEN<br>WNTKHSSGVT RELMRELNGG AYTRYVDNNF CGPDGYPLEC<br>IKDLLARAGK ASCTLSEQLD FIDTKRGVYC CREHEHEIAW<br>YTERSEKSYE LQTPFEIKLA KKFDTFNSEC PNVFVPLNSI<br>IKTIQPRVEK KKLDFGMGRI RSVYPVASPN EENQMCLSTL<br>MKCDHCGETS WQTGDFVKAT CEFCGTENLT KEGATCGYL<br>PQNAVVKIYC PACHNSEVGP EHSLAEYHNE SGLKTIILRKG<br>GRTIAFGGCV FSYVGCHNKC AYWVPRASAN IGCNHTGVVG<br>EGSEGLNDNL LEILQKEKVN INIVGDFKLN EEIAIILASF<br>SASTSAFVET VKGLDYKAFK QIVESCNGFK VTKGKAKKGA<br>WNIGEQQSIL SPLYAFASEA ARVVRISIFSR TLETAQNSVR<br>VLQKAAITIL DGISQYSLRL IDAMMFTSDL ATNNLVVMAY<br>ITGGVVQLTS QWLTNIFGTV YEKLKPVLDW LEEKFKEGVE<br>FLRDGWEIVK FISTCACEIV GGQIVTCAKE IKESVQTFK<br>LVNKFLALCA DSIIIGGAKL KALNLGETFV THSKGLYRKC<br>VKSREETGLL MPLKAPKEII FLEGETLPTE VLTEEVVLKT<br>GDLQPLEQPT SEAVEAPLVG TPVCINGLML LEIKDTEKYC<br>ALAPNMMVTN NTFTLKGGAP TKVTFGDDTV IEVQGYKSVN<br>ITFELDERID KVLNEKCSAY TVELGTEVNE FACVVADAVI<br>KTLQPVSELL TPLGIDLDEW SMATYYLFDE SGEFKLASHM<br>YCSFYPPDED EEEGDCEEEE FEPSTQYEGY TEDDYQGKPL<br>EFGATSAALQ PEEEEQEDWL DDDSQQTVGQ QDGSQEDNQT<br>TIQTIVEVQP QLEMELTPVV QTIEVNSFSG YLKLTDNVYI<br>KNADIVVEAK KVKPTVVVNA ANVYLKHGGG VAGALNKATN<br>NAMQVESDDY IATNGPLKVG GSCVLSGHNH AKHCLHVVGP<br>NVNKGEDIQL LKSAYENFNQ HEVLLAPLLS AGIFGADPIH<br>SLRVCVDTVR TNVYLAVFDK NLYDKLVSSF LEMKSEKQVE<br>QKIAEIPKEE VKPFITESKP SVEQRKQDDK KIKACVEEVT<br>TTLEETKFLT ENLLLYIDIN GNLHPDSATL VSDIDITFLK<br>KDAPYIVGDV VQEGVLTAVV IPTKKAGGTT EMLAKALRKV<br>PTDNYITTYP GQGLNGYTV EAKTVLKKCK SAFYILPSII<br>SNEKQEILGT VSWNLREMLA HAEETRLKMP VCVETKAIVS<br>TIQRKYKGIK IQEGVVDYGA RFYFYTSKTT VASLINTLND |

|  |  |  |             |            |             |             |
|--|--|--|-------------|------------|-------------|-------------|
|  |  |  | LNELVTMPL   | GYVTHGLNLE | EAARYMRSK   | VPATVSVSSP  |
|  |  |  | DAVTAYNGYL  | TSSSKTPEEH | FIETISLAGS  | YKDWSYSGQS  |
|  |  |  | TQLGIEFLKR  | GDKSVYYTSN | PTTFHLDGEV  | ITFDNLKTL   |
|  |  |  | SLREVRTIKV  | FTTVDNINLH | TQVVDMSMTY  | GQQFGPTYLD  |
|  |  |  | GADVTKIKPH  | NSHEGKTFYV | LPNDDTLRVE  | AFEYYHTTDP  |
|  |  |  | SFLGRYMSAL  | NHTKKWKYPQ | VNGLTSIKWA  | DNNCYLATAL  |
|  |  |  | LTLQQIELKF  | NPPALQDAYY | RARAGEAANF  | CALILAYCNK  |
|  |  |  | TVGELGDVRE  | TMSYLFQHAN | LDSCRVNLV   | VCKTCGQQQT  |
|  |  |  | TLKGVEAVMY  | MGTLSYEQFK | KGVPICPTCG  | KQATKYLQQQ  |
|  |  |  | ESPFVMSAP   | PAQYELKHGT | FTCASEYTG   | YQCGHYKHIT  |
|  |  |  | SKETLYCIDG  | ALLTKSSEYK | GPITDVFYKE  | NSYTTTTIKPV |
|  |  |  | TYKLDGVVCT  | EIDPKLDNYY | KKDNSYFTEQ  | PIDLVPNQPY  |
|  |  |  | PNASFDNFKF  | VCDNIKFADD | LNQLTGYYKK  | ASRELKVTFF  |
|  |  |  | PDLNGDVVAI  | DYKHYPSTFK | KGAKLLHKPI  | VWHVNNATNK  |
|  |  |  | ATYKPNWCI   | RCLWSTKPVE | TSNSFDVLKS  | EDAQGMNDLA  |
|  |  |  | CEDLKPVSEE  | VVENPTIQKD | VLECNVKTTE  | VVGDIILKPA  |
|  |  |  | NNSLKITEEV  | GHTDLMAAYV | DNSSLTIKKP  | NELSRVLGLK  |
|  |  |  | TLATHGLAAV  | NSVPWDTIAN | YAKPFLNKVV  | STTTNIVTRC  |
|  |  |  | LNRVCTNYMP  | YFFTLLLQLC | TFTRSTNSRI  | KASMPPTIAK  |
|  |  |  | NTVKSVMKFC  | LEASFNYLKS | PNFSKLINII  | IWFLLLSVCL  |
|  |  |  | GSLIYSTAAL  | GVLMSNLGMP | SYCTGYREGY  | LNSTNVTIAT  |
|  |  |  | YCTGSIPCSV  | CLSGLDSLDT | YPSLETIQIT  | ISSFKWDLTA  |
|  |  |  | FGLVAEWFLA  | YILFTRFFYV | LGLAAIMQLF  | FSYFAVHFIS  |
|  |  |  | NSWLMWLIIN  | LVQMAPISAM | VRMYIFFASF  | YYVWKSYYHV  |
|  |  |  | VDGCNSSTCM  | MCYKRNRRAT | VECTTIVNGV  | RRSFYVYANG  |
|  |  |  | GKGFCCLKHNW | NCVNCDTFCA | GSTFISDEVA  | RDLSLQFKRP  |
|  |  |  | INPTDQSSYI  | VDSVTVKNGS | IHLFYDKAGQ  | KTYERHSLSH  |
|  |  |  | FVNLNLRAN   | NTKGSPLINV | IVFDGSKCE   | ESSAKSASYV  |
|  |  |  | YSQLMCQPIL  | LLDQALVSDV | GDSAFAVAVKM | FDAYVNTFSS  |
|  |  |  | TFNVPMEKLL  | TLVATAEAE  | AKNVSLDNVL  | STFISAARQG  |
|  |  |  | FVDSDVETKD  | VVECLKLSHQ | SDIEVTGDSC  | NNYMLTYNKV  |
|  |  |  | ENMTPRDLGA  | CIDCSARHIN | AQVAKSHNIA  | LIWNVKDFMS  |
|  |  |  | LSEQLRKQIR  | SAAKKNNLPF | KLTCATTRQV  | VNVVTTKIAL  |
|  |  |  | KGGKIVNNWL  | KQLIKVTLVF | LFVAAIFYLI  | TPVHVMSKHT  |
|  |  |  | DFSSEIIGYK  | AIDGGVTRDI | ASTDTCFANK  | HADFDTWFSQ  |
|  |  |  | RGGSYTNDKA  | CPLIAAVITR | EVGFVVPGLP  | GTILRTTNGD  |
|  |  |  | FLHFLPRVFS  | AVGNICYTPS | KLIEYDFAT   | SACVLAAECT  |
|  |  |  | IFKDASGKPV  | PYCYDTNVLE | GSVAYESLRP  | DTRYVLMGDS  |
|  |  |  | IIQFPNTYLE  | GSVRVVTTFD | SEYCRHGTCE  | RSEAGVCVST  |
|  |  |  | SGRWVLNNDY  | YRSLPGVFCG | VDAVNLLTNM  | FTPLIQPIGA  |
|  |  |  | LDISASIVAG  | GIVAIIVTCL | AYYFMRFRRA  | FGEYSHVAVF  |
|  |  |  | NTLLFLMSFT  | VLCLTPVVSF | LPGVYSVIYL  | YLTFYLTNDV  |
|  |  |  | SFLAHIQWMV  | MFTPLVPFWI | TIAYIICIST  | KHFYWFFSNY  |
|  |  |  | LKRRVVFNGV  | SFSTFEEAAL | CTFLLNKEMY  | LKLRSDVLLP  |
|  |  |  | LTQYNRYLAL  | YNKYKYFSGA | MDTTSYREAA  | CCHLAKALND  |
|  |  |  | FSNSGSDVLY  | QPPQTSITSA | VLQSGFRKMA  | FPSGKVEGCM  |
|  |  |  | VQVTCGTTTL  | NGLWLDVVY  | CPRHVICTSE  | DMLNPNYEDL  |
|  |  |  | LIRKSNHNFL  | VQAGNVQLRV | IGHSMQNCVL  | KLKVDATANPK |
|  |  |  | TPKYKFVRIQ  | PGQTFSVLAC | YNGSPSGVYQ  | CAMRPNFTIK  |
|  |  |  | GSFLNGSCGS  | VGFNIDYDCV | SFCYMHMEL   | PTGVHAGTDL  |
|  |  |  | EGNFYGPVVD  | RQTAQAAGTD | TTITVNVLAW  | LYAAVINGDR  |
|  |  |  | WFLNRFTTTL  | NDFNLVAMKY | NYEPLTQDHV  | DILGPLSAQT  |
|  |  |  | GIAVLDMCAS  | LKELLQNGMN | GRTILGSALL  | EDEFTPFDDV  |
|  |  |  | RQCSGVTFQS  | AVKRTIKGTH | HWLLLTILTS  | LLVLVQSTQW  |
|  |  |  | SLFFFLYENA  | FLPFAMGIIA | MSAFAMMFVK  | HKHAFCLCLFL |
|  |  |  | LPSLATVAYF  | NMVYMPASWV | MRIMTWLDMV  | DTSLSGFKLK  |
|  |  |  | DCVMYASAVV  | LLILMTARTV | YDDGARRVWT  | LMNVLTLYVK  |

|  |             |              |             |              |
|--|-------------|--------------|-------------|--------------|
|  | VYYGNALDQA  | ISMWALIISV   | TSNYSGVVTT  | VMFLARGIVF   |
|  | MCVEYCPIFF  | ITGNTLQCIM   | LVYCFLGYFC  | TCYFGLFCLL   |
|  | NRYFRLTLGV  | YDYLVTQEF    | RYMNSQGLLP  | PKNSIDAFKL   |
|  | NIKLLGVGGK  | PCIKVATVQS   | KMSDVKCTSV  | VLLSVLQQLR   |
|  | VESSSKLWAQ  | CVQLHNDILL   | AKDTTEAFEK  | MVSLLSVLLS   |
|  | MQGAVDINKL  | CEEMLDNRAT   | LQAIASEFSS  | LPSYAAFATA   |
|  | QEAYEQAVAN  | GDSEVVLKKL   | KKSLNVAKSE  | FDRDAAMQRK   |
|  | LEKMADQAMT  | QMYKQARSED   | KRAKVTSAMQ  | TMLFTMLRKL   |
|  | DNDALNNIIN  | NARDGCVPLN   | IIPLTAAKL   | MVVIPDYNTY   |
|  | KNTCDGTTFT  | YASALWEIQQ   | VVDADSKIVQ  | LSEISMDNSP   |
|  | NLAWPLIVTA  | LRANSVVKLQ   | NNELSPVALR  | QMSCAAGTTQ   |
|  | TACTDDNALA  | YYNTTKGGRF   | VLALLSDLQD  | LKWARFPKSD   |
|  | GTGTIYTELE  | PPCRFVTDTP   | KGPKVKYLYF  | IKGLNNLNRG   |
|  | MVLGSLAATV  | RLQAGNATEV   | PANSTVLSFC  | AFAVDAAKAY   |
|  | KDYLASGGQP  | ITNCVKMLCT   | HTGTGQAITV  | TPEANMDQES   |
|  | FGGASCCLYC  | RCHIDHPNPK   | GFCDLKGKYV  | QIPTTCANDP   |
|  | VGFTLKNTVC  | TVCGMWKGYG   | CSCDQLREPM  | LQSADAQSFL   |
|  | NRVCGVSAAR  | LTPCGTGTST   | DVVYRAFDIY  | NDKVAGFAKF   |
|  | LKTNCCRFQE  | KDEDDNLIDS   | YFVVKRHTFS  | NYQHEETIYN   |
|  | LLKDCPAVAK  | HDFFKFRIDG   | DMVPHISRQR  | LTKYTMADLV   |
|  | YALRHFDEGN  | CDTLKEILVT   | YNCCDDDYFN  | KKDWDYFVEN   |
|  | PDILRVYANL  | GERVRQALLK   | TVQFCDAMRN  | AGIVGVLTLT   |
|  | NQDLNGNWYD  | FGDFIQTTTPG  | SGVPVVDSSY  | SLLMPILTLT   |
|  | RALTAESHVD  | TDLTkPYIKW   | DLLKYDFTEE  | RLKLFDTRYFK  |
|  | YWDQTYHPNC  | VNCLDDRCIL   | HCANFNVLFS  | TVFPPTSFGP   |
|  | LVRKIFVDGV  | PFVVTGYHF    | RELGVVHNQD  | VNLHSSRLSF   |
|  | KELLVYAADP  | AMHAASGNLL   | LDKRTTCFSV  | AALTNNVAFQ   |
|  | TVKPGNFNKG  | FYDFAVSKGF   | FKEGSSVELK  | HHFFAQDGNA   |
|  | AISDYDYRYR  | NLPTMCDIRQ   | LLFVVEVVDK  | YFDCYDGGCI   |
|  | NANQVIVNNL  | DKSAGFPFNK   | WGKARLYYDS  | MSYEDQDALF   |
|  | AYTKRNV IPT | ITQMNLYKAI   | SAKNRARTVA  | GVSICSTMTN   |
|  | RQFHQKLLKS  | IAATR GATV   | IGTSKFYGGW  | HNMLKTVYSD   |
|  | VENPHLMGWD  | YPKCDRAMPN   | MLRIMASLVL  | ARKHTTCCSL   |
|  | SHRFYRLANE  | CAQVLSEVM    | CGGSLYVKPG  | GTSSGDATTA   |
|  | YANSVFNICQ  | AVTANVNALL   | STDGNKIADK  | YVRNLQHRLY   |
|  | ECLYRNRDVD  | TDFVNEFYAY   | LRKHFSMMIL  | SDDAVVCFNS   |
|  | TYASQGLVAS  | IKNFKSVLYY   | QNNVFMSEAK  | CWTETDLTKG   |
|  | PHEFCSQHTM  | LVKQGDDYVY   | LPYPDPSRIL  | GAGCFVDDIV   |
|  | KTDGTLMIER  | FVSLAIDAYP   | LTKHPNQEYA  | DVFHLYLQYI   |
|  | RKLHDEL TGH | MLDMYSVMLT   | NDNTSRYWEP  | EFYEAMYTPH   |
|  | TVLQAVGACV  | LCNSQTS LRC  | GACIRRPFLC  | CKCCYDHVIS   |
|  | TSHKLVL SVN | PYVCNAPGCD   | VDVTQLYLG   | GMSYYCKSHK   |
|  | PPISFPLCAN  | GQVFGLYKNT   | CVGSDNVTDF  | NAIATCDWTN   |
|  | AGDYILANTC  | TERLKLFAAETL | KATEETFK    | LSYGIATVRE   |
|  | VLSRELHLS   | WEVGKPRPPL   | NRNYVFTGYR  | VTKNSKVQIG   |
|  | EYTFEKGDYD  | DAVVYRGTTT   | YKLNVDGYFV  | LTSHTVMPLS   |
|  | APTLVPQEHY  | VRITGLYPTL   | NISDEFSSNV  | ANYQKVG MQK  |
|  | YSTLQGPPGT  | GKSHFAIGLA   | LYYPSARIVY  | TACSHAAVDA   |
|  | LCEKALKYLP  | IDKCSRIIPA   | RARVECFDKF  | KVNSTLEQYV   |
|  | FCTVNALPET  | TADIVVFDEI   | SMATNYDL SV | VNARLR AKHY  |
|  | VYIGDPAQLP  | APRTLLTKGT   | LEPEYFNSVC  | RLMKTIGPDM   |
|  | FLGTCRRCPA  | EIVDTV SALV  | YDNKLKAHKD  | KSAQC FKM FY |
|  | KGVITHDVSS  | AINRPQIGVV   | REFLTRNPAW  | RKAVFISPYN   |
|  | SQNAVASKIL  | GLPTQTV DSS  | QGSEYDYVIF  | TQTTETAHSC   |
|  | NVNRFNVAIT  | RAKVGILCIM   | SDRDLYDKLQ  | FTSLEIPRRN   |
|  | VATLQAENV T | GLFKDCSKVI   | TGLHPTQAPT  | HLSVDTKFKT   |
|  | EGLCVDIPGI  | PKDMTYRRLI   | SMMGFKMNYQ  | VNGYPNMFIT   |

|  |  |  |                                             |
|--|--|--|---------------------------------------------|
|  |  |  | REEAIRHVRA WIGFDVEGCH ATREAVGTNL PLQLGFSTGV |
|  |  |  | NLVAVPTGYV DTPNNTDFSR VSAKPPPGDQ FKHLIPLMYK |
|  |  |  | GLPWNVVRK IVQMLSDTLK NLSDRVVFL WAHGFELTSM   |
|  |  |  | KYFVKIGPER TCCLCDRRAT CFSTASDTYA CWHHSIGFDY |
|  |  |  | VYNPFMIDVQ QWGFTGNLQS NHDLYCQVHG NAHVASCDAI |
|  |  |  | MTRCLAVHEC FVKRVDWTIE YPIIGDELKI NAACRKVQHM |
|  |  |  | VVKAALLADK FVLHDIGNP KAIKCVQAD VEWKFYDAQP   |
|  |  |  | CSDKAYKIEE LFYSYATHSD KFTDGVCLFW NCNVDRYPAN |
|  |  |  | SIVCRFDTRV LSNLNLPGCD GGSLYVNKHA FHTPAFDKSA |
|  |  |  | FVNLKQLPFF YYSDSPCESH GKQVSDIDY VPLKSATCIT  |
|  |  |  | RCNLGGAVCR HHANEYRLYL DAYNMMISAG FSLWVYKQFD |
|  |  |  | TYNLWNTFTR LQSLNVAFN VVNKGHFDGQ QGEVPVSIIN  |
|  |  |  | NTVYTKVDGV DVELFENKTT LPVNVAFELW AKRNIKPVPE |
|  |  |  | VKILNNLGVD IAANTVIWDY KRDAHAHIST IGVCSMTDIA |
|  |  |  | KKPTETICAP LTVFFDGRVD GQVDLFRNAR NGVLITEGSV |
|  |  |  | KGLQPSVGPK QASLNGVTLI GEAVKTQFNY YKKVDGVVQQ |
|  |  |  | LPETYFTQSR NLQEFKPRSQ MEIDFLELAM DEFIERYKLE |
|  |  |  | GYAFEHIVYG DFSHSQLGGL HLLIGLAKRF KESPFELEDF |
|  |  |  | IPMDSTVKNY FITDAQTGSS KCVCSVIDLL LDDFVEIIKS |
|  |  |  | QDLSVSVKVV KVTIDYTEIS FMLWCKDGHV ETFYPKLQSS |
|  |  |  | QAWQPGVAMP NLYKMQRMLL EKCDLQNYGD SATLPKGIMM |
|  |  |  | NVAKYTQLCQ YLNTLTALVP YNMRVIHFGA GSDKGVAPGT |
|  |  |  | AVLRQWLPTG TLLVDSLDND FVSDADSTLI GDCATVHTAN |
|  |  |  | KWDLIISDMY DPKTKNVTKE NDSKEGFFTY ICGFIQKLA  |
|  |  |  | LGGSVAIKIT EHSWNADLYK LMGHFAWTA FVTNVNASSS  |
|  |  |  | EAFLIGCNLY GKPREQIDGY VMHANYIFWR NTNPIQLSSY |
|  |  |  | SLFDMSKFPL KLRGTAVMSL KEGQINDMIL SLLSKGRLII |
|  |  |  | RENNRVVISS DVLVNN                           |

### 2.1.2. Applications bio-informatiques utilisées

Un nombre d'outils et applications bio-informatiques ont été dans cette étude, tels que : Protein BLAST, Clustal Omega, Prosite, I-TASSER server, outil SOPMA et GPMaw software V 11.

## 2.2. Méthodes

Les analyses bio-informatiques qui ont été qui ont été réalisées sont :

### 2.2.1. Alignement en utilisant BLAST (Basic Local Alignment Search Tool) et l'alignement multiple par COBALT et Clustal Omega

Blast a été utilisé pour l'alignement et recherche d'identité (dans l'archive de Uniprot) pour des protéines qui sont structuralement proches aux protéines de SARS-CoV-2 (mentionnées dans le tableau 01).

L'outil BLAST trouve des régions de similarité locale entre des séquences de protéines ou de nucléotides. Le programme compare les séquences de nucléotides ou de protéines à la séquence dans une base de données et calcule la signification statistique des correspondances.

La valeur statistique, appelée « Expect Value ». La « Expect Value » est le nombre de fois qu'un alignement aussi bon ou meilleur que celui trouvé par BLAST devrait se produire par hasard, compte tenu de la taille de la base de données recherchée. Un seuil de « Expect Value », défini par l'utilisateur, détermine les alignements qui seront signalés. Un seuil de « Expect Value » plus élevé est moins strict et la valeur par défaut de « 10 » de BLAST est conçue pour garantir qu'aucun alignement biologiquement significatif n'est manqué. Cependant, les « Expect Value » comprises entre 0,001 et 0,0000001 sont couramment utilisées pour limiter les alignements affichés à ceux de haute qualité (WHEELER *et al.*, 2007).

Clustal Omega et COBALT ont été utilisés pour la recherche des régions conservatives au niveau des séquences protéiques étudiées, c'est par ce que sur ces régions conservatives que repose les propriétés essentielles des protéines telles que la fonction protéiques, l'interaction de protéines avec leur environnement et etc.

Clustal Omega (également écrit comme Clustal O et Clustal  $\Omega$ ) est un programme rapide et évolutif écrit en C et C++ utilisé pour l'alignement de séquences multiples. Il utilise des arbres guides prédéfinis et un nouveau moteur HMM qui se concentre sur deux profils pour générer ces alignements. Le programme nécessite trois séquences ou plus afin de calculer l'alignement de séquences multiples. Clustal Omega est basé sur la cohérence et est largement considéré comme l'une des implémentations en ligne les plus rapides de tous les outils d'alignement de séquences multiples et se classe toujours parmi les algorithmes basés sur la cohérence et basés sur la matrice.

COBALT (constraint-based alignment tool) a un cadre général qui utilise un alignement multiple progressif pour combiner des contraintes par paires provenant de différentes sources dans un alignement multiple. COBALT n'essaie pas d'utiliser toutes les contraintes disponibles, mais utilise uniquement un sous-ensemble cohérent à score élevé qui peut changer au fur et à mesure que l'alignement progresse, où un ensemble de contraintes est appelé cohérent si toutes les contraintes de l'ensemble peuvent être simultanément satisfaites par un multiple alignement. Nous pouvons rapidement rechercher des domaines dans CDD (conserved domain database) qui correspondent aux régions des séquences d'entrée. Lorsque le même domaine correspond à plusieurs séquences, nous pouvons déduire plusieurs contraintes potentielles par paires basées sur ces correspondances de domaine. En outre, CDD contient également des informations auxiliaires qui permettent à COBALT de créer des profils partiels pour les séquences d'entrée avant le début de l'alignement progressif, ce qui évite des procédures coûteuses en calcul pour la construction de

profils. Nous utilisons les modèles PROSITE pour créer des contraintes uniquement dans l'étape de raffinement, car leur utilisation dans les étapes initiales a donné de moins bonnes performances (données non présentées). Cela est probablement dû au fait que les modèles PROSITE sont beaucoup plus courts que les domaines de CDD et, en tant que tels, sont plus susceptibles de donner de fausses correspondances. COBALT conserve également un sous-ensemble cohérent maximum de toutes les contraintes par paires spécifiées par l'utilisateur en donnant à ces contraintes une priorité supérieure à celle de toute contrainte dérivée par d'autres moyens.

### **2.2.2. Recherche des domaines, motifs et sites fonctionnels en utilisant Prosite**

Nous avons utilisé la base Prosite pour rechercher des domaines protéiques indiquant la famille protéique ou recherche des sites ou motif fonctionnels au sein de protéines de SARS-CoV-2.

PROSITE est une base de données de familles et de domaines de protéines. Elle repose sur le constat que, s'il existe un très grand nombre de protéines différentes, la plupart d'entre elles peuvent être regroupées, sur la base de similitudes dans leurs séquences, en un nombre limité de familles. Les protéines ou domaines protéiques appartenant à une famille particulière partagent généralement des attributs fonctionnels et sont dérivés d'un ancêtre commun.

Il apparaît, lors de l'étude des familles de séquences protéiques, que certaines régions ont été mieux conservées que d'autres au cours de l'évolution. Ces régions sont généralement importantes pour la fonction d'une protéine et/ou pour le maintien de sa structure tridimensionnelle. En analysant les propriétés constantes et variables de tels groupes de séquences similaires, il est possible de dériver une signature pour une famille ou un domaine de protéines, qui distingue ses membres de toutes les autres protéines non apparentées. Une analogie pertinente est l'utilisation des empreintes digitales par la police à des fins d'identification. Une empreinte digitale est généralement suffisante pour identifier un individu donné. De même, une signature protéique peut être utilisée pour attribuer une protéine nouvellement séquencée à une famille spécifique de protéines et ainsi formuler des hypothèses sur sa fonction.

PROSITE contient actuellement des modèles et des profils spécifiques pour plus d'un millier de familles ou de domaines de protéines. Chacune de ces signatures est accompagnée d'une documentation fournissant des informations générales sur la structure et la fonction de ces protéines (SIGRIST *et al.*, 2002).

### **2.2.3. I-TASSER server pour prédiction de structure bidimensionnelle et tridimensionnelle**

On s'est servi de I-TASSER server afin de prédire la structure bidimensionnelle et tridimensionnelle les plus optimales de protéine spike de SARS-CoV-2 et pour mettre en évidence les ses résidus aminoacides qui entrent en interaction avec son ligand naturel.

I-TASSER (Iterative Threading ASSEmbly Refinement) est une méthode bioinformatique permettant de prédire un modèle de structure tridimensionnelle de molécules de protéines à partir de séquences d'acides aminés.[1] Il détecte les modèles de structure de la Protein Data Bank par une technique appelée reconnaissance de pli (ou threading). Les modèles de structure pleine longueur sont construits en réassemblant des fragments structuraux à partir de modèles de filetage à l'aide de simulations Monte Carlo d'échange de répliques. I-TASSER est l'une des méthodes de prédiction de la structure des protéines les plus réussies dans les expériences CASP à l'échelle de la communauté (ROY *et al.*, 2010).

#### **2.2.4. Prédiction des certaines caractéristiques physico-chimiques par le logiciel GPMAW**

La détermination des caractéristiques physico-chimiques des protéines est très essentielle pour pouvoir suggérer un protocole ou stratégie de séparation et purification de protéine ou dans certaines mesures pour pouvoir prédire les types d'interactions de protéine avec son milieu ou même le lieu de sa présence au niveau cellulaire. L'utilisation de l'application GPMAW permet de déterminer certains de ces caractéristiques (Masse molaire, point isoélectrique, hydrophobicité de la molécule...).

GPMAW est un programme d'analyse détaillée de la structure primaire des protéines et des peptides. Bien que l'accent soit mis sur l'analyse par spectrométrie de masse, un grand nombre de données physico-chimiques vous seront également présentées.

# *Résultats et discussions*

### 3. Résultats et discussions

#### 3.1. Alignement avec BLAST (recherche de similarités et régions conservatives).

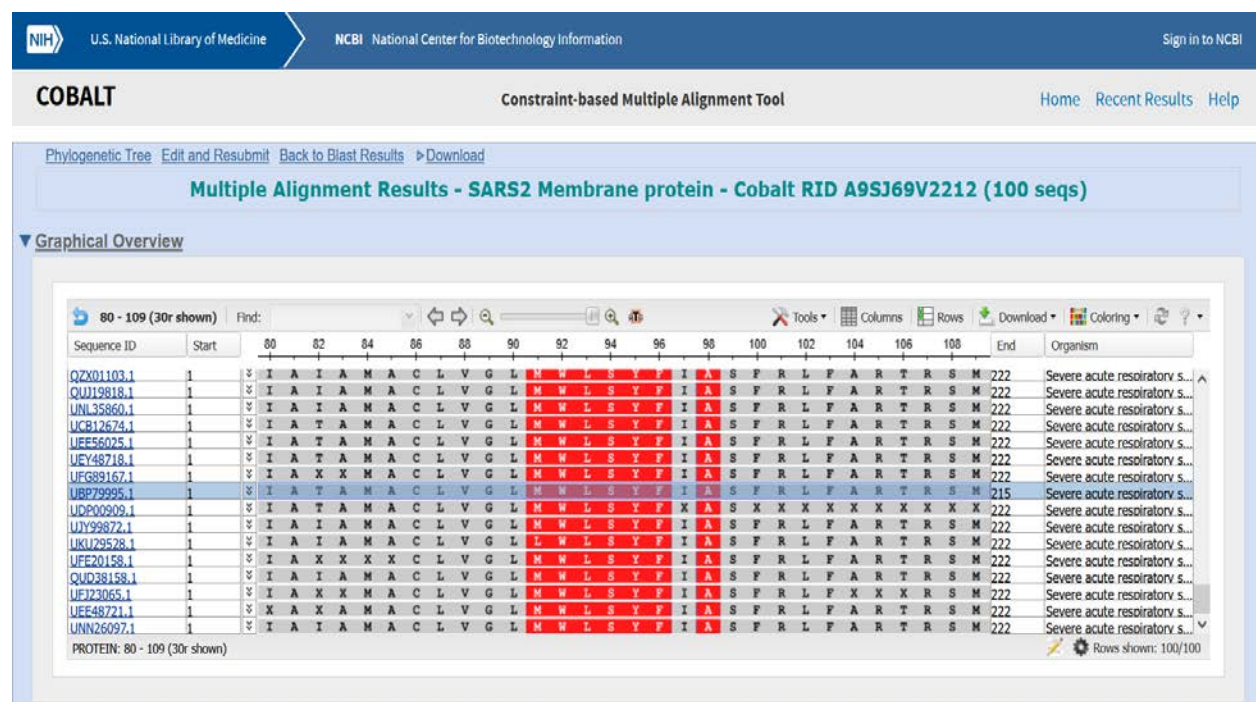
L'alignement a pour objectif de rechercher des protéines dans la nature qui ont des structures proches des celles des protéines virales dont on peut utiliser par exemple les recherches qui ont été faits sur ces protéines afin de les exploiter avec les protéines virales pour rechercher des solutions pour lutter contre ce virus (recherche des médicaments, des inhibiteurs ou antagonistes...).

La recherche sur des motifs conservatifs a un but crucial, parce que c'est sur cet endroit que repose le fonctionnement de la protéine et aucune mutation n'est toléré au niveau de ces régions de la séquence protéiques.

##### 3.1.1. Protéine membranaire

L'alignement de cette protéine avec Protein BLAST n'a montré aucune similarité avec d'autres protéines déjà connues.

Par ailleurs, l'alignement avec 100 autres séquences des corona virus a montré qu'il y une région conservative au niveau de cette protéine qui est entre 91<sup>ème</sup> et 98<sup>ème</sup> acides aminées de la séquence (91)MWLSYFIA(98) comme il montre la figure suivante.

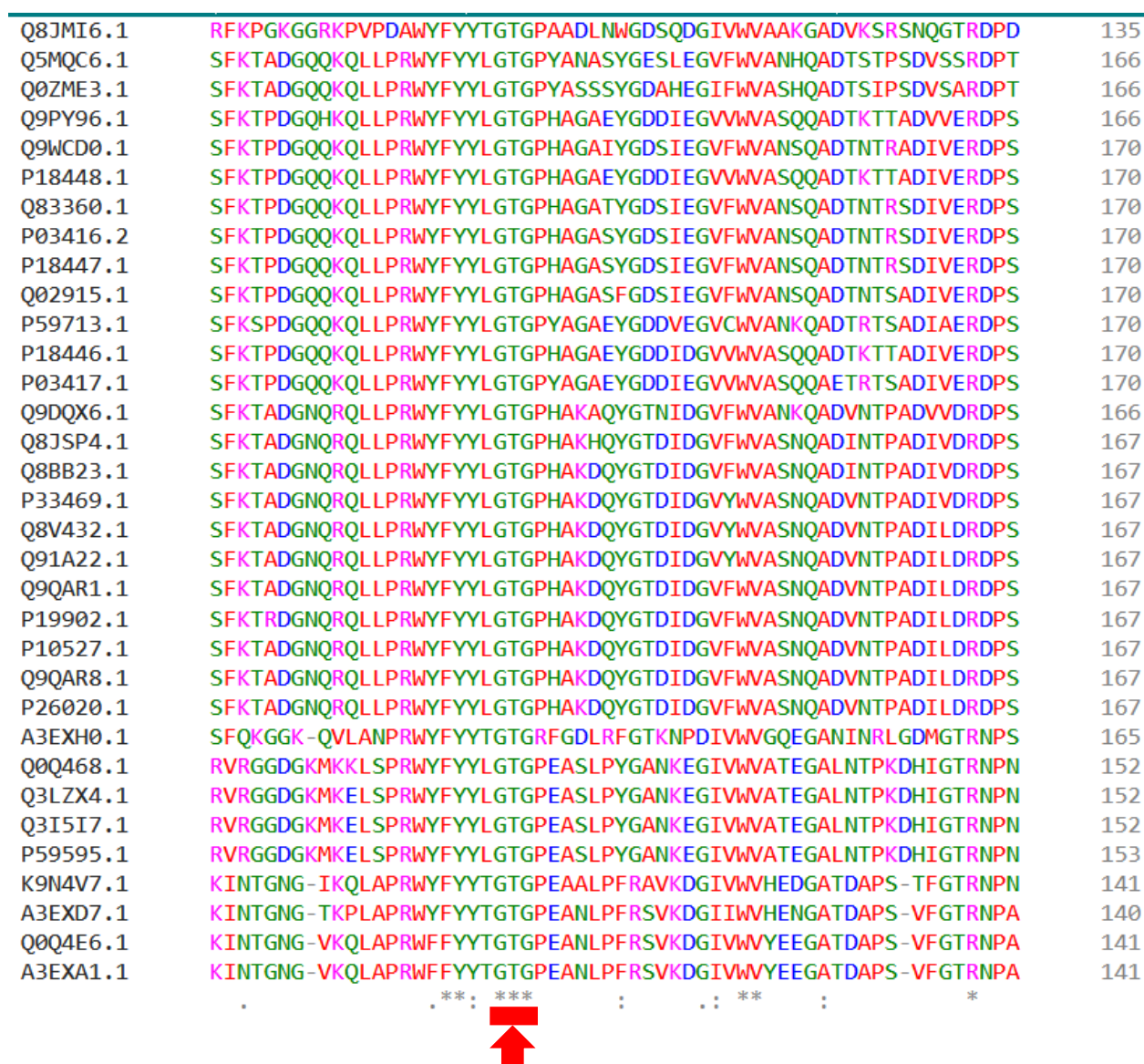


**Figure 12 :** Alignement multiple de protéine de membrane de SARS-CoV-2 avec 100 autres séquences similaires des Corona virus.

### 3.1.2. Nucléoprotéine

La meilleure similarité de cette protéine a été retrouvée avec *Rousettus bat coronavirus* HKU9, appartenant au sous genre *Nobecovirus*, alors que SARS-CoV-2 appartenant au sous genre *Sarbecovirus* (Zhou *et al.*, 2021).

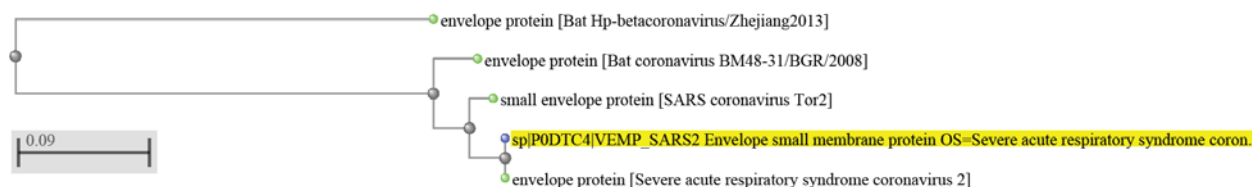
L'alignement multiple a été fait avec BLAST Multiple Alignment Tool qui n'a pas donné un résultat clair, c'est pour cette raison qu'une analyse d'alignement multiple a été réalisée avec Clustal Omega avec 100 séquences protéiques de nucléoprotéine (appartenent a des sources différentes) et qui a montré que la plus grande région conservée pour cette protéine est composée seulement des 3 résidus aminoacides qui sont (114)GTG(116) pour le virus SARS-CoV-2.



**Figure 13 :** Résultat d'alignement de 100 séquences protéiques de nucléoprotéines appartenant à des sources différentes pour détermination des régions conservatives.

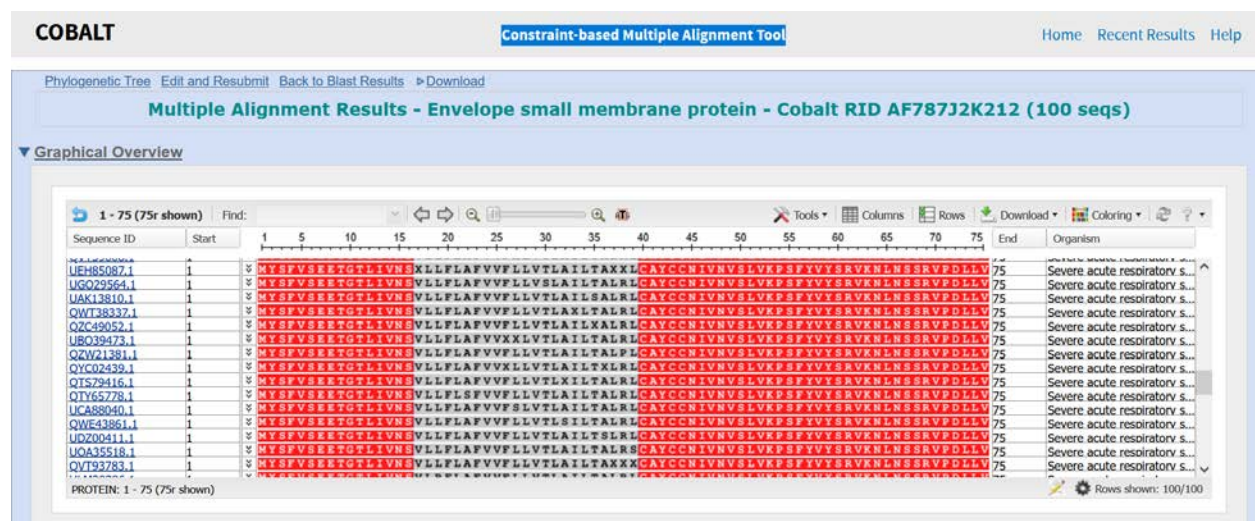
### 3.1.3. Petite protéine d'enveloppe

En dehors de sous genre *Sarbecovirus*, on a enregistré une similarité unique avec la souche Bat Hp-betacoronavirus/Zhejiang2013 appartenant au sous genre de *Hibcovirus* (figure 14) avec un taux de similarité de 54,93%.



**Figure 14 :** L'arbre phylogénétique réalisé par Blast Tree View. La séquence requête est surlignée en jaune.

L'alignement multiple avec COBALT (via BLAST) et avec Clustal, ont montré la présence de deux grande régions conservative au niveau cette protéine, il s'agit de la région (1) MYSFVSEETGTLIVNS(16) et de la région (40)CAYCCNIVNVSLVKPSFYVYS(75), comme il montre la figure 15



**Figure 15 :** Alignement multiple de 100 séquences protéiques de petite protéine d'enveloppe de SARS-CoV par COBALT.

### 3.1.4. Glycoprotéine de spicule (Spike)

L'alignement de cette protéine via BLAST a montré que la meilleure similarité a été avec la même protéine de l'espèce coronavirus de pipistrelle HKU5 avec un taux d'identité de

37,44%, cette espèce appartenant au sous genre *Merbecovirus*, alors que SARS-CoV-2 appartenant au sous genre *Sarbecovirus* (figure 16) (Zhou *et al.*, 2021).

L'alignement multiple de 100 séquences de protéine spike de différentes sources réalisé par Clustal Omega a montré qu'il y a plusieurs régions conservatives qui sont respectivement (860)VLP(862), (972)AIS(974), (1031)ECV(1033), (1042)FCG(1044), (1212)KWPW(1214). Ces régions sont peut-être responsables des propriétés fonctionnelles de cette protéine telle que l'interaction spécifique avec son ligand naturel (ACE2).

| Descriptions                                                         |                                                                                                                                     | Graphic Summary         | Alignments               | Taxonomy                                 |                                    |                            |            |          |           |
|----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|-------------------------|--------------------------|------------------------------------------|------------------------------------|----------------------------|------------|----------|-----------|
| Sequences producing significant alignments                           |                                                                                                                                     |                         |                          |                                          |                                    |                            |            |          |           |
|                                                                      |                                                                                                                                     | Download                | Select columns           | Show                                     | 100                                | ?                          |            |          |           |
| <input checked="" type="checkbox"/> select all 50 sequences selected |                                                                                                                                     | <a href="#">GenPept</a> | <a href="#">Graphics</a> | <a href="#">Distance tree of results</a> | <a href="#">Multiple alignment</a> | <a href="#">MSA Viewer</a> |            |          |           |
|                                                                      | Description                                                                                                                         | Scientific Name         | Max Score                | Total Score                              | Query Cover                        | E value                    | Per. Ident | Acc. Len | Accession |
| <input checked="" type="checkbox"/>                                  | RecName: Full=Spike glycoprotein; Short=S glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Severe acute re...    |                         | 2553                     | 2553                                     | 99%                                | 0.0                        | 100.00%    | 1273     | P0DTC2.1  |
| <input checked="" type="checkbox"/>                                  | RecName: Full=Spike glycoprotein; Short=S glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Severe acute re...    |                         | 1973                     | 1973                                     | 98%                                | 0.0                        | 77.04%     | 1255     | P59594.1  |
| <input checked="" type="checkbox"/>                                  | RecName: Full=Spike glycoprotein; Short=S glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Bat SARS coron...     |                         | 1935                     | 1935                                     | 98%                                | 0.0                        | 76.57%     | 1242     | Q3LZX1.1  |
| <input checked="" type="checkbox"/>                                  | RecName: Full=Spike glycoprotein; Short=S glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Bat SARS CoV...       |                         | 1924                     | 1924                                     | 98%                                | 0.0                        | 75.93%     | 1241     | Q3I5J5.1  |
| <input checked="" type="checkbox"/>                                  | RecName: Full=Spike glycoprotein; Short=S glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Bat CoV 279/2005      |                         | 1910                     | 1910                                     | 98%                                | 0.0                        | 75.38%     | 1241     | Q0Q475.1  |
| <input checked="" type="checkbox"/>                                  | RecName: Full=Spike glycoprotein; Short=S glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Pipistrellus bat c... |                         | 582                      | 582                                      | 77%                                | 0.0                        | 37.44%     | 1352     | A3EXD0.1  |
| <input checked="" type="checkbox"/>                                  | RecName: Full=Spike glycoprotein; Short=S glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Rousettus bat co...   |                         | 572                      | 572                                      | 94%                                | 0.0                        | 32.08%     | 1274     | A3EXG6.1  |
| <input checked="" type="checkbox"/>                                  | RecName: Full=Spike glycoprotein; Short=S glycoprotein; AltName: Full=E2; AltName: Full=Peplomer protein; ... Bat coronavirus ...   |                         | 561                      | 561                                      | 78%                                | 3e-176                     | 36.61%     | 1350     | Q0Q4F2.1  |

**Figure 16 :** Résultat d'alignement via BLAST de la séquence de protéine de spécule de SARS-CoV-2.

### 3.1.5. Polyprotéine réplicase 1a

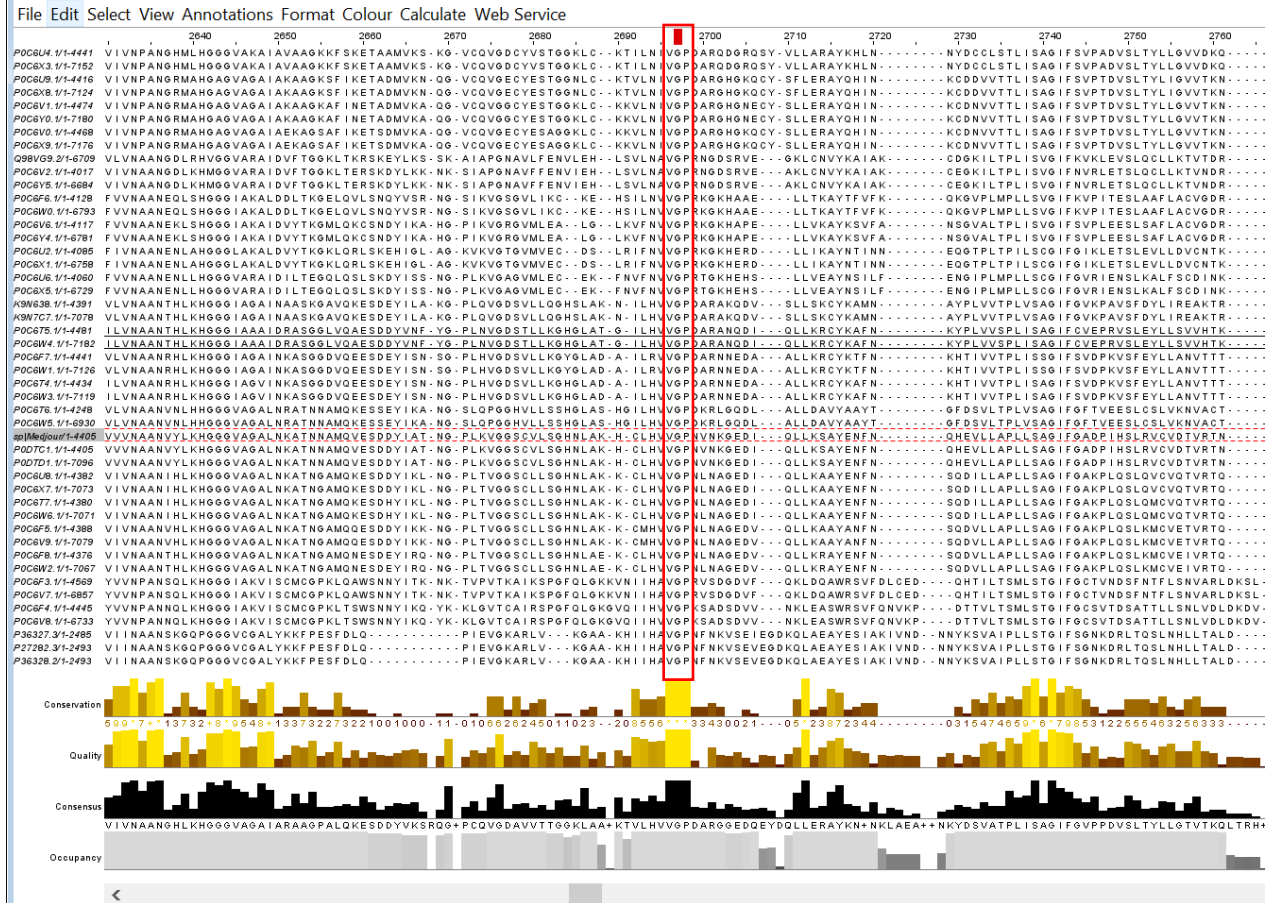
L'alignement en utilisant BASLT a montré que la similarité de structure la plus proche en dehors de sous *Sarbecovirus* est celle avec l'espèce *Rousettus bat coronavirus HKU9*, de sous genre *Nobecovirus*. Une telle similarité montre un taux d'identité de 40,66%.

D'autre part, l'alignement multiple avec 100 autres séquences de polyprotéines réplicase 1a via clustal omega en utilisant Jalview pour visualiser le résultat a montré la présence d'une région conservative qui est (1118)VGP(1120) sur la séquence requête de polyprotéine réplicase 1a (figure 17)

Jalview 2.11.2.2

File Tools Help Window

http://www.ebi.ac.uk/Tools/services/rest/clustalo/result/clustalo-E20220616-110059-0593-64937624-p2m/aln-clustal\_num



**Figure 17 :** Résultat d'alignement avec 100 séquences protéiques des différentes sources de polyprotéine 1a le motif conservé est encadré en rouge.

### 3.1.6. Polyprotéine réplicase 1ab

L'alignement en utilisant BASLT a montré que la similarité de structure la plus proche en dehors de sous *Sarbecovirus* est celle avec l'espèce *Rousettus bat coronavirus HKU9* (Zhou *et al.*, 2021), de sous genre *Nobecovirus*. Une telle similarité montre un taux d'identité de 52,58%.

D'autre part, l'alignement multiple avec 100 autres séquences de polyprotéines réplicase 1ab via clustal omega en utilisant Jalview pour visualiser le résultat n'a montré aucune région conservative sur la séquence.

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

DownloadSelect columnsShow100?

☒

select all

100 sequences selected

GenPept

Graphics

Distance tree of results

Multiple alignment

MSA Viewer

|                                     | Description                                                                                              | Scientific Name       | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|-------------------------------------|----------------------------------------------------------------------------------------------------------|-----------------------|-----------|-------------|-------------|---------|------------|----------|-----------|
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: Rec... | Severe acute re...    | 14365     | 14365       | 100%        | 0.0     | 100.00%    | 7096     | P0DTC1.1  |
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: Rec... | Severe acute re...    | 12658     | 12658       | 100%        | 0.0     | 86.01%     | 7073     | P0C6X7.1  |
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: Rec... | Bat SARS CoV...       | 12599     | 12599       | 100%        | 0.0     | 85.67%     | 7071     | P0C6W8.1  |
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: Rec... | Bat SARS coron...     | 12587     | 12587       | 100%        | 0.0     | 85.69%     | 7067     | P0C6W2.1  |
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: Rec... | Bat CoV 279/2005      | 11202     | 11202       | 85%         | 0.0     | 88.65%     | 7079     | P0C6V9.1  |
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1a; Short=pp1a; AltName: Full=ORF1a polyprotein; Contains: RecNam... | Severe acute re...    | 8693      | 8693        | 62%         | 0.0     | 100.00%    | 4405     | P0DTC1.1  |
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1a; Short=pp1a; AltName: Full=ORF1a polyprotein; Contains: RecNam... | Severe acute re...    | 7160      | 7160        | 62%         | 0.0     | 80.13%     | 4382     | P0C6U8.1  |
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1a; Short=pp1a; AltName: Full=ORF1a polyprotein; Contains: RecNam... | Bat SARS coron...     | 7140      | 7140        | 62%         | 0.0     | 80.05%     | 4376     | P0C6F8.1  |
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1a; Short=pp1a; AltName: Full=ORF1a polyprotein; Contains: RecNam... | Bat SARS CoV ...      | 7133      | 7133        | 62%         | 0.0     | 79.66%     | 4380     | P0C6T7.1  |
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: Rec... | Rousettus bat c...    | 5984      | 6258        | 95%         | 0.0     | 52.58%     | 6930     | P0C6W5    |
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: Rec... | Betacoronavirus...    | 5849      | 6129        | 96%         | 0.0     | 50.84%     | 7078     | K9N7C7    |
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1ab; Short=pp1ab; AltName: Full=ORF1ab polyprotein; Contains: Rec... | Pipistrellus bat c... | 5845      | 6150        | 95%         | 0.0     | 51.41%     | 7182     | P0C6W4    |
| <input checked="" type="checkbox"/> | RecName: Full=Replicase polyprotein 1a; Short=pp1a; AltName: Full=ORF1a polyprotein; Contains: RecNam... | Bat CoV 279/2005      | 5726      | 5726        | 47%         | 0.0     | 83.31%     | 4388     | P0C6F5.1  |

Feedback

**Figure 18 :** Résultat d'alignement via BLAST de la séquence de polyprotéine réplacase de SARS-CoV-2.

### 3.2. Analyse des protéines via Prosite

Prosite a été utilisé pour la recherche de domaine indiquant l'appartenance à une famille protéique et la recherche de sites et motif de fonctionnement protéique.

#### 3.2.1. Protéine membranaire

La figure suivante montre les motifs protéiques retrouvés suite à l'analyse par Prosite.

|                                                                                                                                                |  |                                                                                            |  |
|------------------------------------------------------------------------------------------------------------------------------------------------|--|--------------------------------------------------------------------------------------------|--|
| sp-P0DTC5-VME1_SARS2<br>(sp-P0DTC5-VME1_SARS2)<br> (222 aa) |  | <b>PS00005 PKC_PHOSPHO_SITE</b> Protein kinase C phosphorylation site :                    |  |
| <b>PS00001 ASN_GLYCOSYLATION</b> N-glycosylation site :                                                                                        |  | <b>99 - 101:</b> SFR<br>Predicted feature:<br>MOD_RES 99 Phosphoserine [condition: S]      |  |
| <b>5 - 8:</b> NGTI<br>Predicted feature:<br>CARBOHYD 5 N-linked (GlcNAc...) asparagine [condition: N]                                          |  | <b>172 - 174:</b> TsR<br>Predicted feature:<br>MOD_RES 172 Phosphothreonine [condition: T] |  |
| <b>PS00006 CK2_PHOSPHO_SITE</b> Casein kinase II phosphorylation site :                                                                        |  | <b>184 - 186:</b> SqR<br>Predicted feature:<br>MOD_RES 184 Phosphoserine [condition: S]    |  |
| <b>9 - 12:</b> TveE<br>Predicted feature:<br>MOD_RES 9 Phosphothreonine [condition: T]                                                         |  |                                                                                            |  |
| <b>212 - 215:</b> SssD<br>Predicted feature:<br>MOD_RES 212 Phosphoserine [condition: S]                                                       |  |                                                                                            |  |
| <b>PS00008 MYRISTYL</b> N-myristoylation site :                                                                                                |  |                                                                                            |  |
| <b>79 - 84:</b> GIaiAM                                                                                                                         |  |                                                                                            |  |
| <b>126 - 131:</b> GTILTR                                                                                                                       |  |                                                                                            |  |

**Figure 19 :** Résultats d'analyse de la séquence protéique de protéine membranaire de SARS-CoV-2 avec Prosite.

Les motifs retrouvés sont :

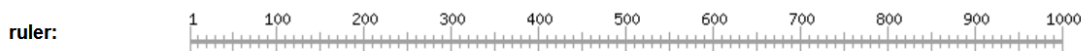
- Un site de glycosylation porté par l'asparagine (N) qui occupe la 5<sup>ème</sup> position au niveau de la séquence, donc N(5) et le motif est NGTI ;
- Deux sites probables pour phosphorylation, l'un est porté par T(9) pour motif TVEE et l'autre est porté par S(212), pour le motif SSSD ;
- Deux sites de N-myristoylation, porté chacun par G(79) et G(126) pour les motifs : GIAIAM et GTILTR. Ces sites servent probablement à la fixation de la protéine avec la membrane de l'enveloppe viral ;
- Et finalement, trois sites de phosphorylation par protéines kinase C, chacun est porté par résidu sérine (S) et qui sont : S(99), S(172) et S(184).

### 3.2.2. Nucléoprotéine

La figure suivante montre les domaines et motifs protéiques retrouvés suite à l'analyse par Prosite.

**hits by profiles:** [2 hits (by 2 distinct profiles) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.



sp-P0DTC9-  
NCAP\_SARS2  
(sp-P0DTC9-NCA  
P\_SARS2 )

(419 aa)

**PS51928 COV\_N\_NTD** Coronavirus nucleocapsid (CoV N) protein N-terminal (NTD) domain profile :

**48 - 175:** score = 55.653

NTASWFTALTQH GK - EDLKFPRGQGV PINTNSPDDQIGYYRATRrIRGGDGKMKDLSP  
RWYFYLLGTGPEAGLPYGANKDGIWVATEGALNTPKDHIGTRNPANNAIVLQLPQGtT  
LPKGFYAEG

**Predicted feature:**

| DOMAIN | 48 | 175 | CoV N NTD | [condition: none] |
|--------|----|-----|-----------|-------------------|
|        |    |     |           |                   |

**PS51929 COV\_N\_CTD** Coronavirus nucleocapsid (CoV N) protein C-terminal (CTD) domain profile :

**247 - 364:** score = 50.581

TKKSAAEASKPRQKRATKAYNVTQAFGRRGpeQTQGNFGDQELIRQGT DYKHW P QIAQ  
FAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTFP

**Predicted feature:**

| DOMAIN | 247 | 364 | CoV N CTD | [condition: none] |
|--------|-----|-----|-----------|-------------------|
|        |     |     |           |                   |

**hits by patterns with a high probability of occurrence or by user-defined patterns:** [33 hits (by 6 distinct patterns) on 1 sequence]



sp-P0DTC9-  
NCAP\_SARS2  
(sp-P0DTC9-NCA  
P\_SARS2 )

(419 aa)

**Figure 20 :** Résultats d'analyse de la séquence protéique de nucléoprotéine de SARS-CoV-2 avec Prosite.

Les domaines et motifs retrouvés sont :

- Deux domaines de nucléocapside pour Coronavirus (l'un est N-terminal 48-175 et l'autre est C-terminal 247-364) ;
- 12 sites de N-myristoylation, porté chacun par G(25), G(44), G(71), G(120), G(175), G(178), G(179), G(200), G(204), G(214), G(244) et G(328), respectivement pour les motifs : GSNQNG, GLPNNT, GVPINT, GLPYGA, GSRGGS, GGSQAS, GSQASS, GSSRGT, GTSPAR, GGDAAL, GQTVTK et GTWLTY. Ces sites servent probablement à la fixation de la protéine avec la membrane de l'enveloppe viral ;
- 3 sites de glycosylation porté par l'asparagine (N) qui occupe les positions 47, 192 et 269 au niveau de la séquence avec les motifs NNTA, NSSR et NVTQ ;
- 6 sites de phosphorylation par caséine kinase II, qui sont portés par : S(78), S(79), T(115), T(141), S(250) et S(412) ;
- Un site de phosphorylation par protéine kinase AMPc et GMPc dépendent pour le motif RRAT ;
- 10 sites de phosphorylation par protéines kinase C, portés par résidu sérine (S) ou thréonine (T) et qui sont : T(91), S(105) et T(141), S(183), S(187), S(193), S(201), S(235), T(247) et S(255) ;
- Et finalement, un site d'amidation pour le motif FGRR.

### 3.2.3. Petite protéine d'enveloppe

La figure 21 montre les motifs protéiques retrouvés suite à l'analyse par Prosite pour la petite protéine d'enveloppe.

Les domaines et motifs retrouvés sont :

- Au niveau de la séquence de la position 1 à 75 est occupé par le profil de protéine de l'enveloppe de *Coronavirus* ;
- Deux sites de glycolisation de type N-liée, portés chacun par asparagine : N(48) et N(66) ;
- Et finalement, un site de phosphorylation par protéine kinase C porté par sérine S(67).

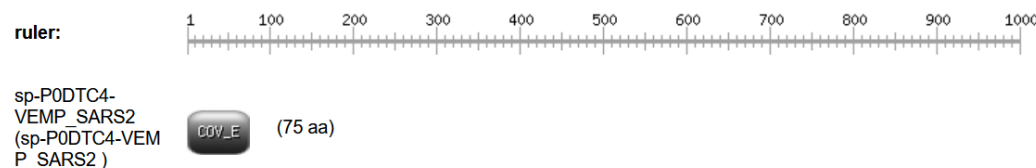
### 3.2.4. Glycoprotéine de spicule (Spike)

La figure 22 A, 22 B et 22 C montrent les domaines et motifs protéiques retrouvés suite à l'analyse par Prosite pour la glycoprotéine de spicule.

Les domaines et motifs retrouvés sont :

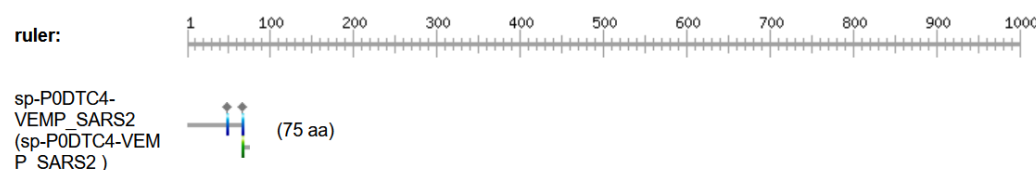
**hits by profiles: [1 hit (by 1 profile) on 1 sequence]**

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.

**PS51926 COV\_E Coronavirus envelope (CoV E) protein profile :**

**1 - 75:** score = 27.870

---MYSFVSEETGTLIVNSVLFLAFVVFLLVTLAILTALRLCAYCCNIVNLSLVKPSFY  
VYSRVKNLNSS---RVPDLLV-----

**hits by patterns with a high probability of occurrence or by user-defined patterns: [3 hits (by 2 distinct patterns) on 1 sequence]****PS00001 ASN\_GLYCOSYLATION N-glycosylation site :**

**48 - 51:** NVSL

**Predicted feature:**

CARBOHYD 48 N-linked (GlcNAc...) asparagine [condition: N]

**66 - 69:** NSSR

**Predicted feature:**

CARBOHYD 66 N-linked (GlcNAc...) asparagine [condition: N]

**PS00005 PKC\_PHOSPHO\_SITE Protein kinase C phosphorylation site :**

**67 - 69:** SSR

**Predicted feature:**

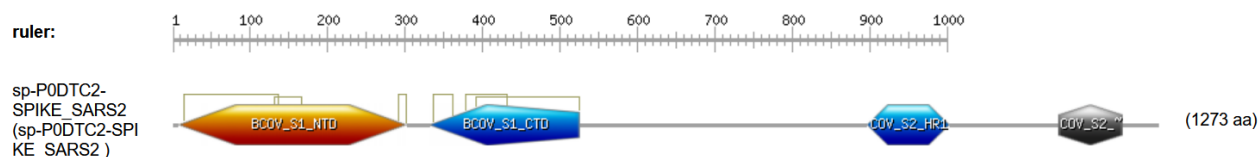
MOD\_RES 67 Phosphoserine [condition: S]

**Figure 21 :** Résultats d'analyse de la séquence protéique de petite protéine d'enveloppe de SARS-CoV-2 avec Prosite.

- Deux domaines pour glycoprotéine spike de *Betacoronavirus*, un domaine de sous unité S1 N-terminal de 9-303, contenant 3 ponts disulfures, respectivement entre cystéines : 15-136, 131-166 et 291-301, un domaine de sous unité S1 C-terminal de 334-527 contenant également 3 ponts disulfures entre cystéines : 336-361, 379-432 et 391-525 ;
- Deux domaines pour glycoprotéine spike de *Coronavirus*, un pour le profil de la région de sous unité 2 heptade répétée 1 (HR1) entre 896-1001 et l'autre est pour le profil de la région de sous unité 2 heptade répétée 2 (HR2) entre 1143-1225 ;

## hits by profiles: [4 hits (by 4 distinct profiles) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.



## PS51922 BCOV\_S1\_NTD Betacoronavirus spike (S) glycoprotein S1 subunit N-terminal (NTD) domain profile :

9 - 303: score = 49.266

PLVSSQCQVNLTR-----TQLPAYTNSFTRGVYYPDKVFRSSVLHSTQdIPLPFFSN  
VTWFAHIVSGTNGTKRFDP--VLPFNDGVYFASTEKSNIR-----GWIFGTTLDS  
KTQSLILIV---NNATNVVIVKVEFQFCNDPFLGVYHKNKSMWESEFRVYSSANNCTF  
EYvsqPFLMDLEGGKQGNFKnIrefvFNIDGYFKIYSKHTPInIvrdlpQGFSALEPLVD  
LPtGINITRFQTLALHRSYLTPgdssgWTAGAAAYVGYLQPRTFLLKYNEGTITDA  
VDcALDPLSETKcTL

## Predicted features:

|          |     |     |                |                     |
|----------|-----|-----|----------------|---------------------|
| DOMAIN   | 9   | 303 | BetaCoV S1-NTD | [condition: none]   |
| DISULFID | 15  | 136 |                | [condition: C-x*-C] |
| DISULFID | 131 | 166 |                | [condition: C-x*-C] |
| DISULFID | 291 | 301 |                | [condition: C-x*-C] |

## Absent feature:

|          |     |     |  |                              |
|----------|-----|-----|--|------------------------------|
| DISULFID | 143 | 245 |  | [condition not true: C-x*-C] |
|----------|-----|-----|--|------------------------------|

## PS51921 BCOV\_S1\_CTD Betacoronavirus spike (S) glycoprotein S1 subunit C-terminal (CTD) domain profile :

334 - 527: score = 64.510

NLCpFGEVFNATRFASVYAWNRRKRISNCVADYSVLVNSASFSTFKCYGVSPTKLNDLCFT  
NVYADSFVIRGDEVQRQIAPGQTGKIADYNYKLDPDFTGcVIAWNSNN----LDSKVGNY  
NYLYRLFRKSNLkpFERDISTEiYQAGSTPCNG-----VE-----GF  
NCYFPLQSYGFQPTNGVGYPYRWWLSFELLHAP-ATVcGP

## Predicted features:

|          |     |     |                |                     |
|----------|-----|-----|----------------|---------------------|
| DOMAIN   | 334 | 527 | BetaCoV S1-CTD | [condition: none]   |
| DISULFID | 336 | 361 |                | [condition: C-x*-C] |
| DISULFID | 379 | 432 |                | [condition: C-x*-C] |
| DISULFID | 391 | 525 |                | [condition: C-x*-C] |

## PS51923 COV\_S2\_HR1 Coronavirus spike (S) glycoprotein S2 subunit heptad repeat 1 (HR1) region profile :

896 - 1001: score = 50.311

IPFAMQMAYRFNGIGVTONVLYENQKLIANQFNSAIGKIQDSLSTASALGKLQDVVNQN  
AQLNLTLVKQLSSNFGAISSVLNDILSRDLKVEAEVQIDRLITGRL

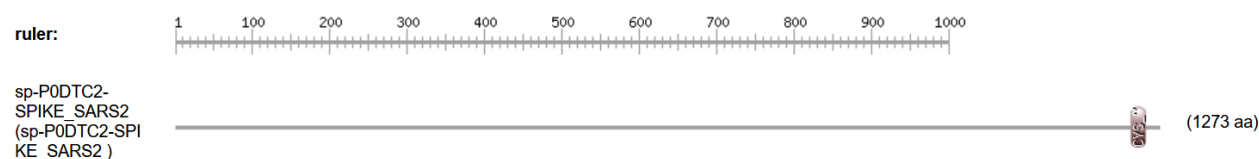
## PS51924 COV\_S2\_HR2 Coronavirus spike (S) glycoprotein S2 subunit heptad repeat 2 (HR2) region profile :

1143 - 1225: score = 29.395

PELDSFKEELDKYFKNHTSPDVLGDISGinASVVNIQKEIDRLNEVAKNLNESLIDLQE  
LGKYEQYIKWPYIWLGFIAGLI

## hits by profiles with a high probability of occurrence: [1 hit (by 1 profile) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.



## PS50311 CYS\_RICH Cysteine-rich region profile :

1235 - 1254: score = 14.622

CcmtscscclkgccscgscC

**Figure 22 A :** Résultats d'analyse de la séquence protéique de la protéine de spicule de SARS-CoV-2 avec Prosite.

- 22 sites de glycolisation de type N-liée, portés chacun par asparagine : 17, 61, 74, 122, 149, 165, 234, 282, 331, 343, 603, 616, 657, 709, 717, 801, 1074, 1098, 1134, 1158, 1173 et 1194 ;

hits by patterns with a high probability of occurrence or by user-defined patterns: [78 hits (by 6 distinct patterns) on 1 sequence]



**Figure 22 B :** Résultats d'analyse de la séquence protéique de la protéine de spicule de SARS-CoV-2 avec Prosite.

- 14 sites de phosphorylation par protéine kinase C portés par Thréonines : 19, 76, 95, 302, 376, 415, 632, 998, 1105 et sérines : 555, 680, 813, 1037 et 1147 ;

- 14 sites de phosphorylation par caséine kinase II, qui sont portés par sérines : 50, 151, 221, 816, 982, 1147, 1196 et thréonines : 108, 250, 284, 734, 827, 1136 et 1160 ;

|                                                                                |        |                  |                |                                                                                            |        |                  |                |
|--------------------------------------------------------------------------------|--------|------------------|----------------|--------------------------------------------------------------------------------------------|--------|------------------|----------------|
| <b>PS00005 PKC_PHOSPHO_SITE</b> <i>Protein kinase C phosphorylation site :</i> |        |                  |                | <b>PS00006 CK2_PHOSPHO_SITE</b> <i>Casein kinase II phosphorylation site :</i>             |        |                  |                |
| 19 - 21:                                                                       | TtR    |                  |                | 50 - 53:                                                                                   | StqD   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 19     | Phosphothreonine | [condition: T] | MOD_RES                                                                                    | 50     | Phosphoserine    | [condition: S] |
| 76 - 78:                                                                       | TkR    |                  |                | 108 - 111:                                                                                 | TtLD   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 76     | Phosphothreonine | [condition: T] | MOD_RES                                                                                    | 108    | Phosphothreonine | [condition: T] |
| 95 - 97:                                                                       | TeK    |                  |                | 151 - 154:                                                                                 | SvME   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 95     | Phosphothreonine | [condition: T] | MOD_RES                                                                                    | 151    | Phosphoserine    | [condition: S] |
| 302 - 304:                                                                     | TlK    |                  |                | 221 - 224:                                                                                 | SalE   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 302    | Phosphothreonine | [condition: T] | MOD_RES                                                                                    | 221    | Phosphoserine    | [condition: S] |
| 376 - 378:                                                                     | TfK    |                  |                | 250 - 253:                                                                                 | TpgD   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 376    | Phosphothreonine | [condition: T] | MOD_RES                                                                                    | 250    | Phosphothreonine | [condition: T] |
| 415 - 417:                                                                     | TgK    |                  |                | 284 - 287:                                                                                 | Tltd   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 415    | Phosphothreonine | [condition: T] | MOD_RES                                                                                    | 284    | Phosphothreonine | [condition: T] |
| 555 - 557:                                                                     | SnK    |                  |                | 734 - 737:                                                                                 | TsvD   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 555    | Phosphoserine    | [condition: S] | MOD_RES                                                                                    | 734    | Phosphothreonine | [condition: T] |
| 632 - 634:                                                                     | TwR    |                  |                | 816 - 819:                                                                                 | SfIE   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 632    | Phosphothreonine | [condition: T] | MOD_RES                                                                                    | 816    | Phosphoserine    | [condition: S] |
| 680 - 682:                                                                     | SpR    |                  |                | 827 - 830:                                                                                 | TlAd   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 680    | Phosphoserine    | [condition: S] | MOD_RES                                                                                    | 827    | Phosphothreonine | [condition: T] |
| 813 - 815:                                                                     | SkR    |                  |                | 982 - 985:                                                                                 | SrLD   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 813    | Phosphoserine    | [condition: S] | MOD_RES                                                                                    | 982    | Phosphoserine    | [condition: S] |
| 998 - 1000:                                                                    | TgR    |                  |                | 1136 - 1139:                                                                               | TvyD   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 998    | Phosphothreonine | [condition: T] | MOD_RES                                                                                    | 1136   | Phosphothreonine | [condition: T] |
| 1037 - 1039:                                                                   | SkR    |                  |                | 1147 - 1150:                                                                               | SfKE   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 1037   | Phosphoserine    | [condition: S] | MOD_RES                                                                                    | 1147   | Phosphoserine    | [condition: S] |
| 1105 - 1107:                                                                   | TqR    |                  |                | 1160 - 1163:                                                                               | TspD   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 1105   | Phosphothreonine | [condition: T] | MOD_RES                                                                                    | 1160   | Phosphothreonine | [condition: T] |
| 1147 - 1149:                                                                   | SfK    |                  |                | 1196 - 1199:                                                                               | SlID   |                  |                |
| Predicted feature:                                                             |        |                  |                | Predicted feature:                                                                         |        |                  |                |
| MOD_RES                                                                        | 1147   | Phosphoserine    | [condition: S] | MOD_RES                                                                                    | 1196   | Phosphoserine    | [condition: S] |
| <b>PS00008 MYRISTYL</b> <i>N-myristoylation site</i>                           |        |                  |                | <b>PS00004 CAMP_PHOSPHO_SITE</b> <i>cAMP-dependent protein kinase phosphorylation site</i> |        |                  |                |
| 476 - 481:                                                                     | GStpCN |                  |                | 356 - 359:                                                                                 | KRIS   |                  |                |
| 72 - 77:                                                                       | GTngTK | 545 - 550:       | GLtgTG         | 528 - 531:                                                                                 | KKST   |                  |                |
| 89 - 94:                                                                       | GVyFAS | 601 - 606:       | GTntSN         |                                                                                            |        |                  |                |
| 184 - 189:                                                                     | GNfKNL | 648 - 653:       | GClIGA         |                                                                                            |        |                  |                |
| 232 - 237:                                                                     | GInItr | 667 - 672:       | GAgICA         |                                                                                            |        |                  |                |
| 311 - 316:                                                                     | GIyqTS | 669 - 674:       | GICAyS         |                                                                                            |        |                  |                |
| 381 - 386:                                                                     | GVspTK | 700 - 705:       | GAenSV         |                                                                                            |        |                  |                |
| 431 - 436:                                                                     | GCvIAW | 757 - 762:       | GSfcTQ         |                                                                                            |        |                  |                |
| 446 - 451:                                                                     | GGnyNY | 880 - 885:       | GTitSG         |                                                                                            |        |                  |                |
|                                                                                |        |                  |                | 889 - 894:                                                                                 | GAgAAL |                  |                |
|                                                                                |        |                  |                | 908 - 913:                                                                                 | GIGvTQ |                  |                |
|                                                                                |        |                  |                | 910 - 915:                                                                                 | GVtqNV |                  |                |
|                                                                                |        |                  |                | 971 - 976:                                                                                 | GAisSV |                  |                |
|                                                                                |        |                  |                | 1093 - 1098:                                                                               | GVfvSN |                  |                |
|                                                                                |        |                  |                | 1131 - 1136:                                                                               | GIVnNT |                  |                |
|                                                                                |        |                  |                | 1171 - 1176:                                                                               | GInaSV |                  |                |
|                                                                                |        |                  |                | 1246 - 1251:                                                                               | GCcsCG |                  |                |

**Figure 22 C :** Résultats d'analyse de la séquence protéique de la protéine de spicule de SARS-CoV-2 avec Prosite.

- 25 sites de N-myristoylation caractérisant les glycines : 72, 89, 184, 232, 311, 381, 431, 446, 476, 545, 601, 648, 667, 669, 700, 757, 880, 889, 908, 910, 971, 1093, 1131, 1171 et 1246 et qui servent probablement à l'encrage avec la membrane de l'enveloppe viral ;
- Deux sites de phosphorylation par protéine kinase AMPc dépendent pour le motif (356)KRiS(359) et (528)KKsT(531) ;
- Et finalement, un motif d'attachement cellulaire représenté par la séquence (403)RGD(405).

### 3.2.5. Polyprotéine réplique 1a

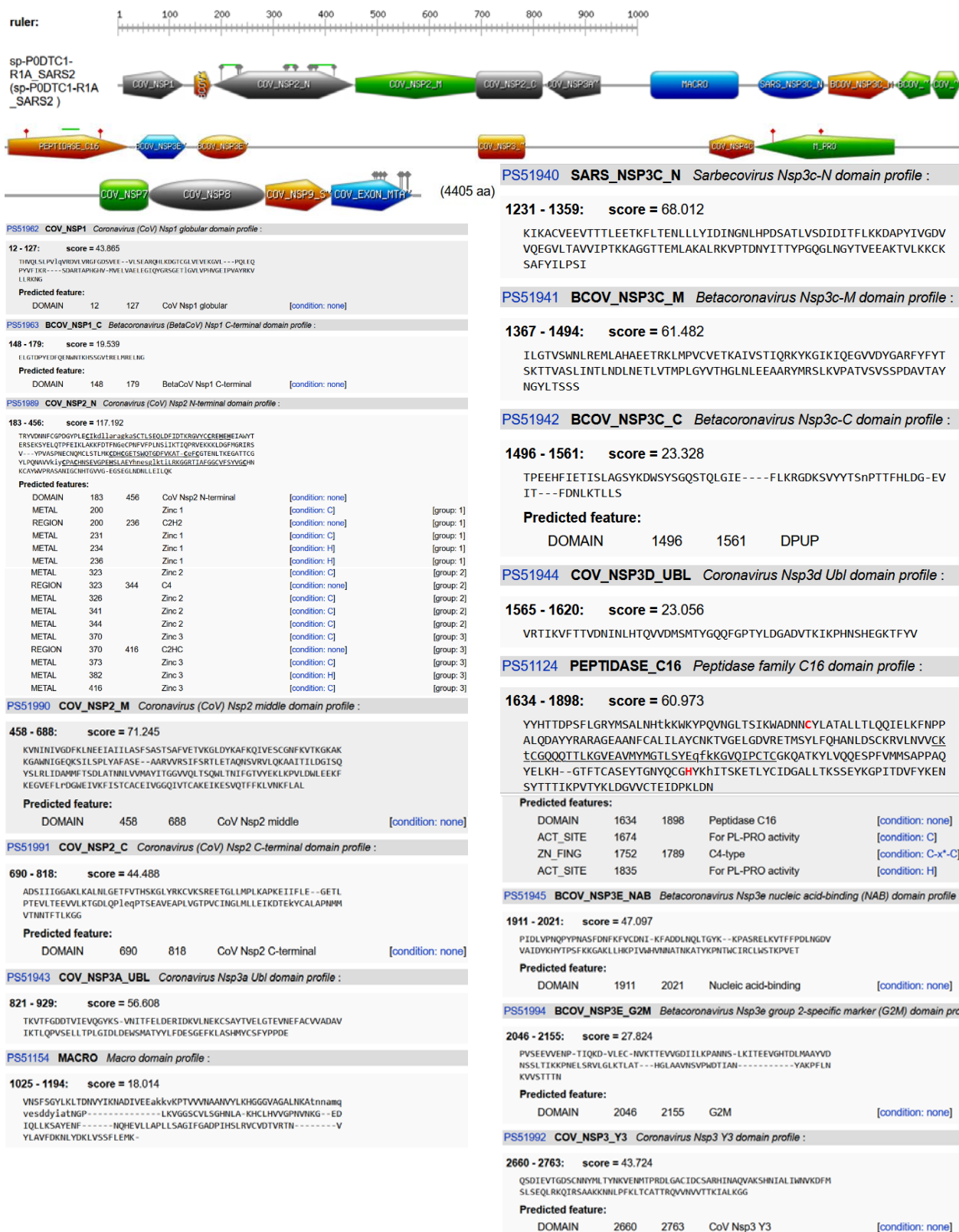
La figure 23 A et 23 B montrent les domaines et motifs protéiques retrouvés suite à l'analyse par Prosite pour la polyprotéine réplique 1a.

Les domaines et motifs retrouvés sont :

- Domaine globulaire Nsp1 de *Coronavirus* de la position 12 à 127 ;
- Domaine Nsp1 C-terminal de *Betacoronavirus* de la position 148 à 179 ;
- Domaine Nsp2 N-terminal de *Coronavirus* de la position 183 à 456 contenant 3 régions de doigts de zinc occupant les positions : 200-236, 323-344 et 370-416 regroupant au total 12 résidus portant de zincs : C(200), C(231), H(234), H(236), C(323), C(326), C(341), C(344), C(370), C(373), H(382) et C(416) ;
- Domaine moyen Nsp2 de *Coronavirus* de la position 458 à 688 ;
- Domaine Nsp2 C-terminal de *Coronavirus* de la position 690 à 818 ;
- Domaine Nsp3a Ubl de *Coronavirus* de la position 821 à 929 ;
- Macro domaine de la position 1025 à 1194 ;
- Domaine Nsp3c-N de *Sarbecovirus* de la position 1231 à 1359 ;
- Domaine Nsp3c-M de *Betacoronavirus* de la position 1367 à 1494 ;
- Domaine Nsp3c-C de *Betacoronavirus* de la position 1496 à 1561 ;
- Domaine Nsp3d Ubl de *Coronavirus* de la position 1565 à 1620 ;
- Domaine de l'appartenance à la famille de peptidases C16 de la position 1634 à 1898 contenant : un site actif qui composé de deux résidus d'acide aminé C(1674) et H(1835) d'activité papaine like protéase, un doigt de zinc de type C4 de la position 1752 à 1789 ;
- Domaine de liaison à l'acide nucléique Nsp3e de *Betacoronavirus* de la position 1911 à 2021 ;

## hits by profiles: [21 hits (by 21 distinct profiles) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.



**Figure 23 A :** Résultats d'analyse de la séquence protéique de polyprotéine réplase 1a de SARS-CoV-2 avec Prosite.

- Domaine Nsp3e de groupe de marqueur spécifique 2 de *Betacoronavirus* de la position 2046 à 2155 ;
- Domaine Nsp3 Y3 de *Coronavirus* de la position 2660 à 2763 ;
- Domaine Nsp4 C-terminal de *Coronavirus* de la position 3165 à 3263 ;
- Domaine de protéase principale de *Betacoronavirus* de la position 3264 à 3569, contenant un site actif à activité protéase 3CL composé de deux résidus aminoacides : H(3304) et C(3408) ;
- Domaine du cofacteur Nsp7 de l'ARN polymérase dépendante de l'ARN du *Coronavirus* de la position 3860 à 3942 ;

PS51946

COV\_NSP4C

Coronavirus Nsp4 C-terminal (Nsp4C) domain profile :

3165 - 3263:

score = 40.913

VVFNGVSFSTFEAALCTFLNKENYLLKRSVDLPLTQYRRLALYNYKYFSGAPDIT  
SYREAACHLAKALNDFSHSGSDVLYQPPTTSITSAVLQ

Predicted feature:

DOMAIN

3165

3263

Nsp4C

[condition: none]

PS51442

M\_PRO

Coronavirus main protease (M-pro) domain profile :

3264 - 3569:

score = 154.193

SGFRKMAFPSGKVEGCMVQTCGTTTLNGLHLDVVYCPRICTSEDLNPNYEDLLIR  
KSNHNF LQAGNVQLRVIGHSMQNCVLLKVDITANPKTPKYFVRIQGGTFSVLACYNG  
SPSGVYQCAHRPNFTIKGSFLNGSCGSGVFNIDYDCVSFCYBHHMLPTGVHAGTDLEGN  
FYGPFDVRQTAQAAGTDTTITVNVLAHLVAAVINGDRHFLNRFTTTLNDFNLVAMKYNYE  
PLTQDHVGLGLPLSAQTGIAVLNDCASLKEELLQNGHNGRTILGSLLEDEFPTFDVVRQC  
SGVTTFQ

Predicted features:

DOMAIN

3264

3569

Peptidase C30

[condition: none]

ACT\_SITE

3304

For 3CL-PRO activity

[condition: H]

ACT\_SITE

3408

For 3CL-PRO activity

[condition: C]

PS51949

COV\_NSP7

Coronavirus RNA-dependent RNA polymerase (RdRp) Nsp7 cofactor domain profile :

3860 - 3942:

score = 32.810

SKMSDVKCTSVLLSVLQURVSSSKLWACVQLHNDILLAKDTTEAFKRWSSLLSVLL  
SHQGAVDINKLCEEMDRATLQ

Predicted feature:

DOMAIN

3860

3942

RdRp Nsp7 cofactor

[condition: none]

PS51950

COV\_NSP8

Coronavirus RNA-dependent RNA polymerase (RdRp) Nsp8 cofactor domain profile :

3943 - 4140:

score = 75.721

ATASEFSSLPSYAAAFATAQEAYEQAVANGDSFVLLKKLKSUNAKSEFORDAANQRKLE  
KHADQANTQGYKQARSSEKRAKVTSAQTMFLFHLRLDNDALNINIINARGDCVPLNII  
PLTTAAKLHVVIPQYNTYKNTCDGTTFTYASALHEIQQVVDADSKIVQLSEISMDNSPIL  
AWPLIVTALRAHSAVKLQ

Predicted feature:

DOMAIN

3943

4140

RdRp Nsp8 cofactor

[condition: none]

PS51951

COV\_NSP9\_SSRNA\_BD

Coronavirus Nsp9 single-stranded RNA (ssRNA)-binding domain profile :

4141 - 4253:

score = 47.646

HNELSPVALRQHSAAAGTTQTACTDNALGYYNTTKGGRFVLLALLSDQLKMARFPKSD  
GTGTIYTELEPPCRFVTDTPKGPVKVLYFIKGLNHLNRGNVLGSLAATVRLQ

Predicted feature:

DOMAIN

4141

4253

Nsp9 ssRNA-binding

[condition: none]

PS51952

COV\_EXON\_MTASE\_COACT

Coronavirus (CoV) ExoN/MTase coactivator domain profile :

4254 - 4392:

score = 61.109

AGNATEVPANISTVLSCAFVDAKAYKDYASGGQPIITNCVQMLCTHTGTGQAITVTPE  
AHMDQSFSGGASCCLYCRCHIDHPNPKGFCDLKGKVVQIPTTCANDPVGFTLKNTVCTVC  
GNWKGVCSCDQLREPPLQ

Predicted features:

DOMAIN

4254

4392

ExoN/MTase coactivator

[condition: none]

METAL

4327

Zinc 1

[condition: C]

METAL

4330

Zinc 1

[condition: C]

METAL

4336

Zinc 1

[condition: H]

METAL

4343

Zinc 1

[condition: C]

METAL

4370

Zinc 2

[condition: C]

METAL

4373

Zinc 2

[condition: C]

METAL

4381

Zinc 2

[condition: C]

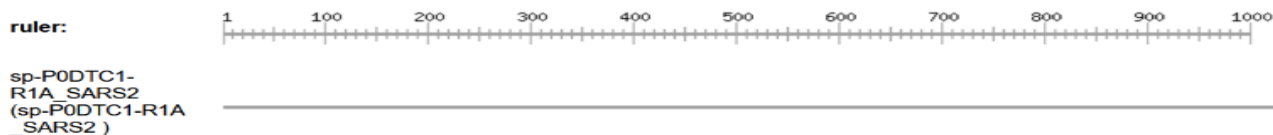
METAL

4383

Zinc 2

[condition: C]

hits by patterns: [1 hit (by 1 pattern) on 1 sequence]



**PS00867 CPSASE\_2** Carbamoyl-phosphate synthase subdomain signature 2 :

**2061 - 2068:** [confidence level: (-1)] VLECNVKT

(4405 aa)

**Figure 23 B :** Résultats d'analyse de la séquence protéique de polyprotéine réplacase 1a de SARS-CoV-2 avec Prosite.

- Domaine du cofacteur Nsp8 de l'ARN polymérase dépendante de l'ARN du *Coronavirus* de la position 3943 à 4140 ;
- Domaine Nsp9 de liaison à l'ARN simple brin de *Coronavirus* de la position 4141 à 4253 ;
- Domaine coactivateur ExoN/MTase de *Coronavirus* de la position 4254 à 4392, contenant 8 résidus porteurs de Zn : C(4327), C(4330), H(4336), C(4343), C(4370), C(4373), C(4381), C(4383) ;
- Sous-domaine de signature de la carbamoyl-phosphate synthase 2 de motif VLECNVKT, occupant les positions de 2061 à 2068.

### 3.2.6. Polyprotéine réplase 1ab

La figure 24 A et 24 B montrent les domaines et motifs protéiques retrouvés suite à l'analyse par Prosite pour la polyprotéine réplase 1ab.

Les domaines et motifs retrouvés sont :

- Domaine globulaire Nsp1 de *Coronavirus* de la position 12 à 127 ;
- Domaine Nsp1 C-terminal de *Betacoronavirus* de la position 148 à 179 ;
- Domaine Nsp2 N-terminal de *Coronavirus* de la position 183 à 456 contenant 3 régions de doigts de zinc occupant les positions : 200-236, 323-344 et 370-416 regroupant au total 12 résidus portant de zincs : C(200), C(231), H(234), H(236), C(323), C(326), C(341), C(344), C(370), C(373), H(382) et C(416) ;
- Domaine moyen Nsp2 de *Coronavirus* de la position 458 à 688 ;
- Domaine Nsp2 C-terminal de *Coronavirus* de la position 690 à 818 ;
- Domaine Nsp3a Ubl de *Coronavirus* de la position 821 à 929 ;
- Macro domaine de la position 1025 à 1194 ;
- Domaine Nsp3c-N de *Sarbecovirus* de la position 1231 à 1359 ;
- Domaine Nsp3c-M de *Betacoronavirus* de la position 1367 à 1494 ;
- Domaine Nsp3c-C de *Betacoronavirus* de la position 1496 à 1561 ;
- Domaine Nsp3d Ubl de *Coronavirus* de la position 1565 à 1620 ;
- Domaine de l'appartenance à la famille de peptidases C16 de la position 1634 à 1898 contenant : un site actif qui composé de deux résidus d'acide aminé C(1674) et H(1835) d'activité papaine like protéase, un doigt de zinc de type C4 de la position 1752 à 1789 ;
- Domaine de liaison à l'acide nucléique Nsp3e de *Betacoronavirus* de la position 1911 à 2021 ;

- Domaine Nsp3e de groupe de marqueur spécifique 2 de *Betacoronavirus* de la position 2046 à 2155 ;
- Domaine Nsp3 Y3 de *Coronavirus* de la position 2660 à 2763 ;
- Domaine Nsp4 C-terminal de *Coronavirus* de la position 3165 à 3263 ;
- Domaine de protéase principale de *Betacoronavirus* de la position 3264 à 3569, contenant un site active à activité protéase 3CL composé de deux résidus aminoacides : H(3304) et C(3408) ;
- Domaine du cofacteur Nsp7 de l'ARN polymérase dépendante de l'ARN du *Coronavirus* de la position 3860 à 3942 ;
- Domaine du cofacteur Nsp8 de l'ARN polymérase dépendante de l'ARN du *Coronavirus* de la position 3943 à 4140 ;
- Domaine Nsp9 de liaison à l'ARN simple brin de *Coronavirus* de la position 4141 à 4253 ;
- Domaine coactivateur ExoN/MTase de *Coronavirus* de la position 4254 à 4392, contenant 8 résidus porteurs de Zn : C(4327), C(4330), H(4336), C(4343), C(4370), C(4373), C(4381), C(4383) ;
- Domaine de la nucléotidyl transférase (NiRAN) associée au *Nidovirus* RdRp de la position 4399 à 4653 ;
- Domaine Nsp12 de l'ARN polymérase dépendante de l'ARN (RdRp) de *Coronavirus* de la position 4757 à 5324, contenant un site actif composé de 3 résidus aminoacides S(5151), D(5152) et D(5153) ;
- Domaine catalytique RdRp de ssRNA positif des virus de la position 5004 à 5166 ;
- Domaine de liaison au zinc des *Coronaviridae* (CV ZBD) de la position 5325 à 5408, contenant 12 résidus porteurs de zincs : C(5329), C(5332), C(5340), C(5343), C(5350), C(5353), H(5357), H(5363), C(5374), C(5379), C(5396) et H(5399) ;
- Domaine de l'hélicase centrale de virus à ARN (+), de la position 5581 à 5932, contenant un domaine hélicase liaison à l'ATP pour des virus à ARN (+) de la position 5581 à 5781 et un motif de liaison au nucléotide triphosphate de la position 5606 à 5613 ;
- Domaine exoribonucléase 3'-5' (ExoN) de *Nidovirus*, de la position 5997 à 6212, contenant un site actif à 5 résidus aminoacides : D(6015), E(6017), E(6116), H(6193) et D(6198) et 8 résidus aminoacides pour liaison au zinc : C(6132), C(6135), C(6151), H(6154), H(6182), C(6186), H(6189) et C(6204) ;

- Domaine guanine-N7-méthyltransférase (N7-MTase) du *Coronavirus* de la position 6221 à 6452, contenant une région de liaison au sterile alpha motif (SAM) de la position 6256 à 6262, une région de liaison au Guanosine-P3-adenosine-5',5'-triphosphate (GpppA-binding) et 4 résidus aminoacides pour liaison au zinc : C(6377), C(6398), C(6409) et H(6412) ;
- Domaine Nsp15 d'oligomérisation N-terminal de *Coronavirus* de la position 6453 à 6513 ;
- Domaine Nsp11 N-terminal de *Arterivirus* / NSP15 moyen *Coronavirus* (AV-Nsp11N/CoV-Nsp15M), de la position 6514 à 6639 ;
- Domaine endoribonucléase spécifique de l'uridylate nidoviral (NendoU), de la position 6656 à 6795, contenant un site actif de 3 résidus aminoacides : H(6686), H(6701) et K(6741) ;
- Domaine 2'-O-méthyltransférase (2'-O-MTase) de *Nidovirus*, de la position 6800 à 7094, contenant un site actif composé de 4 résidus aminoacides : K(6844), D(6928), K(6968) et E(7001) ;
- Sous-domaine de signature de la carbamoyl-phosphate synthase 2 de motif VLECNVKT, occupant les positions de 2061 à 2068 ;
- Un domaine de lipocaline de motif GTS..KFYGGWHNM, occupant la position de 4982 à 4993.

**ruler:**

sp-P0DTD1-  
R1AB\_SARS2  
(sp-P0DTD1-R1A  
B\_SARS2 )

(7096 aa)

**3165 - 3263:** score = 40.913

VVFNGVSTFEEALCTFLINKEHYKLRSDVILPLTQYHRYLALYNYKYFSGANDTT  
SYREACCHLAKALNDFSHSGSDVLQPPQTSITSAVLQ

**Predicted feature:**

| DOMAIN | 3165 | 3263 | Nsp4C |                   |
|--------|------|------|-------|-------------------|
|        |      |      |       | [condition: none] |

47

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |  |  |  |  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|
| <b>PS51442 M_PRO</b> Coronavirus main protease (M-pro) domain profile :<br><br>3264 - 3569: score = 154.193<br><br>SGRFRAIPSGVGEVGVQVTCGTTTLNGLWLDVVYCPRIWICTSEDMNPVYDILLR<br>KSHRIELVQAGVQLRVIEGSHQICVLLKLVQDANPKPKYKFKIPGQITFSLACVNG<br>SPSGVQVCARRPIETEGSF LIGSGVSGVGLRDYDVSVCYVRIHELPTQWAGTDLGN<br>FYGPVVGQTAGAGCTDITITVWVLAAYVAGIDIRHILHRTITTLIDHRLVAPRYVE<br>PLIQDRIWELGPIASQITGAVDRCASKEELLEQWERTITLQALIDETIPIDVINGE<br>SVYFIQ<br><br>Predicted features:<br>DOMAIN 3264 3569 Peptidase C30 [condition: none]<br>ACT_SITE 3304 For 3CL-PRO activity [condition: H] [group: 1]<br>ACT_SITE 3408 For 3CL-PRO activity [condition: C] [group: 1]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |  |  |  |  |
| <b>PS51949 COV_NSP7</b> Coronavirus RNA-dependent RNA polymerase (RdRp) Nsp7 cofactor domain profile :<br><br>3860 - 3942: score = 32.810<br><br>SKPSDYKTSVLSVLQQLRVSSSKLMAQCVLHIDELLAKDTFAEKWYSLLSVLL<br>SHQAVDINKLCEHLDNRATLQ<br><br>Predicted feature:<br>DOMAIN 3860 3942 RdRp Nsp7 cofactor [condition: none]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |  |  |  |  |
| <b>PS51950 COV_NSP8</b> Coronavirus RNA-dependent RNA polymerase (RdRp) Nsp8 cofactor domain profile :<br><br>3943 - 4140: score = 75.721<br><br>ATASEFSLSPTAAFAATQAEAYEQWANGDSYVLKKLKSINAKSEFDRDAWQRIE<br>YMGQWATQPYQASSEDAKAVTSANQITLFTPLRLIDRALNITIRARDCVPLNLT<br>PLTTAAKLHVIVDYNVYKTCGGTETTYASALMEIQGVVADSKVQLSEISDKSPNHL<br>AHLPLVTALRAISAVKLQ<br><br>Predicted features:<br>DOMAIN 3943 4140 RdRp Nsp8 cofactor [condition: none]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |  |  |  |  |
| <b>PS51951 COV_NSP9_SSRNA_BD</b> Coronavirus Nsp9 single-stranded RNA (ssRNA)-binding domain profile :<br><br>4141 - 4253: score = 47.646<br><br>NHLSFVALRQPSACAGTQTACTDNNALYNTTKGGRVLLASDLQDLMANPKPSD<br>GTGTITTELPPECFREVDITPKGPKVCKYVETKGLNMLNGVLSGLAATVRLQ<br><br>Predicted feature:<br>DOMAIN 4141 4253 Nsp9 ssRNA-binding [condition: none]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |  |  |  |  |
| <b>PS51952 COV_EXON_MTASE_COACT</b> Coronavirus (CoV) ExoNMTase coactivator domain profile :<br><br>4254 - 4392: score = 61.109<br><br>AGNATVPANSTVLSTCAFAVDAKAYKYLASGGGPTINCVMCTHTTGQATVTPTPE<br>AMQDESGFGASCLYCRCHDHPNPKGFDLKGKVGQVTPDTCANDPVGFTLQVTCVC<br>GPAKVGSGCDQLREPMQL<br><br>Predicted features:<br>DOMAIN 4254 4392 ExoNMTase coactivator [condition: none]<br>METAL 4327 Zinc 1 [condition: C] [group: 1]<br>METAL 4330 Zinc 1 [condition: C] [group: 1]<br>METAL 4336 Zinc 1 [condition: H] [group: 1]<br>METAL 4343 Zinc 1 [condition: C] [group: 1]<br>METAL 4370 Zinc 2 [condition: C] [group: 2]<br>METAL 4373 Zinc 2 [condition: C] [group: 2]<br>METAL 4381 Zinc 2 [condition: C] [group: 2]<br>METAL 4383 Zinc 2 [condition: C] [group: 2]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |  |  |  |  |
| <b>PS51947 NIRAN</b> Nidovirus RdRp-associated nucleotidyl transferase (NIRAN) domain profile :<br><br>4399 - 4653: score = 30.066<br><br>FLIRVCG-VSAAR-LTPCGTGSTGVVYRMQDYNKDVAGNKLCTRCRFLQIKEDDIN<br>LIGSYVYVORHITSNQHREELVNLIGGCPWAKCHDFKRLKGGWPHLSHQRLTKYRM<br>ADLVYALRVDFGRGCTLKELVLYNYCCDDYRCKQKQSGFVengdLLVymALGGIRMQ<br>ALLCTQVCDANNACTQVALLDMQDLNBNADYGGTQTTRSGVAVGVDSYSLILRPL<br>TLTTRALTAEMNDETL<br><br>Predicted features:<br>DOMAIN 4399 4653 NIRAN [condition: none]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |  |  |  |  |
| <b>PS51948 COV_NSP12_RDRP</b> Coronavirus Nsp12 RNA-dependent RNA polymerase (RdRp) domain profile :<br><br>4757 - 5324: score = 314.272<br><br>RLSFKELLYVADNPVHNSGBILLDQRTTCTSVNALNHWAFQIVQVQNFNDQYDAFY<br>SKGFTREGSSVLIWFTTAGDGAALSDQYPRYALPTPCEDIRGLFVQVADNYTCVD<br>GCCTANQVTVRIIDCSAGFPFNGAGKLVYDSVYEDDALFAYTQVRYVPTTITQRIEL<br>KYATSANIRATVAGUSCTSTNIRGFHQELKLSAATGATVAVTGTSCFYGVANRHLKT<br>VYSCVIRPLKADPYPCORANRPHLRKASLVLRKHCTCSSLSHRYRLANCAQVLS<br>ENPFGASGAYRQKGSAGRTTAYAVDSVRLCRVFAHRAVLSLSGAGLDRYVIRMLQ<br>HRLVELCYRRADGSDTDRVLSYAYLRGHSPHLSGGVACVRSYASQVSLSDHKS<br>VLYVQRYRPSKACTETDLTKGPIEFCSQHTLVKGGQGVVYVPYPSATLGAGCTV<br>DDVKTGDTLHFERFVSLATDAVPLTQPHQFVAVTHVLYQYTRHDLTGNRDRYS<br>VNLTHRTSYRGEFFPYEARVPTHTVQLQ<br><br>Predicted features:<br>DOMAIN 4757 5324 Nsp12 RNA-dependent RNA polymerase [condition: none]<br>ACT_SITE 5151 [condition: S] [group: 1]<br>ACT_SITE 5152 [condition: D] [group: 1]<br>ACT_SITE 5153 [condition: D] [group: 1]                                                                                                                                                                                                                                                                             |  |  |  |  |
| <b>PS50507 RDRP_SSRNA_POS</b> RdRp of positive ssRNA viruses catalytic domain profile :<br><br>5004 - 5166: score = 8.290<br><br>PHLMGMDYPPKCRANPRH-LRIPASLVLRKHITCCSLSHRYRLANCAQVLSERWKG<br>GSLYKVGPGTSSGDAATAYAVSVRLCQAVTANNVLLstdgnLadkymLqhrLYec<br>LymrdvdtfdmefyAVLRHSPHLLSDQAVCVNSYASQG<br><br>PS51653 CV_ZBD Coronaviridae zinc-binding (CV ZBD) domain profile :<br><br>5325 - 5408: score = 33.035<br><br>AVGACVLCISQTSILRGACTRRPFLCCCKCYDMVTLSKRLVLSNPVYVAPACGVTVV<br>TQLVLGGHSYVCKSNKPTSPFLC<br><br>Predicted features:<br>DOMAIN 5325 5408 CV ZBD [condition: none]<br>METAL 5329 Zinc 1 [condition: C] [group: 1]<br>METAL 5332 Zinc 1 [condition: C] [group: 1]<br>METAL 5340 Zinc 2 [condition: C] [group: 1]<br>METAL 5343 Zinc 1 [condition: C] [group: 1]<br>METAL 5350 Zinc 1 [condition: [CH]] [group: 1]<br>METAL 5353 Zinc 2: via tele nitrogen [condition: [CH]] [group: 2]<br>METAL 5357 Zinc 2: via pros nitrogen [condition: [CH]] [group: 2]<br>METAL 5363 Zinc 2 [condition: [CH]] [group: 2]<br>METAL 5374 Zinc 3 [condition: C] [group: 2]<br>METAL 5379 Zinc 3: via pros nitrogen [condition: [CH]] [group: 3]<br>METAL 5396 Zinc 3 [condition: C] [group: 3]<br>METAL 5399 Zinc 3 [condition: [CH]] [group: 3] |  |  |  |  |
| <b>PS51657 PSRV_HELICASE</b> (+)RNA virus helicase core domain profile :<br><br>5581 - 5932: score = 27.299<br><br>NLSDEFSSHVAVYVGVG-----MKEYSTLQGPSTGSHFAIGLALVYPSARIVYAC<br>SHAANDALCEALYKLPDKCSLIpararvecDKKYNISLLeqyVCTynaIactITAD<br>LVVDELISPAINYLLSVYMARIRAKIPVYIGDPAOLPAIRTLITGLTLepeyfnsvRLM<br>KTIGDPELGTCHRCPAIIVDTYSALVYVAKLKARDKSAGCT--KRPVKGVTI-----<br>-HDVvaItrpLIGVYVFTItrpwrKavFIsgyngprvarkLIGLITQTVDSQSGE<br>YDVVITFTQTETAS--DNARFINATRAKVGILCTESORDLYKLGFTSLERPRNAT<br>LQENVTGL<br><br>Predicted features:<br>DOMAIN 5581 5762 (+)RNA virus helicase ATP-binding [condition: none]<br>NP_BIND 5806 5613 NTP [condition: none]<br><br>Absent feature:<br>- [condition not true: unknown]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |  |  |  |  |
| <b>PS51953 NIV_EXON</b> Nidovirus 3'-5' exonuclease (ExoN) domain profile :<br><br>5997 - 6212: score = 62.876<br><br>NFITREAAIRVIRAMZGIVEGCHATREAVGTRILPLQLGFSGVHVAWPTGYVDTPMIT<br>DIFSRSAKPPGQGHKILPLHYKGLPHNVRVILIKVQLSDTLKHLSORWVYVLMHKGFE<br>LTSQYVYKIGPETERCTCLCORRATCFSTASDYAGHHISIGFQVYVYPRKIDVQKMGFTG<br>NLGSIMDLVYQ-VHGNVAVSCDATTRELAEHCTV<br><br>Predicted features:<br>DOMAIN 5997 6212 ExoN [condition: none]<br>ACT_SITE 6015 [condition: D] [group: 1]<br>ACT_SITE 6017 [condition: E] [group: 1]<br>ACT_SITE 6116 [condition: [ED]] [group: 1]<br>METAL 6132 Zinc 1 [condition: C] [group: 2]<br>METAL 6135 Zinc 1 [condition: C] [group: 2]<br>METAL 6151 Zinc 1 [condition: C] [group: 2]<br>METAL 6154 Zinc 1 [condition: H] [group: 2]<br>METAL 6182 Zinc 2 [condition: C] [group: 3]<br>METAL 6186 Zinc 2 [condition: C] [group: 3]<br>METAL 6189 Zinc 2 [condition: H] [group: 3]<br>ACT_SITE 6193 [condition: H] [group: 1]<br>ACT_SITE 6198 [condition: D] [group: 1]<br>METAL 6204 Zinc 2 [condition: C] [group: 3]                                                                                                                                                                                                    |  |  |  |  |
| <b>PS51954 COV_N7_MTASE</b> Coronavirus (CoV) guanine-N7-methyltransferase (N7-MTase) domain profile :<br><br>6221 - 6452: score = 110.466<br><br>VPTIGDELKINAKRCVQ#RVVVAALLADKFFVLVHDLG#EADKCVYQADQWRFYDQAP<br>CSQKAKLEELFYVATYHSGKFTDQCLPARKWRYVWMSICRFDTRVLSMLKGLQ<br>SGSLYVNHKAEHTPAFQKSAFVNLKQLPFYFYSDSCESHGQVSDIDYVPLKSACTET<br>RCLLGGAVCRIMAEYRLYLDVNMPLSAGFLMWYKQFDYVNLNMTITRLQ<br><br>Predicted features:<br>DOMAIN 6221 6452 N7-MTase [condition: none]<br>BINDING 6256 6262 SAM [condition: D-x-G-x-P-x-[GA]] [group: 1]<br>REGION 6339 6353 GpppA-binding [condition: none]<br>METAL 6377 Zinc [condition: C] [group: 1]<br>METAL 6398 Zinc [condition: C] [group: 1]<br>METAL 6409 Zinc [condition: C] [group: 1]<br>METAL 6412 Zinc [condition: H] [group: 1]                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |  |  |  |  |
| <b>PS51960 COV_NSP15_NTD</b> Coronavirus (CoV) Nsp15 N-terminal oligomerization domain profile :<br><br>6453 - 6513: score = 31.105<br><br>SLENAVRHWKGFDSQGGQVPSVITINNTVYKGVSGVFNELKTTLPWVAWFLMAK<br>R<br><br>Predicted feature:<br>DOMAIN 6453 6513 Nsp15 N-terminal oligomerization [condition: none]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |  |  |  |  |
| <b>PS51961 AV_NSP11N_COV_NSP15M</b> Arterivirus Nsp11 N-terminal/coronavirus NSP15 middle (AV-Nsp11N/CoV-Nsp15M) domain profile :<br><br>6514 - 6639: score = 42.039<br><br>NTPKPPVKKLLNNLGVSDIARIYVMDPKRDPAHEISTGVCSNEDIAKKPTEITICAPLTIV<br>FFRGRDQVQLFRNARWVLTIGSSVKGILQPSVGPKASLNGVTLTGAV----KTRQWY<br>YKVDGQVQ<br><br>Predicted features:<br>DOMAIN 6514 6639 AV-Nsp11N/CoV-Nsp15M [condition: none]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |  |  |  |  |
| <b>PS51958 NENDOU</b> Nidoviral uridylyate-specific endonuclease (NendoU) domain profile :<br><br>6656 - 6795: score = 70.476<br><br>KPRSGREIDFLEARDFTIRYKLEGVAFHIVYGGFSHQQLGGHLLTGLAKRCEKSP<br>LELDFRPSVYVMTITDAGTDSKVCNVDLLEDDVEITKSQLVAPSNVNTID<br>VYTESPLCKDGGVETYP<br><br>Predicted features:<br>DOMAIN 6656 6795 NendoU [condition: none]<br>ACT_SITE 6686 [condition: H] [group: 1]<br>ACT_SITE 6701 [condition: H] [group: 1]<br>ACT_SITE 6741 [condition: K] [group: 1]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |  |  |  |
| <b>PS51955 NIV_2_O_MTASE</b> Nidovirus 2'-O-methyltransferase (2'-O-MTase) domain profile :<br><br>6800 - 7094: score = 126.727<br><br>SQAKGQVAPVHLYVQWRLKELCDLQVYGDSATLPKGLP#HVAKYTLQCYLNTLAV<br>PYMIRVTHFGAGSGQVAPGATVLRQLPFGTLVDSGLNDVYSADSLTIGCATVhta<br>NKQDLITISQVDPKTNVikendskGFFTYCGTQKQLLGGSAVKTTEHSMADLY<br>KLHMFAMATVATVNWASSSEAFILGCHVLG--KPRGIDQVWNAVYFURNTMNPV<br>SSYSLFRSKFPLKLGATWASLKGGQNDKLLSLSGRLIIRHNHVRVSSQVNLV<br><br>Predicted features:<br>DOMAIN 6800 7094 Nidovirus-type SAM-dependent 2'-O-MTase [condition: none]<br>ACT_SITE 6844 [condition: K] [group: 1]<br>ACT_SITE 6928 [condition: D] [group: 1]<br>ACT_SITE 6968 [condition: K] [group: 1]<br>ACT_SITE 7001 [condition: E] [group: 1]<br><br>ruler:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |  |  |  |  |
| <b>PS00867 CPASE_2</b> Carbamoyl-phosphate synthase subdomain signature 2 :<br><br>2061 - 2068: [confidence level: (-1)] VLECNKVT<br><br>PS00213 LIPOCALIN Lipocalin signature :<br><br>4982 - 4993: [confidence level: (-1)] GTS...KFGYGRHPR                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |  |  |  |  |

**Figure 24 B** : Résultats d'analyse de la séquence protéique de polyprotéine réplicase 1ab de SARS-CoV-2 avec Prosite.

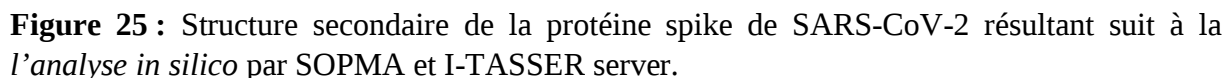
### 3.3. Analyses des protéines via I-TASSER server et l'outil SOPMA

Le serveur I-TASSER est une plate-forme en ligne appartenant à l'université de Michigan aux Etats-Unis et qui implémente les algorithmes basés sur I-TASSER pour les prédictions de la structure et de la fonction des protéines. Il permet aux utilisateurs universitaires de générer automatiquement des prédictions de modèle de haute qualité de la structure 3D et de la fonction biologique des molécules de protéines à partir de leurs séquences d'acides aminés.

Malheureusement, c'est à cause de grands nombres des requêtes reçu à chaque instant par cette plateforme que nous n'avons obtenu que la réponse à notre requête relative à la protéine spike, en attendant toujours les réponses aux autres requêtes qui peuvent durer de quelques heures, voire quelques semaines ou mois.

#### 3.3.1. Glycoprotéine de spicule (spike glycoprotein)

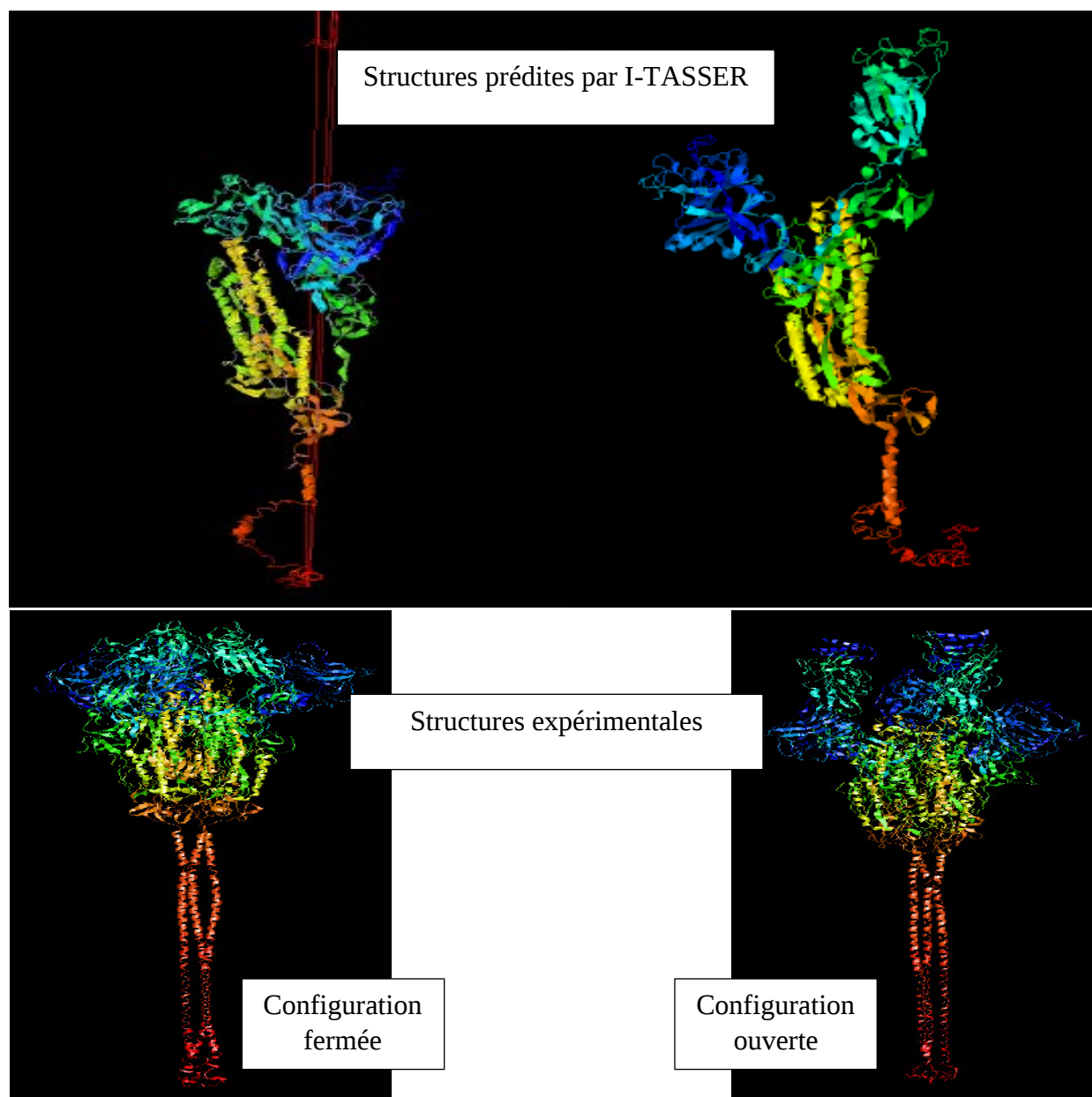
Dans les protéines, la structure secondaire (figure 25) est définie souvent par des arrangements de liaisons hydrogène entre les groupements amide et carbonyle du squelette peptidique (les liaisons hydrogène impliquant les chaînes latérales des acides aminés ne sont pas prises en considération dans la structure secondaire). Ces arrangements peuvent donner naissance à des structures en hélice, feuillet et coude qui représentent pour la protéine spike (28,59%, 23,25% et 3,38% respectivement, ainsi que le 44,78% restant est occupé par de pelote aléatoire ou Randon coil, d'après l'analyse par l'outil SOPMA). La prédiction de la structure secondaire de protéine peut permettre, dans certaines mesures, d'exprimer la structure tridimensionnelle des protéine la figure 26 montre la structure secondaire de la protéines spike résultant de l'analyse via I-TASSER server.



La structure tridimensionnelle des protéines est la structure fonctionnelle et qui permet d'expliquer les différentes interactions de la protéine avec son milieu.

Les structures tridimensionnelles de la glycoprotéine de spicule de SARS-CoV-2 prédite par I-TASSER server possédant les score TM les plus importants sont mentionnées dans la figure 26

Ces structures sont comparables avec celles qui sont établis expérimentalement dans la même figure en bas.



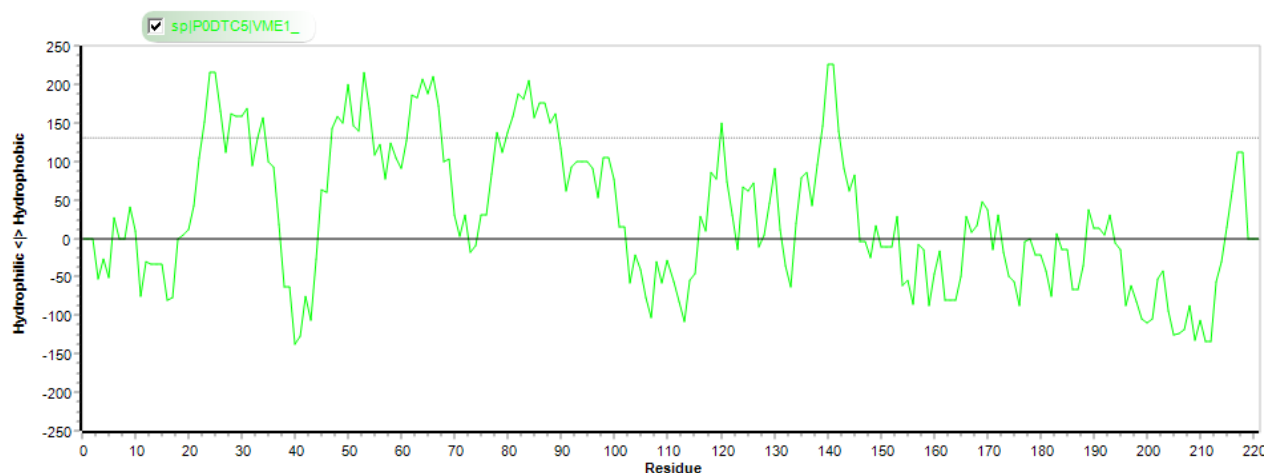
**Figure 26 :** Structures tridimensionnelles de la protéine de spicule de SARS-CoV-2. Les deux structures de haut sont les résultats de I-TASSER server ayant le score TM le plus élevé. Les deux structures d'en bas sont obtenus d'une manière in vitro. Les structures du haut sont comparables avec les structure du bas.

### 3.4. Prédiction des certaines caractéristiques physico-chimiques par le logiciel GPMW

#### 3.4.1. Protéine membranaire

Est une protéine de taille théorique moyenne de 25126,411 Da. Le pH isoélectrique théorique de la forme oxydée de cette protéine est 10,17, alors que sa forme réduite (c'est-à-dire dans le cas où tous les résidus cystéines sont porteurs des groupements thiols avec réduits avec « H » est 9,69. Elle est une protéine basique.

La figure suivante montre que cette protéine a un caractère hydrophobe plus important, ce qui lui permet d'accrocher à la membrane de l'enveloppe viral.

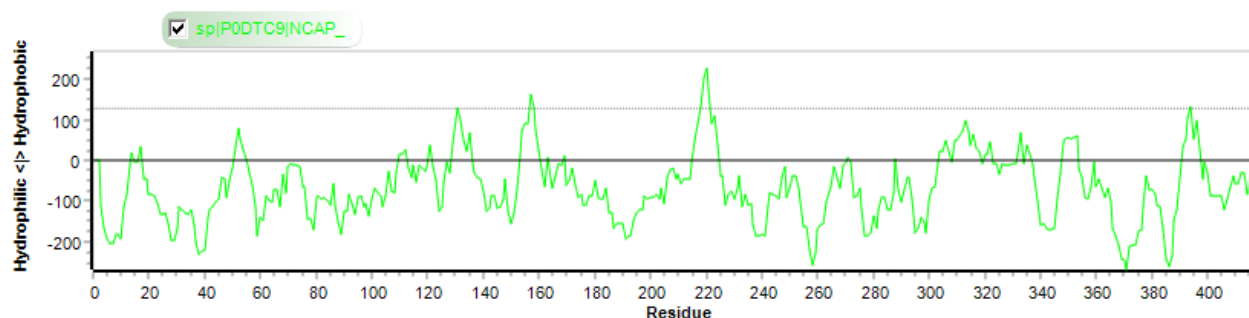


**Figure 27 :** Graphe d'hydrophobicité de protéine membranaire de virus SARS-CoV-2.

#### 3.4.2. Nucléoprotéine

Est une protéine de taille théorique moyenne de 45598,0187 Da. Le pH isoélectrique théorique de la forme oxydée de cette protéine est 10,48, il fait de même pour sa forme réduite. Elle est une protéine basique.

La figure suivante montre que cette protéine a un caractère hydrophile plus important, ce qui lui permet d'être en contact avec l'environnement aqueux de la cavité virale.

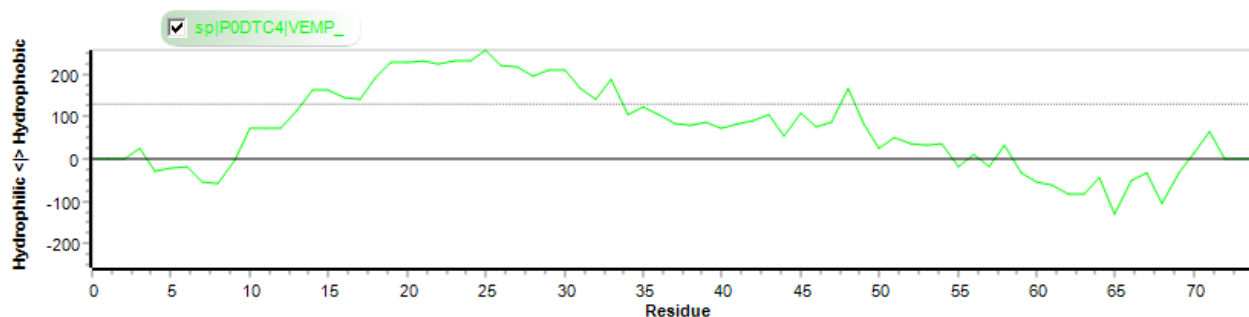


**Figure 28 :** Graphe d'hydrophobicité de nucléoprotéine de virus SARS-CoV-2.

### 3.4.3. Petite protéine d'enveloppe

Est une protéine de taille théorique moyenne de 8356,51 Da. Le pH isoélectrique théorique de la forme oxydée de cette protéine est 9,21, alors que pour sa forme réduite est 8,46. Elle est une protéine basique.

La figure suivante montre que cette protéine a un caractère hydrophobe plus important, ce qui lui permet d'accrocher à la membrane de l'enveloppe viral.

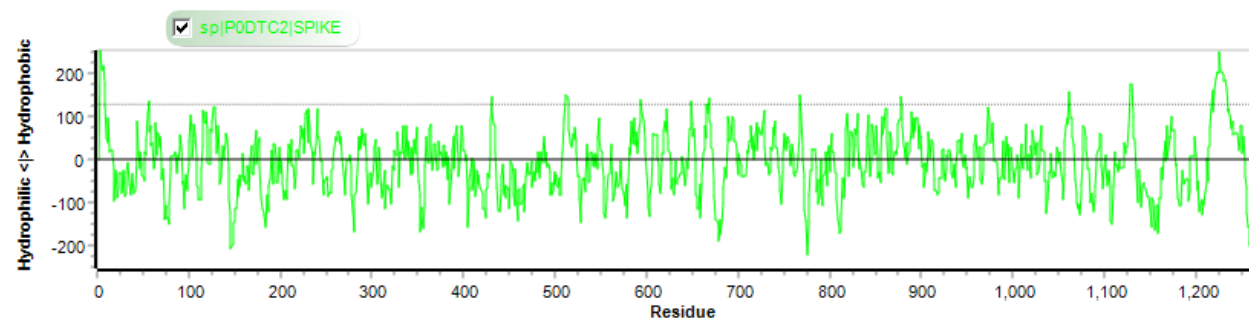


**Figure 29 :** Graphe d'hydrophobicité de petite protéine d'enveloppe de virus SARS-CoV-

### 3.4.4. Glycoprotéine de spicule

Est une protéine de taille théorique moyenne de 141048,09 Da. Le pH isoélectrique théorique de la forme oxydée de cette protéine est 6,24, alors que pour sa forme réduite est 6,21. Elle est une protéine acide.

La figure suivante montre que cette protéine a un caractère à la fois hydrophobe et hydrophile en partie, c'est une caractéristique qui lui permet de s'accrocher à la membrane de l'enveloppe viral d'une part et d'interagir via son fragment externe avec son ligand (ACE2) qui se trouve dans l'environnement externe aqueux.

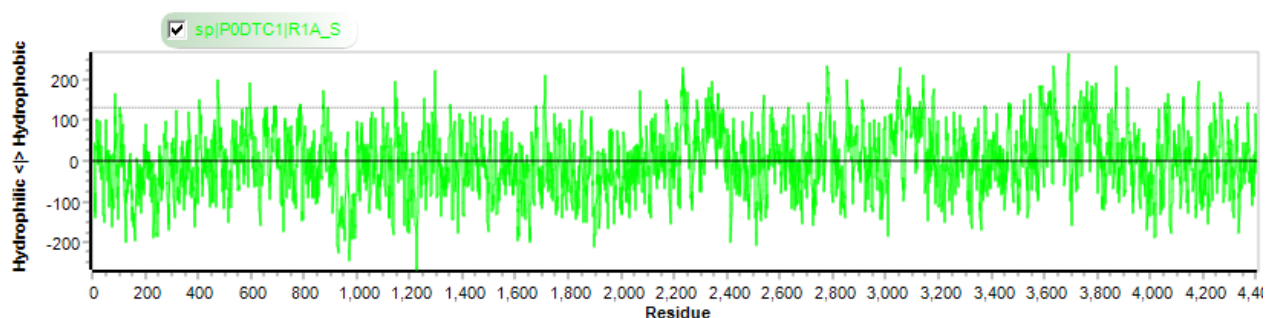


**Figure 30 :** Graphe d'hydrophobicité de glycoprotéine de spicule de virus SARS-CoV-2.

### 3.4.5. Polyprotéine réplacase 1a

Est une protéine de taille théorique moyenne de 489532,94 Da. Le pH isoélectrique théorique de la forme oxydée de cette protéine est 6,01, alors que pour sa forme réduite est 6. Elle est une protéine acide.

La figure suivante montre que cette protéine a un caractère polaire lui permettant de s'interagir avec son environnement.

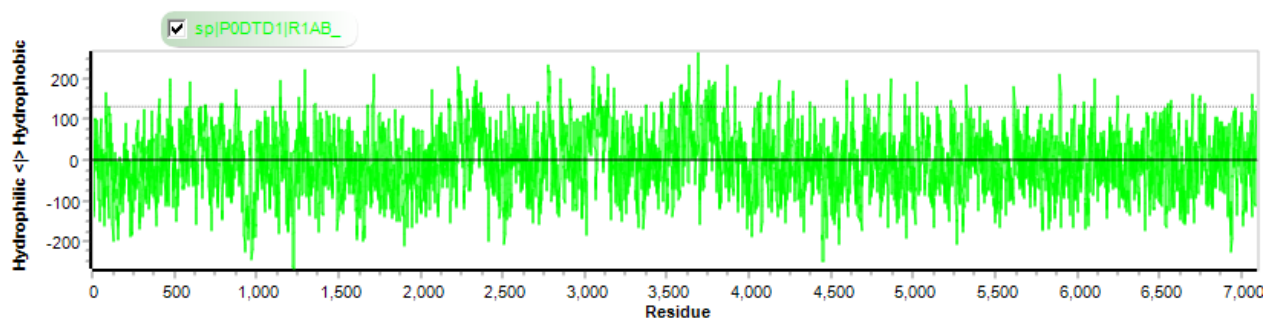


**Figure 31 :** Graphe d'hydrophobicité de polyprotéine réplacase 1a de virus SARS-CoV-2.

### 3.4.6. Polyprotéine réplacase 1ab

Est une protéine de taille théorique moyenne de 793316,28 Da. Le pH isoélectrique théorique de la forme oxydée de cette protéine est 6,33, alors que pour sa forme réduite est 6,30. Elle est une protéine acide

La figure suivante montre que cette protéine a un caractère polaire lui permettant de s'interagir avec son environnement.



**Figure 32 :** Graphe d'hydrophobicité de polyprotéine réplacase 1ab de virus SARS-CoV-

*Conclusion*

## Conclusion

Les protéines de SARS-CoV-2 jouent un rôle crucial dans l'infection par le SRAS-CoV-2, la virulence l'autoassemblage et la continuation de cycle de vie de ce virion. Pour comprendre les différents mécanisme différences entre les variants, déterminer et caractériser certaines régions protéiques conservatives essentielles au bon fonctionnement des protéines et la survie de ce virus, identifier les motifs et domaines protéiques fonctionnels et bien d'autres caractérisations, nous avons, alors, procédé à prélever les séquences des plus importantes protéines à partir de la banque des donnée Uniprot, les protéines choisies était au nombre de 6 protéines (protéine membranaire, nucléoprotéine, petite protéine d'enveloppe, protéine spike, polyprotéine réplicase 1a et polyprotéine réplicase 1ab). Ces protéines ont été affronté face aux plusieurs outils et programmes bio-informatiques (BLAST Protein, Clustal Omega, COBALT, Jalview, Prosite, SOPMA, I-TASSER server, GPMaw) afin de les caractériser. Ainsi, plusieurs conclusions ont été tirées :

- Pours au moins près de la moitié des protéines examinées via BLAST l'espèce structuralement la plus proche de SARS-CoV-2 est celle de *Rousettus bat coronavirus HKU9*, appartenant au sous genre de *Nobecovirus* ;
- Plusieurs régions conservées essentielles à la fonction de ces protéines ont été mise à l'évidence : la région (91)MWLSYFIA(98) pour la protéine membranaire, (114)GTG(116) pour la nucléoprotéine, (1) MYSFVSEETGTLIVNS(16) et (40)CAYCCNIVNVSLVKPSFYVYS(75) pour la petite protéine d'enveloppe, (860)VLP(862), (972)AIS(974), (1031)ECV(1033), (1042)FCG(1044), (1212)KWPW(1214) pour la protéine spike, (1118)VGP(1120) pour la polyprotéine réplicase 1a, alors qu'on a pas pu caractériser aucune région conservative pour la polyprotéine réplicase 1ab ;
- L'analyse via Prosite a permis de souligner plusieurs domaines et motifs fonctionnels ayant une grande importance pour ces protéines, les plus importants de ces domaines et motifs sont : les sites de N-myristoylation avec les glycines G(79) et G(126) possiblement pour l'encrage membranaire pour la protéine membranaire, Deux domaines de nucléocapside pour Coronavirus (N-terminal 48-175 et C-terminal 247-364) et 12 sites de N-myristoylation pour la nucléoprotéine, le domaine protéine de l'enveloppe de Coronavirus occupant la position de 1 à 75 pour cette même protéine, domaine de sous unité S1 N-terminal, domaine de 9-303 de sous unité S1 C-terminal de 334-527, domaine sous unité 2 heptade répétée 1 (HR1) entre 896-1001, sous unité 2 heptade répétée 2 (HR2) entre 1143-1225, 25 sites de N-myristoylation

utiles pour la fixation à membrane de l'enveloppe viral et le motif (403)RGD(405) nécessaire à la fixation avec la protéine de surface ACE 2 de pneumocyte pour la protéine spike, domaine Nsp1 de la position 12 à 127, domaine Nsp1 C-terminal entre 148 à 179, domaine Nsp2 N-terminal entre 183 à 456, domaine moyen Nsp2 entre 458 à 688, domaine Nsp2 C-terminal entre 690 à 818, domaine Nsp3a Ubl entre 821 à 929, macro domaine de la position 1025 à 1194, domaine Nsp3c-N entre 1231 à 1359, domaine Nsp3c-M entre 1367 à 1494, domaine Nsp3c-C entre 1496-1561, domaine Nsp3d Ubl de la position 1565 à 1620, un site actif qui composé de deux résidus C(1674) et H(1835) d'activité papaine like protéase, domaine de liaison à l'acide nucléique Nsp3e entre 1911-2021, Domaine Nsp3e de groupe de marqueur spécifique 2 entre 2046-2155, domaine Nsp3 Y3 entre 2660-2763, domaine Nsp4 C-terminal 3165-3263, site actif à activité protéase 3CL composé de deux résidus H(3304) et C(3408), domaine du cofacteur Nsp7 de l'ARN polymérase entre 3860-3942, domaine du cofacteur Nsp8 de l'ARN polymérase entre 3943 à 4140, domaine Nsp9 entre 4141 à 4253, domaine coactivateur ExoN/MTase entre 4254 à 4392 tous ces domaines et motifs caractérisent la polyprotéine réplique 1a alors que pour la polyprotéine réplique 1ab on a tous les domaines et motifs caractéristiques de la polyprotéines réplique 1a plus le domaine de la nucléotidyl transférase (NiRAN) associée au Nidovirus RdRp entre 4399-4653, domaine Nsp12 entre 4757-5324, domaine catalytique RdRp de ssRNA positif entre 5004-5166, domaine de liaison au zinc entre 5325 à 5408, domaine de l'hélicase centrale entre 5581-5932, domaine exoribonucléase 3'-5' (ExoN) entre 5997-6212, domaine guanine-N7-méthyltransférase (N7-MTase) entre 6221-6452, domaine Nsp15 entre 6453-6513, domaine (AV-Nsp11N/CoV-Nsp15M) entre 6514 à 6639, domaine (NendoU) entre 6656-6795, domaine (N7-MTase) entre 6221-6452, domaine Nsp15 d'oligomérisation N-terminal entre 6453 à 6513, domaine Nsp11 (AV-Nsp11N/CoV-Nsp15M) entre 6514-6639 domaine (NendoU) entre 6656-6795, domaine (2'-O-MTase) entre 6800-7094 ;

- D'autre part la structure secondaire obtenue montre la protéine est formée de 28,59% d'hélice alpha, 23,25% de feuillet beta et 3,38% de coude, par ailleurs la structure tridimensionnelle obtenue est comparable avec la même structure résultante de l'expérimentation ;
- Finalement, les caractéristiques physicochimiques des protéines tirées par GPMW peuvent être utilisées pour aider dans des essais expérimentaux ou la séparation et la purification des protéines de SARS-CoV-2.

# *Références bibliographiques*

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

Adedeji, a.o., et al., novel inhibitors of severe acute respiratory syndrome coronavirus entry that act by three distinct mechanisms. *Journal of virology*, 2013. 87(14): p. 8017-8028.

Adnan shereen, m., khan, s., kazmi, a., bashir, n., &siddique, r. Covid-19 infection: origin, transmission, and characteristics of human coronaviruses. *Journal of advanced research*. (2020). Doi:10.1016/j.jare.2020.03.005

Ahmadpoor p, rostaing l. Why the immune system fails to mount an adaptive immune response to a covid-19 infection. *Transpl int*. 25 avr 2020.

Ahmedou, abdelmoumine. "conception de nouveaux inhibiteurs de 3CLPRO-sars-cov-2 par approche in silico." (2021).

Ahmedou, abdelmoumine. "conception de nouveaux inhibiteurs de 3CLPRO-sars-cov-2 par approche in silico." (2021).

Alipoor, s. D., mortaz, e., jamaati, h., tabarsi, p., bayram, h., varahram, m., and adcock, i. M. Covid-19: molecular and cellular response. *Front cell infect microbiol*. (2021). 11: 563085.doi: 10.3389/fcimb.2021.563085

Alles, roxane. Developpement de nouveaux nanovecteurs pour les thérapies anti-hcv/hcc. Diss. Université de strasbourg, 2013.

Anzi, h. (2021). Pcr sars-cov-2: comparaison entre le kit mascir sars-cov-2 m 0.2 et le kit genefindertm covid-19 plus realamp kit.

Anzi, hasna. "pcr sars-cov-2: comparaison entre le kit mascir sars-cov-2 m 0.2 et le kit genefindertm covid-19 plus realamp kit." (2021).

Bachir, djaballah. "methode bio-informatique pour l'analyse de proteine." (2021).

Bartlam, mark, haitao yang, and zihe rao. "structural insights into sars coronavirus proteins." *current opinion in structural biology* 15.6 (2005): 664-672.

Bay bay, yassine. Caracteristiques des patients covid-19 admis en services de reanimation a l'hôpital ibn sina de rabat: etude descriptive a propos de 230 cas. Diss. 2021.

Bayle, benoit. Savoir pour guerir: soigner ses phobies ou ses toc par l'intention paradoxale. Editions ellipses, 2019.

Beaudoin C. A., Pandurangan A. P., Kim S. Y., Hamaia S. W., Huang C. L-H., Blundell T. L., Vedithi S. C. et Jackson A. P. (2022): In silico analysis of mutations near S1/S2 cleavage site in SARS-CoV-2 spike protein reveals increased propensity of glycosylation in Omicron strain. Med Virol. 2022;1–12. P: 1-12.

Bekkali, n., et al., eczematiform eruption after pfizer-biontech covid19 vaccine. Therapie, 2021. 76(4): p. 364-365.

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

Benmoussa, nadia. Risque du voyage et conseils medicaux aux voyageurs. Diss. 2014.

Benzakour, lamyae, frederic assal, and julie anne peron. "covid long neuropsychologique: origine neurologique ou psychiatrique?." revue medicale suisse 17.736 (2021): 822-826.

Birgand, gabriel, solen kerneis, and jean-christophe lucet. "modes de transmission du sars-cov-2: que sait-on actuellement?." m decine et maladies infectieuses

Bitar, dounia, and julien emmanuelli. "syndrome respiratoire aigu severe. L'epidemie de sras en 2003 en france." sras: www. Invs. Santeubliquefrance. Fr/publications/2004/sras\_021204/sras. Pdf (2005).

Bonnin, ariane. Caracterisation de la proteine s du coronavirus humain 229E. Diss. Universite de lille, 2018.

Bonny v, maillard a, mousseaux c, plaçais l, richier q. Covid-19 : physiopathologie d'une maladie a plusieurs visages. Rev med interne. Juin 2020;41(6):375-89.

Bonny, v., et al. "covid-19: physiopathologie d'une maladie a plusieurs visages." la revue de médecine interne 41.6 (2020): 375-389.

Bonny, v., et al. "covid-19: physiopathologie d'une maladie a plusieurs visages." la revue de medecine interne 41.6 (2020): 375-389

Bouguelli, malak, yousra guermat, and saida karouche. "etude epidemiologique sur le cov-19 dans la commune d'ain m'lila et recherche des activites et recherche des activites biologiques de l'artemisia herbabiologiques de l'artemisia herba alba pour des applications therapeutique." (2021).

Bouzidi, b. (2020). Enquete de la sero prevalence covid 19 chez les donneurs de sang dans la region marrakech-safi.

Brinis, houda, and asma abderrezak. Contribution a l'etude de coronavirus (sars-cov2) a m'sila. Diss. Universite mohamed boudiaf-m'sila, 2021..

Brugere-picoux, j., et al. "rapport 21-11. Covid-19 et monde animal, d'une origine encore mysterieuse vers un futur toujours incertain." bulletin de l'academie nationale de medecine 205.8 (2021): 879-890.

Brugere-picoux, j., et al. "rapport 21-11. Covid-19 et monde animal, d'une origine encore mysterieuse vers un futur toujours incertain." bulletin de l'academie nationale de medecine 205.8 (2021): 879-890.

Catania, l. J. Sars-cov-2 and the covid-19 pandemic. Foundations of artificial intelligence in healthcare and bioscience. (2021). Chapter 8, sars-cov-2 and the covid19 pandemicp 445–467 doi:10.1016/b978-0-12-824477-7.00004-3

Chaib, rachid. "les risques lies aux agresseurs biologiques." journal article 14 (2020).

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

Chau, c.h., j.d. Strobe, and w.d. Figg, covid-19 clinical diagnostics and testing technology. Pharmacotherapy, 2020. 40(8): p. 857-868.

Chehboub, faten, and rofaida zitouni. "analyse viro-informatique des variants sars-cov-2 portee sur la proteine spike."2021

Corbett, k.s., et al., evaluation of the mrna-1273 vaccine against sarscov-2 in nonhuman primates. The new england journal of medicine, 2020. 383(16): p. 1544-1555

Cornec, jérémy. Imaginaires de la dystopie et du posthumain dans les series d'anticipation science-fictionnelles contemporaines (2009-2019). Diss. Brest, 2021.

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

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

Coulibaly, f., et al. "structures cristallographiques des birnavirus: implications pour l'evolution des virus a arn double brin."

Covid-19 science report: pathogenesis and host immune response(phr) to sars-cov-

Cui j., li f., shi z. L. 2019. Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. 17 (3) : 181–192

De haan c. A., rottier p. J. 2005. Molecular interactions in the assembly of coronaviruses. Adv. Virus res. 64 : 165–230

Delaplace, manon. Role des cyclophilines dans la replication des coronavirus. Diss. Paris, agroparistech, 2021.

Dhama, k., et al., coronavirus disease 2019-covid-19. Clinical microbiology reviews, 2020. 33(4): p. E00028-20.

Doroftei, b., et al., mini-review discussing the reliability and efficiency of covid-19 vaccines. Diagnostics (basel, switzerland), 2021. 11(4): p. 579.

Doucet, alain. "relation entre la structure et la fonction de la preelafine et son implication au niveau pulmonaire." (2006).

De wilde ah, snijder ej, kikkert m, van hemert mj. Host factors in coronavirus replication. Roles of host gene and non-coding rna expression in virus infection. 23 juin 2017; 419:1-42

Drouaz, saadia, and selma kadri. Etude descriptive de la situation actuelle de covid-19 a m'sila. Diss. Universite mohamed boudiaf-m'sila, 2020.

Dussol, alexis. "covid-19: la vengeance du pangolin: les 180 jours qui ebranlerent la planete." covid-19 (2020): 1-304.

El moussaoui, kamal. "covid 19 et grossesse." (2021).

El ouazzani, kawtar. "les pathologies hivernale infectieuses et conseil a l'officine." (2021).

Elharouchi, sofia. Les mesures de lutte contre le sars-cov-2. Experience du service de sterilisation et d'hygiene du cvmit de l'hmimv durant l'annee 2020. Diss. 2022.

Elharouchi, sofia. Les mesures de lutte contre le sars-cov-2. Experience du service de sterilisation et d'hygiene du cvmit De l'hmimv durant l'annee 2020. Diss. 2022.

Eschenbrenner, etienne. Roles des tetraspanines tspan5 et tspan15 dans le controle de l'endocytose et du niveau d'expression de la metalloprotease adam10. Diss. Université paris-saclay (comue), 2019.

Es-saad, salwa. "role de la cttnb1 et de ses nouveaux partenaires dans la regulation de l'immunité antivirale innée et de la voie wnt." (2019).

Essaadouni, l., and k. Krati. "physiopathologie de la maladie covid-19." rmmad (2020): 3

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

Fellah, sofia. Revue de la litterature sur l'immunogenicite, l'innocuite et l'efficacite des differents vaccins contre la covid-19. Diss. 2022. Formation (2022).

From bio wiki. Middle east respiratory syndrome–related coronavirus 2012.last edited on 12 august 2021, at 20:57 (utc). Disponiblesur : <http://rule of 6IXFIELD of science.com./2012/09/a-newcoronavirus-should-youcare.htm>

Gabant, guillaume. Spectrometrie de masse des modifications induites ou post-traductionnelles de proteines: methodologie et application a des proteines d'interet therapeutique. Diss. Universite d'orleans, 2014.

Gala, j.-l., et al., methodes diagnostiques du covid-19. Louvain med 2020 mai-juin; 139 (05-06) : 228-23, 2020.

Galmes, johanna. Isolement et caracterisation de nouvelles especes de torque teno mini virus (ttmv): implication potentielle dans la pathogenese de la pneumonie. Diss. Lyon, école normale superieure, 2013.

Garbaye, jean. "la symbiose mycorhizienne: une association entre les plantes et les champignons." la symbiose mycorhizienne (2013): 1-280.

Graham rl, sparks js, eckerle ld, sims ac, denison mr. Sars coronavirus replicase proteins in pathogenesis. Virus res 2008;133:88–100

Gralinski l. E., menachery v. D. 2020. Return of the coronavirus: 2019-ncov. Viruses. 12 (2) : 135

Hadjeris, roumaissa, manel boudiaf, and ichraf haouas. "tests de depistage et protocole therapeutique sars-cov-2 (covid-19)." (2021).

Hamidi, redha malek, et al. "covid-19 en reanimation: quelle prise en charge?." revue algerienne d'allergologie. Vol 5.01 (2020): 2543-3555.

Haute autorite de sante (has) ,2020• aspects immunologiques et virologiques de l'infection par le sars-cov-2 •– novembre 2020 – isbn :978-2-11-155666-9.

Helmy, y. A., mohamed, f., elaswad, a., sobieh, a., kenney, s. P., and shehata, a. A. The covid-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. Journal of clinical medicine .2020. 9,1225; doi:10.3390/jcm9041225

Henry bm, oliveira mhs, benoit s, plebani m, lippi g. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (covid-19): a meta-analysis. Clin chem lab med. 10 avr 2020.

Huang y, yang c, xu x, xu w, liu s. Structural and functional properties of sarscov-2 spike protein: potential antivirus drug development for covid-19. *Acta pharmacologica sinica*. Sept 2020;41(9):1141-9.

Hui D.S., Azhar E.I., Memish Z.A. and Zumla A. (2022): Human Coronavirus InfectionsdSevere Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS), and SARS-CoV-2. *Encyclopedia of Respiratory Medicine*, 2nd edition, Volume 4. P 146-161.

Jamai amir, i., et al., covid-19 : virologie, epidemiologie et diagnostic biologique. *Option/bio*, 2020. 31(619) : p. 15-20.

Jayaram, h., fan, h., bowman, b. R., ooi, a., jayaram, j., collisson, e. W., ... prasad, b. V. V. X-ray structures of the n- and c-terminal domains of a coronavirus nucleocapsidprotein: implications for nucleocapsid formation. *Journal of virology*. (2006).80(13), 6612–6620. Doi:10.1128/jvi.00157-06

Josse, evelyne. "le coronavirus pour les nuls." (2020).

Juckel, dylan, jean dubuisson, and sandrine belouzard. "les coronavirus, ennemis incertains." *medecine/sciences* 36.6-7 (2020): 633-641.

Juckel, dylan, jean dubuisson, and sandrine belouzard. "les coronavirus, ennemis incertains." *medecine/sciences* 36.6-7 (2020): 633-641.

Kabore, mikaila. "confinement des passagers aeroportes en situation de pandemie de covid-19 au bf: epidemiologie et efficacite d'intervention." (2021).

Kampf, g., et al., persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. *The journal of hospital infection*, 2020. 104(3): p. 246-251

Khammoum, r., touati, s., & fetni, s. (2021). Enquete ethnobotanique sur l'infection au covid19.

Kadkhoda k. Covid-19: an immunopathological view. *Msphere*. 22 avr 2020;5(2):e00344-20.2.2020.

Kin, nathalie, and astrid vabret. "les infections a coronavirus humains." *revue francophone des laboratoires* 2016.487 (2016): 25-33.

Klaus, julia, et al. "detection and genome sequencing of sars-cov-2 in a domestic cat with respiratory signs in switzerland." *viruses* 13.3 (2021): 496.

Knoll, m.d. And c. Wonodi, oxford-astrazeneca covid-19 vaccine efficacy. *Lancet* (london, england), 2021. 397(10269): p. 72-74.

Kohnert, dirk. "sur l'impact socio-economique des pandemies en afrique: leçons tirees du covid-19, de la trypanosomiase, du vih, de la fievre jaune, du cholera." (2021): 1-46.

Lahnikate, sarah. "coronavirus émergents: état des lieux et actualités." (2021).

Laude h. D. B., eleouet j. F. 1998 le coronavirus porcin prcv: un virus émergent pas comme les autres. *Virologie*. 2 : 305–16

Le coupanec, alain. Conséquences neurologiques des clivages de la glycoprotéine s du coronavirus neuroinvasif humain oc43. Diss. Université du québec, institut national de la recherche scientifique, 2018.

Li f. 2016. Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* 3(1), 237–261

Li h, liu l, zhang d, xu j, dai h, tang n, et al. Sars-cov-2 and viral sepsis: observations and hypotheses. *The lancet*. Avr 2020.

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

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) : 367–7

Liu j, li s, liu j, liang b, wang x, wang h, et al. Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of sars-cov-2 infected patients. *Ebiomedicine*. Mai 2020;55:102763.

Liu, j., et al., hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting sars-cov-2 infection in vitro. *Cell discovery*, 2020. 6: p. 16-16

Lorin, valerie. Identification, caracterisation moleculaire et etude des proprietes fonctionnelles des anticorps neutralisants a large spectre contre le vih-1. Diss. Universite de paris, 2020.

Loubna, ramdani. "etude en population sur l'evolution des determinants du choix alimentaires des eleves du secondaire: etude de l'experience des adolescents de guelma."

Machado, b.a.s., et al., the main molecular and serological methods for diagnosing covid-19: an overview based on the literature. *Viruses*, 2020. 13(1): p. 40

Marais, claire-line. Role de la snare memb11 comme «recepteur» de la gtpase arf1 a l'appareil de golgi chez arabidopsis thaliana. Diss. Bordeaux 2, 2013.

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

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

Mertens, pascal, et al. "development and potential usefulness of the covid-19 ag respi-strip diagnostic assay in a pandemic context." *frontiers in medicine* 7 (2020): 225.

Meryam, djamaa; hawa, boulekroune. Les differentes strategies therapeutiques impliquees dans le controle et le traitement de l'epidemie covid-19. 2021. Phd thesis. University center of abdalhafid boussouf-mila.

Migne, camille victoire, et al. "quels outils pour la surveillance des arbovirus émergents au sein de leurs hotes mammiferes et vecteurs arthropodes?." *virologie* 25.1 (2021): 12-28

Mihoubi, manel, and assia madad. Utilisation de la chloroquine et l'hydroxychloroquine dans le traitement du covid-19. Diss. Universite mohamed boudiaf-m'sila, 2021.

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

Mondiale de la sante, organisation. Surveillance de la sante publique dans le contexte de la covid-19: orientations provisoires, 14 fevrier 2022. No. Who/2019-ncov/surveillancguidance/2022.1. Organisation mondiale de la sante, 2022.

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

Mulangu, s., et al., a randomized, controlled trial of ebola virus disease therapeutics. 2019. 381(24): p. 2293-2303

Muller, m., i. Bulubas, and t. Vogel. "les facteurs pronostiques dans la covid-19." npg neurologie-psychiatrie-geriatrie 21.125 (2021): 304-312.

Nafti, kenza, and hala zerrouk. "etude epidemiologique et evaluation des facteurs des risques du covid 19 dans la wilaya d'el-oued." (2021).

Narayanan, krishna, cheng huang, and shinji makino. "sars coronavirus accessory proteins." virus research 133.1 (2008): 113-121.

Narayanan, krishna, cheng huang, and shinji makino. "sars coronavirus accessory proteins." virus research 133.1 (2008): 113-121.

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

Oms. Les differents types de vaccins contre la covid-19. 2021; available from: <https://www.who.int/fr/news-room/feature-stories/detail/the-race-for-acovid-19-vaccine-explained> (consulté le 11/09/2021).

Oms. Prise en charge clinique des infections respiratoires aiguës severes en cas de suspicion d'infection par un nouveau coronavirus (ncov). 2020; available from: <https://www.who.int/publications/i/item/10665-332299> ( consulté 05/11/2021)

Parodi, andre laurent. "animal et societe." sante (2021).

Parrochia, daniel. "sur les origines possibles de l'epidemie covid-19." (2021).

Personnic S., Duhamel S., Bettarel Y., Sime-Ngando T. et Jacquet S. (2006) : les virus planctoniques : un compartiment biologique clé des milieux aquatiques. *Courrier de l'environnement de l'INRA* n°53, décembre 2006. P: 19-34

Poisson-benatouil, catherine, and jean-luc galabert. "évaluer le potentiel de l'artemisia annua dans le cadre de la lutte contre le covid-19."

Prabhakar, h., kapoor, i., & mahajan, c. (eds.). *Clinical synopsis of covid-19*. chap1 , introduction: history of coronavirus disease pandemic. (2020). doi:10.1007/978-981-15- 8681-1

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

Rabaan, ali a., et al. "sars-cov-2, sars-cov, and mers-cov: a comparative overview." *infez med* 28.2 (2020): 174-184.

Rechidi, nidal, et al. "l'integration pedagogique des tic a l'epreuve de la crise covid-19: quels enseignements a tirer?." *revue internationale du chercheur* 1.2 (2020).

Redha, aidel mohamed. *Etude epidemiologique de l'efficacite des differents traitements utilises contre le coronavirus (covid-19)*. Diss. 2020.

Rim, c. H. E. R. K. A. O. U. I. "place du test panbio covid-19AG rapid dans le diagnostic du sarscov 2 (etude retrospective au laboratoire de virologie de l'hmimv)." (2021).

Rogers, r., et al., the covid-19 diagnostic dilemma: a clinician's perspective. *Journal of clinical microbiology*, 2020. 58(8): p. E01287-20

Rogers, ralph, et al. "the covid-19 diagnostic dilemma: a clinician's perspective." *journal of clinical microbiology* 58.8 (2020): e01287-20.

Rossignol, jean-michel, cecile lagaudriere-gesbert, and delphine sitterlin. "1. Le virus de l'hepatite b." *hepatite b*. Edp sciences, 2021. 1-26.

Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2010 Apr;5(4):725-38. doi: 10.1038/nprot.2010.5. Epub 2010 Mar 25. PMID: 20360767; PMCID: PMC2849174.

Samake, d., et al. "la covid-19 a mopti: aspects epidemiologique, clinique, therapeutique et evolutif." (2021).

Sami, zemmouchi ramzi, and f. E. N. D. I. Nahla. "intitulé."

Sarkar, m., &saha, s. Structural insight into the role of novel sars-cov-2 e protein: a potential target for vaccine development and other therapeutic strategies. Plos one.(2020). 15(8), e0237300. Doi:10.1371/journal.pone.0237300.

Satarker, s., nampoothiri, m. Structural proteins in severe acute respiratory syndrome coronavirus-2.archives of medical research. (2020). Doi:10.1016/j.arcmed.2020.05.012

Savarino, a., et al., new insights into the antiviral effects of chloroquine. The lancet. Infectious diseases, 2006. 6(2): p. 67-69

Segondy, michel. "les coronavirus humains." revue francophone des laboratoires 2020.526 (2020): 32-39

Sigrist C.J.A., Cerutti L., Hulo N., Gattiker A., Falquet L., Pagni M., Bairoch A., Bucher P. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform. 3:265-274(2002).

Singh, v., luthra, a., chauhan, r., and meena, s. C. Clinical synopsis of covid-19. (2020). Chap 2. Basic virology and pathophysiology of covid-19.doi:10.1007/978-981-15-8681-1 p 5-30

Snijder ej, decroly e, ziebuhr j. The nonstructural proteins directing coronavirus rna synthesis and processing. Adv virus res. 2016; 96:59-126.

Stitra, nassiha, et al. La protease alcaline: interets biotechnologiques. Diss. Universite de jijel, 2008.

Tahar, examineur mr nouadri, and encadreur mme benlounissiaïcha. "memoire."

Tahar, examineur mr nouadri, and encadreur mme benlounissiaïcha. "memoire."2021

Tay mz, poh cm, rénia l, macary pa, ng lfp. The trinity of covid-19: immunity, inflammation and intervention. *Nat rev immunol.* 28 avr 2020.

Togola, mahamadou. Spectre des virus des voies respiratoires dans l'aire sanitaire de l'hôpital du mali et du cscom de yirimadio. Diss. Usttb, 2021.

Traore, abdouramane. "reactivite de la proteine spike du sars-cov-2 et de ses domaines de liaison rbd et p3 aux serums des patients covid-19 et avant covid-19 au mali." (2021).

Tsai ph, wang ml, yang dm, liang kh, chou sj, chiou sh, lin th, wang ct, chang tj. Variance genomique des cadres de lecture ouverts (orf) et de la proteine spike dans le coronavirus 2 du syndrome respiratoire aigu severe (sras-cov-2). *J chin med assoc.* 2020 août;83(8):725-732. Doi : 10.1097/jcma.0000000000000387. Pmid : 32773643 ; pmcid : pmc7493783.

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

Vardhana sa, wolchok jd. The many faces of the anti-covid immune response. *J exp med.* 1 juin 2020; 217(6):e20200678.

Velazquez-salinas l, verdugo-rodriguez a, rodriguez ll, borca mv. The role of interleukin 6 during viral infections. *Front microbiol.* 10 mai 2019;10:1057.

Wahida, boulakhiout, and mohammed nadhir elaggoune marwa sekfali. "place de la phytothérapie dans le traitement contre le covid-19." (2021).

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

Wang, l., et al., review of the 2019 novel coronavirus (sars-cov-2) based on current evidence. *International journal of antimicrobial agents*, 2020. 55(6): p. 105948-105948

Wang, ning, et al. "subunit vaccines against emerging pathogenic human coronaviruses." *frontiers in microbiology* 11 (2020): 298.

Wang, y., et al., unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (covid-19) implicate special control measures. *Journal of medical virology*, 2020. 92(6): p. 568-576

Wheeler D, Bhagwat M. BLAST QuickStart: Example-Driven Web-Based BLAST Tutorial. In: Bergman NH, editor. *Comparative Genomics: Volumes 1 and 2*. Totowa (NJ): Humana Press; 2007.

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

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

Yaqinuddin, a., et al., effect of sars-cov-2 mutations on the efficacy of antibody therapy and response to vaccines. *Vaccines*, 2021. 9(8): p. 914

Yaqinuddin, ahmed, et al. "effect of sars-cov-2 mutations on the efficacy of antibody therapy and response to vaccines." *vaccines* 9.8 (2021): 914.

Zeng, w., liu, g., ma, h., zhao, d., yang, y., liu, m., ... jin, t. Biochemical characterization of sars-cov-2 nucleocapsid protein. *Biochemical and biophysical research communications*. (2020). Doi:10.1016/j.bbrc.2020.04.136

Zheng m, gao y, wang g, song g, liu s, sun d, et al. Functional exhaustion of antiviral lymphocytes in covid-19 patients. *Cell mol immunol*. Mai 2020;17(5):533-5.

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*.

Zhou, p., et al., a pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 2020. 579(7798): p. 270-273.

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

Zi-wei ye et shuofeng yuan, « zoonotic origins of human coronaviruses » [archive], sur international journal of biological sciences, 2020 (issn 1449-2288, pmid 32226286, pmcid pmc7098031, doi 10.7150/ijbs.45472, consulte le 22 mai 2020), p. 1686–1697

*Annexes*

## **Annexe**

### **Code international des acides aminés**

- G - Glycine (Gly)
- P - Proline (Pro)
- A - Alanine (Ala)
- V - Valine (Val)
- L - Leucine (Leu)
- I - Isoleucine (Ile)
- M - Methionine (Met)
- C - Cysteine (Cys)
- F - Phenylalanine (Phe)
- Y - Tyrosine (Tyr)
- W - Tryptophan (Trp)
- H - Histidine (His)
- K - Lysine (Lys)
- R - Arginine (Arg)
- Q - Glutamine (Gln)
- N - Asparagine (Asn)
- E - Glutamic Acid (Glu)
- D - Aspartic Acid (Asp)
- S - Serine (Ser)
- T - Threonine (Thr)