Immune Epitope Map of the Reported Protein Sequences of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)

CURRENT STATUS: POSTED

Leonardo A. Guevarra Jr.
University of Santo Tomas
✉ laguevarra@ust.edu.ph
Corresponding Author

Gianne Eduard L. Ulanday
National Institutes of Health - University of the Philippines Manila
ORCID: https://orcid.org/0000-0002-1093-9287

DOI: 10.21203/rs.3.rs-18689/v1

SUBJECT AREAS
Virology  Bioinformatics

KEYWORDS
coronavirus, SARS-CoV-2, novel coronavirus, 2019-nCoV, epitope map, COVID-19, severe acute respiratory syndrome coronavirus 2
Abstract
Identifying immunogenic sequences of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) proteins is important in developing epitope-based vaccine and diagnostics. This step is critical in designing potent vaccines and highly specific diagnostic tools which can help prevent the spread of this disease.

In this study, we identified, using in silico analysis tools, immunogenic epitopes of the reported sequences of SARS-CoV-2 proteins and determined similar sequences with known viral proteins. The amino acid sequences of the SARS-CoV-2 proteins were acquired from the National Center for Biotechnology Information (NCBI) database. B-cell epitope prediction was done using in silico analysis tools available at the Immune Epitope Database and Analysis Resources (IEDB). Blastp was performed on the identified immunogenic sequences to determine similarities with known viral proteins and deduce possible locations in the coronavirus.

We were able to identify B-cell epitopes of the SARS-CoV-2 polyprotein, surface glycoprotein, envelop, membrane glycoprotein, nucleocapsid phosphoprotein, orf3, orf7a and orf8. No epitope was identified in orf6 and orf10. High similarities of the predicted immunogenic epitopes of the SARS-CoV-2 were observed with the 2003 SARS-CoV. However, unique epitopes were identified in non-structural proteins (NSP) 1 and 3 and surface glycoprotein of the SARS-CoV-2.

Introduction
Coronaviruses are single-stranded, positive-sense RNA viruses which are classified into four genera; namely, alpha, beta, delta, and gamma coronaviruses. The former two genera primarily infect mammals, whereas the latter two primarily infect birds [1,2]. Its genome is the largest among the RNA viruses and includes a variable number (around 6 to 11) of open reading frames (orf). Coronavirus replication is somewhat unique wherein; it involves ribosomal frameshifting or slippage and having a large replicase gene with an open reading frame (orf1ab) . The replicase gene occupies around two-thirds of its genome and encodes the 16 nonstructural proteins (NSPs). The remaining one-third of the genome (~10kb) encodes for the structural and accessory proteins [1,3]. The main structural proteins include the viral envelope-bound membrane protein (M), envelope protein (E) and
spike protein (S) and the RNA-bound nucleocapsid (N) [3,4]. A fifth structural protein, the hemagglutinin esterase (HE), may be present but only among betacoronaviruses [5]. Aside from the structural proteins, its gene encodes 16 non-structural proteins which are responsible either in viral gene replication, protein scaffold formation, proteolytic maturation of proteins, and protection from host’s immune response [6].

Until recently, there were six coronaviruses (CoVs) known to infect humans; HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, SARS-CoV and MERS-CoV which evolved between 1960 and 2015 [7]. By the end of 2019, however, a new coronavirus was detected in China among individuals suffering from acute respiratory distress [8]. From the initial cases identified to have links with the Huanan seafood and wildlife market in Wuhan City at the Hubei Province in Central China, this zoonotic emerging infection, has now reached 25 countries in Asia, North America, Europe, and Australia [9,10]. The exact source of exposure leading to this event is still under investigation.

Researchers worldwide rushed to sequence the viral genome to aid state authorities in building their diagnostic and rapid containment capabilities. This emerging threat has caused an unprecedented alarm among states and was immediately recognized by the World Health Organization (WHO) as a Public Health Emergency of International Concern [9,11]. As of 15 March 2020, the global confirmed cases of coronavirus disease 2019 (COVID-19) has already reached more than 153 thousand cases and has claimed 5,375 lives [13].

Coronaviruses have been notoriously implicated in recent high-profile, cross-border outbreaks affecting human populations. Phylogenetic studies of these viral family suggest a high capacity for transmission across species barriers having been found in bats, pigs, camels, and humans. The increasing frequency of its genetic recombination coupled with profound human-animal interface activities leads to higher probabilities of zoonotic spillover events [13–15]. The emergence of novel pathogens, such as the SARS-CoV–2, poses a serious threat to human health of up to global proportions because of the knowledge gaps on the pathogen causing the disease and the lack of pre-formed immunity among individuals [16]. This knowledge gap, particularly on the molecular characteristics of SARS-CoV–2, is a barrier in creating strategies in controlling the spread of the
infection including the development of rapid diagnostic devices and designing of vaccines [17]. Fortunately, bioinformatics tools such as epitope analysis resources and sequence identity analysis tools can be exploited in identifying and mapping immunogenic sequences and their possible locations in viral polyproteins [18,19].

In an effort to contribute to the existing knowledge gap on the identity and genomic characteristics of the SARS-CoV-2, we aimed to identify, using in silico prediction tools, B-cell epitopes of the of the SARS-CoV-2 which can serve as basis for future recombinant engineering work and vaccine design studies. We also aim to determine similarities in the identity of the in silico-predicted epitopes with other viral proteins found in public databases, especially those which are closely related to SARS-CoV-2. Focus has been established on SARS-related coronaviruses (SARS-CoV) and other significant members of betacoronavirus as these were the apparent nearest relative of SARS-CoV-2 based on current phylogenetic data.

Results

We were able to identify, using in silico epitope prediction, tools available in the Immune Epitope Database and Analysis Resources (IEDB), potentially immunogenic epitopes of the reported amino acid sequences of SARS-CoV-2 polyprotein, surface glycoprotein, orf3, envelop protein, membrane glycoprotein, orf7a, orf8, and nucleocapsid phosphoprotein. For the polypeptide sequence of orf6 and orf10, none was found to be potentially immunogenic, and all values are lower than the cut-off.

Supplementary Table 1 and Supplementary Table 2 presents the position, sequences, antigenicity, surface accessibility, and hydrophilicity scores of the predicted epitopes.

The 10-mer peptide sequences with the highest antigenicity scores are both found in the envelop protein. These sequences are located at positions 50—59 (SLVKPSFYVY) and 51—60 (LVKPSFYVYS) which can actually form a single immunogenic epitope of the envelop proteins of SARS-CoV-2. The sequence which has the highest surface accessibility score is located at position 382—391 (LPQRQKKQQT) of the nucleocapsid phosphoprotein while the sequence, which is predicted to be most hydrophilic, is located at position 237—246 (KGQQQGQT) also of the nucleocapsid phosphoprotein. Combining continuous adjacent sequences of the predicted 10-mer epitopes generated 111 epitopes
for the polyprotein, 22 for the surface glycoprotein, three for orf3, a single 11-mer epitope for the envelop protein, five for membrane glycoprotein, four for orf7a, five for orf8, and six for the nucleocapsid phosphoprotein. These sequences are presented in Table 1 and Table 2.

A high homology was observed in the predicted immunogenic epitopes of SARS-CoV-2 with the proteins of the SARS-related coronavirus (SARS-CoV). The epitopes of SARS-CoV-2 polyproteins have homologous sequences with the non-structural proteins (NSP) 1, 3 (replicase and proteinase domains), 7 (replicase light chain), NSP 8 (replicase heavy chain), NSP 9 (replicase), 10, 12, 13 (helicase), 14 (guanine-n7 methyltransferase) and 15. This was also observed for the epitopes predicted for the reported sequences of SARS-CoV-2 glycoprotein, envelop, and orf7a which were found to have homologous sequences in the SARS-CoV spike glycoprotein, small envelop protein, and orf7a accessory protein, respectively. Unique epitopes were also found at positions 488—497 (VETVKGLDYK), 555—564 (AQNSVRVLQKA), 713—725 (SKGLYRKCVKSRE), 1006—1015 (VEVQPQLEME), 1045—1054 (IVEEAKKVKP), 1048—1057 (EAKKVKTTPV), 1227—1236 (QDDKKIKACV), 2041—2051 (CEDLKVSEEV), 2045—2056 (KPVSEEVVENPT), 2551—2564 (ESSAKSASVYYSQL) and 2655—2665 (LKLSHQSDIEV) of the SARS-CoV-2 polyprotein and positions 44—53 (RSSVLHSTQD), 319—328 (RVQPTESIVR) and 321—330 (QPTESIVRF) of the SARS-CoV-2 surface glycoprotein.

Discussion
In the last two decades prior to the current SARS-CoV-2 outbreak, two coronaviruses gained prominence due to its novelty, infectivity, and virulence - the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) in 2002—2003 and the Middle East Respiratory Syndrome coronavirus (MERS-CoV) in 2012 (Son et al., 2017). The lessons learned in both epidemics are being applied by scientists around the world in the current SARS-CoV-2 outbreak as evidenced by increased data transparency and broader information sharing among stakeholders [20].

The availability of current technologies has also paved the way for a quicker response to human diseases. Molecular biology-based technologies, including advancement of sequencing methods, helped in the characterization of pathogens. Whole genome sequences can be done with remarkable speed, accuracy, and depth of information [21]. In addition, bioinformatics tools and global genomic
and proteomic databases haver aided scientists worldwide in understanding molecular structures and characteristics, hence developing strategies to control human diseases [22].

The application of computational methods in immunology, such as in silico epitope prediction, enabled researchers to focus on and prioritize immune targets for experimental epitope mapping, saving time and resources, which are crucial in providing expedient epidemic containment and response [23–25]. In silico epitope mapping helped researchers to expeditiously identify epitopes essential in rational vaccine design and development of epitope-based diagnostic serological devices [26–27]. In this paper, we present putative epitopes of SARS-CoV-2 proteins, including sequence similarities with other viral proteins, which may potentially be used in the development of epitope-based vaccine against this recent emerging infection. One of the findings presented in this paper that may have impact in the disease control strategies is the high homology between the immune epitopes of SARS-CoV-2 and the 2003 SARS-CoV which also originated in China. We were able to identify high sequence homology between SARS-CoV-2 NSP1, NSP3, NSP7, NSP8, NSP9, NSP10, NSP12, NSP13, NSP14, NSP15, and surface glycoprotein of the SARS-CoV-2 with the corresponding proteins of SARS-CoV. This has been consistent with previous reports on the phylogenetic relatedness of SARS-CoV-2 with SARS-CoV, although, the highest genetic sequence similarity was observed with bat-derived SARS-like virus (~88% genetic identity) which proves its zoonotic origin [2,28, 29]. These observations may have possible implications on the therapeutic and surveillance strategies since protein similarities in NSPs and surface glycoprotein between these two betacoronaviruses may yield cross-protection between SARS-CoV-2 and SARS-CoV as previously observed in cases of other human coronavirus infection; explain possible similarities in the mechanism of infection, hence, treatment; and prevent the error of using SARS-CoV-2 and SARS-CoV homologous epitopes in antibody-based detection which, in serological assays, have been known to be the cause inability to correctly discriminate closely-related pathogens, thus, decreased specificity of the serological test [30–32]. The protein with the greatest number of homologous epitopes with SARS-CoV, based on the blastp performed, is the surface glycoprotein. Seventeen out of the 21 in silico-predicted epitopes of the SARS-CoV-2 surface glycoprotein are at least 64% homologous with the epitopes of the SARS-CoV
spike glycoprotein. This observation is very important to note since the surface glycoprotein is pathogenically and serologically important because of its role in viral and host cell membrane fusion, hence, a good prospect as epitope-based vaccine due to its ability to produce viral-neutralizing antibodies [33–34].

In the polyprotein, the portion which has the highest number of predicted epitopes is at the putative position of the NSP3 protein located between amino acid position 920 to 2665. This portion also contains the most number (8 of 11) of SARS-CoV–2 unique epitopes not only for the polyprotein but for all the reported proteins analyzed based on the blastp analysis we performed. The finding is not surprising knowing that the NSP3 is the largest nonstructural protein of CoVs and has been reported to be heavily involved in proteolytic processing and polyprotein maturation. Furthermore, it was reported that NSP3 is involved in multiple interactions with other NSPs providing cooperative enzymatic functions. Surprisingly, the NSP3 is highly divergent among CoVs with mutations leading to evolutionary adaptations specific to certain coronaviruses [35–37]. The NSP3/4 macrodomain and transmembrane units are also critical for the ability of coronaviruses to evade the immune system. Experimental studies in both SARS-CoV and MERS-CoV revealed that subunits of NSP3/4 induced the formation of double-membrane vesicles (DMVs), which are specialized replicative organelles (ROs), that enhances viral RNA synthesis while hiding double-stranded RNA from detection by the innate immune system [6,38,39]. A study mentioned the detection of proteinases NSP3 and NSP5 in the mature virion along with the structural proteins [40]. This phenomenon should be elucidated further as data on NSP3 is relatively scarce compared to its structural counterparts.

On the other hand, the identified unique residues, especially for relevant proteins such as the surface glycoprotein, can be further explored experimentally to confirm its feasibility and uniqueness against other viruses, particularly, coronaviruses. During the SARS outbreak, there was difficulty in identifying actual SARS cases from common cold viruses based on serological tests as there was a high seroprevalence in the population of antibodies against the common cold, aggravated by the presence of cross-reactive antibodies against conserved coronavirus epitopes. Nevertheless, serological testing has its advantage of detecting asymptomatic infections, monitoring disease progression and study of
post-infection transmission dynamics [25,41].

Methods
The amino acid sequence of the SARS-CoV–2 polyprotein (GenBank: QHO60603.1), surface glycoprotein (GenBank: QHO60594.1), orf3 (QHO60595.1), envelop protein (QHO60596.1), membrane glycoprotein (QHO60597.1), orf6 (QHO60598.1), orf7a (QHO60599.1), orf8 (QHO60560.1), nucleocapsid phosphoprotein (QHO60561.1) and orf10 (QHO60595.1) were acquired from the National Center for Biotechnology Information (NCBI). The reported Genpept sequences were used in the identification of the linear continuous B-cell epitopes.

The criteria used in identifying putative immunogenic epitopes are antigenicity, surface accessibility, and hydrophilicity based on known computational tools available at the Immune Epitope Database and Analysis Resources (IEDB) [42-44]. The window size was set to 10 amino acids. Antigenicity, surface accessibility, and hydrophilicity scores, derived from IEDB analysis were compared to the computed cut-off value set by each parameter. Peptides which score below the cut-off of at least one of the three parameters were excluded. Adjacent predicted immunogenic sequences, which are positioned continuously, were considered and reported as one epitope.

After epitope prediction, sequence homology of the predicted immunogenic epitope was done to identify related viral proteins. Proteins reported to have similarity with the predicted immunogenic epitope, its origin, and percent identity with the query sequence were noted and reported. Putative amino acid positions in the SARS-CoV–2 were compared with reference alignment of a bat coronavirus sequence (data not shown) and positions reported in recent literature [18,19,29].

Declarations

Author’s Contributions
LAG – performed in silico epitope mapping, blastp analysis and preparation of manuscript
GLU – performed identification possible location of identified epitopes in coronavirus proteins and preparation of manuscript

Competing Interests
The authors declare no competing interests.

References

1. Song, Z. et al. From SARS to MERS, thrusting coronaviruses into the spotlight.Viruses. 11,59; https://doi.org/10.3390/v11010059 (2019)

2. Tang, J. W., Tambyah, P. A., & Hui, D. S. C. Emergence of a novel coronavirus causing respiratory illness from Wuhan, China. J Infect. https://doi.org/10.1016/j.jinf.2020.01.014 (2020)

3. Fehr, A., & Perlman, S. Coronaviruses: An Overview of Their Replication and Pathogenesis. Methods Mol Biol, 1282, 1-282 (2015)

4. Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu Rev Virol, 3(1), 237–261 (2016)

5. de Groot et al. Part II—The Positive Sense Single Stranded RNA Viruses Family Coronaviridae. In Virus Taxonomy: Ninth report of the International Committee on Taxonomy of Viruses (pp. 806–828) (2012)

6. Fehr, A. R. et al. The nsp3 Macrodomain Promotes Virulence in Mice with Coronavirus-Induced Encephalitis. J Virol, 89(3), 1523–1536 (2015).

7. Sharmin, R., & Islam, A. B. M. M. K. Conserved antigenic sites between MERS-CoV and Bat-coronavirus are revealed through sequence analysis. Source Code Biol Med, 11(1), 1–6. https://doi.org/10.1186/s13029-016-0049-7 (2016).

8. Chen, N. et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet, 395(10223), 507-513 (2020).

9. Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet, 6736(20), 1-10. https://doi.org/10.1016/S0140-
10. Nishiura H. et al. The Extent of Transmission of Novel Coronavirus in Wuhan, China, 2020. J Clin Med, 9(2), 330. https://doi.org/10.3390/jcm9020330 (2020).

11. World Health Organization (WHO). Novel Coronavirus (2019-nCoV) Situation report 16. WHO Bulletin, (JANUARY), 1–7 (2020).

12. World Health Organization (WHO). Coronavirus disease 2019 (COVID-19) Situation Report-55 (2020).

13. Cui, J., Li, F., & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol, 17(3), 181-192 (2019).

14. Li, H. et al. Human-animal interactions and bat coronavirus spillover potential among rural residents in Southern China. Biosafety and Health, 1(2), 84-90 (2019).

15. Zhu Z. et al. Predicting the receptor-binding domain usage of the coronavirus based on kmer frequency on spike protein. Infect Genet Evol, 61, 183-184 (2018).

16. Nishiura, H., Linton, N. M., & Akhmetzhanov, A. R. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis, 113332. https://doi.org/10.1016/j.jelechem.2019.113332 (2020).

17. European Center for Disease Prevention and Control (ECDC). Outbreak of acute respiratory syndrome associated with a novel coronavirus, China ; First cases imported in the EU / EEA ; second update (2020).

18. Hassan, A. et al. Pangenome and immuno-proteomics analysis of Acinetobacter baumannii strains revealed the core peptide vaccine targets. BMC Genomics, 17(732). https://doi.org/10.1186/s12864-016-2951-4 (2016).

19. Potocnakova, L., Bhide, M., & Pulzova, L. B. An Introduction to B-Cell Epitope Mapping and in Silico Epitope Prediction. J Immunol Res,. https://doi.org/10.1155/2016/6760830 (2016).
20. Liu, S. L., & Saif, L. Emerging Viruses without Borders: The Wuhan Coronavirus. Viruses, 12(2), 9-10 (2020).

21. Heather, J. M., & Chain, B.. The sequence of sequencers: The history of sequencing DNA. Genomics, 107(1). https://doi.org/10.1016/j.ygeno.2015.11.003 (2016)

22. Diniz, W. J. S., & Canduri, F. Bioinformatics: An overview and its applications. Genet Mol Res, 16(1), https://doi.org/10.4238/gmr16019645 (2017).

23. Backert, L., & Kohlbacher, O. Immunoinformatics and epitope prediction in the age of genomic medicine. Genome Med, 7(1), https://doi.org/10.1186/s13073-015-0245-0 (2015).

24. Kringeluma, J. V., Nielsena, M., Padkjær, S. B., & Lunda, O. Structural analysis of B-cell epitopes in antibody:protein complexes. Mol Immunol, 53(1–2), 24–34 (2013).

25. Meyer, B., Drosten, C., & Müller, M. A. Serological assays for emerging coronaviruses: Challenges and pitfalls. Virus Res, 194, 175-183 (2014).

26. Hua, C. et al. Computationally-driven identification of antibody epitopes. ELife, 6, https://doi.org/10.7554/eLife.29023 (2017).

27. Shey, R. A. et al. In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. Scientific Reports, 9(1). https://doi.org/10.1038/s41598-019-40833-x (2019)

28. Lu, R. et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet, 6736(20), https://doi.org/10.1016/S0140-6736(20)30251-8 (2020)

29. Wu, A. et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe, https://doi.org/10.1016/j.chom.2020.02.001 (2020)

30. Capriles, P. V. S. Z. et al. Structural modelling and comparative analysis of
homologous, analogous and specific proteins from Trypanosoma cruzi versus Homo sapiens: Putative drug targets for chagas’ disease treatment. BMC Genomics, 11(1), https://doi.org/10.1186/1471-2164-11-610 (2010).

31. Chan, K. H. et al. Serological responses in patients with severe acute respiratory syndrome coronavirus infection and cross-reactivity with human coronaviruses 229E, OC43, and NL63. Clin Diag Lab Immunol, 12(11), 1317-1321 (2005).

32. Che, X. et al. Antigenic Cross-Reactivity between Severe Acute Respiratory Syndrome-Associated Coronavirus and Human Coronaviruses 229E and OC43. J Infect Dis, 191(12), 2033-2037 (2005).

33. Guillen, J., Perez-Berna, A. J., Moreno, M. R., & Villalain, J.. Identification of the Membrane-Active Regions of the Severe Acute Respiratory Syndrome Coronavirus Spike Membrane Glycoprotein Using a 16/18-Mer Peptide Scan: Implications for the Viral Fusion Mechanism. J Virol, 11(8), 781-783 (2005).

34. Zhu Z. Potent cross-reactive neutralization of SARS coronavirus isolates by human monoclonal antibodies. PNAS, 104(29), 12123-12128 (2007).

35. Forni, D. et al. Extensive Positive Selection Drives the Evolution of Nonstructural Proteins in Lineage C Betacoronaviruses. J Virol, 90(7), 3627-3639 (2016).

36. Imbert, I. et al. The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. Virus Res, 133(2), 136-148 (2008).

37. Lei, J., Kusov, Y., & Hilgenfeld, R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. Antiviral Res, 149, 58-74 (2018).

38. Angelini, M., Akhlaghpour, M., Neuman, B. W., & Buchmeier, M. J. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Proteins 3, 4, and 6 Induce Double-Membrane Vesicles. MBio, 4(4), 165, https://doi.org/10.1128/mBio.00524-13 (2013).

39. Oudshoorn, D. et al. Expression and cleavage of middle east respiratory syndrome
coronavirus nsp3–4 polyprotein induce the formation of double-membrane vesicles that mimic those associated with coronaviral RNA replication. MBio, 8(6), https://doi.org/10.1128/mBio.01658-17 (2017).

40. Neuman, B. W. et al. Proteomics Analysis Unravels the Functional Repertoire of Coronavirus Nonstructural Protein 3. J Virol, 82(11), 5279–5294 (2008).

41. Niedrig, M. et al. First external quality assurance of antibody diagnostic for SARS-new coronavirus. J Clin Virol, 34(1), 22–25 (2005).

42. Kolaskar, S., & Tongaonkar, P. C. A semi-empirical method for prediction of antigenic determinants on protein antigens. Febbs Lett, 276(1-2), 172-174 (1990).

43. Parker, J. M. R., Guo, D., & Hodges, R. S. New Hydrophilicity Scale Derived from High-Performance Liquid Chromatography Peptide Retention Data: Correlation of Predicted Surface Residues with Antigenicity and X-ray-Derived Accessible Sites. Biochemistry, 25(19), 5425–5432 (1986).

44. Vita, R. et al. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res, 47(D1), D339–D343. https://doi.org/10.1093/nar/gky1006 (2019)

Tables
Table 1. Peptide Sequence of Continuous Epitopes of SARS-CoV-2 Polyprotein and Similar Proteins of Viral Origins as Identified by blastp Analysis

| Position in SARS-CoV-2 | Potential protein in SARS-CoV-2 | SARS-CoV-2 Epitope Sequence | Subject Peptide Sequence | % Homology | Organism Source and Protein of Related Sequence |
|-------------------------|--------------------------------|----------------------------|--------------------------|------------|-----------------------------------------------|
| 9-19                   | NSP1                           | NEKTHVQLSLP                | ****HVQLSLP             | 64%        | SARS-CoV NSP1                                  |
| 35-44                  | NSP7                           | VEEVLSEARQ                 | VEEALSEARE              | 80%        | SARS-CoV NSP1                                  |
| 38-47                  | NSP7                           | VLSEARQHLK                 | LSEAREHLK               | 80%        | SARS-CoV NSP1                                  |
| 57-67                  | NSP8                           | EKGVLQPQLEQP               | EKGVLQPQLEP             | 100%       | SARS-CoV NSP1                                  |
| 66-75                  | NSP8                           | QPYVFIKRSD                 | QPYVFIKRSD              | 100%       | SARS-CoV NSP1                                  |
| NSP  | Range       | Sequence        | Identity | Description                                      |
|------|-------------|-----------------|----------|--------------------------------------------------|
| 2     | 301-311     | RSVYPVSPNE      | 45%      | Human alphaherpesvirus envelop glycoprotein      |
| 3     | 488-497     | VETVKGLDYK      | -        | -                                                |
| 3     | 495-504     | **KAFMQVVE      | 60%      | Escherichia phage T7 RNA polymerase              |
| 3     | 555-564     | AQNSVRVLQKA     | -        | -                                                |
| 3     | 618-627     | GTVYEKLPV       | 50%      | Sindbis virus envelop polyprotein               |
| 3     | 711-720     | THSKGLRYKC      | 60%      | Rotavirus NSP2                                  |
| 3     | 713-725     | SKGLYRCKVKSRE   | -        | -                                                |
| 3     | 765-774     | PLEOPTSEAV      | 50%      | Human alphaherpesvirus capsid                    |
| 3     | 794-803     | KDTEKCALA       | 50%      | Influenza A(H5N1) virus RNA Polymerase           |
| 4     | 920-929     | MYCSFYPPDE      | 91%      | SARS-CoV NSP3                                  |
| 4     | 1006-1015   | VEVPQPEME       | -        | -                                                |
| 4     | 1045-1054   | IVEEAKVKP       | -        | -                                                |
| 4     | 1048-1057   | EAKKVPTVV       | -        | -                                                |
| 4     | 1180-1189   | KNLYDKLVSS      | 60%      | Influenza A(H5N2) virus hemagglutinin            |
| 4     | 1198-1207   | QVEQKIAEIP      | 60%      | SARS-CoV spike glycoprotein                      |
| 4     | 1213-1222   | PFITESKPSV      | 70%      | Rotavirus A capsid protein                       |
| 4     | 1227-1236   | QDDKKIKACV      | -        | -                                                |
| 4     | 1338-1352   | TVEEAKTLKCKS    | 87%      | SARS-CoV unique domain                          |
| 4     | 1398-1407   | IVSTIQRKYK      | 80%      | SARS-CoV NSP3 (Replicase)                        |
| 4     | 1423-1432   | YFYTSKTTVA      | 70%      | SARS-CoV NSP3 (Replicase)                        |
| 4     | 1425-1434   | YTSKTTVSL       | 70%      | SARS-CoV NSP3 (Replicase)                        |
| 4     | 1528-1537   | LKRGDKSVYV      | 80%      | SARS-CoV NSP3 (Replicase)                        |
| 4     | 1628-1638   | RVEAFEEYHTT     | 82%      | SARS-CoV NSP3 (Replicase)                        |
| 4     | 1691-1702   | NPPALQDAYRA     | 83%      | SARS-CoV NSP3 (Replicase)                        |
| 4     | 1774-1783   | LSYEQQKKGV      | 50%      | Crimean-Congo Hemorrhagic Fever virus nucleoprotein |
| 4     | 1788-1797   | TCGKQATKYL      | 60%      | SARS-CoV NSP3 (Proteinase)                       |
| 4     | 1790-1803   | GQATKLYVQQES    | 71%      | SARS-CoV NSP3 (Proteinase)                       |
| 4     | 1808-1818   | SAPPAQYELKH     | 64%      | SARS-CoV NSP3 (Proteinase)                       |
| 4     | 1827-1828   | YTGNYQCHG      | 100      | SARS-CoV NSP3 (Proteinase)                       |
| Year  | Peptide Sequence     | Similarity (%) | Protein or Virus Origin                                      |
|-------|----------------------|----------------|-------------------------------------------------------------|
| 1836  | YTTIKPVTY            | 60%            | SARS-CoV NSP3 (Proteinase)                                  |
| 1873  | YTTIK****            |                | SARS-CoV NSP3 (Proteinase)                                  |
| 1882  |                     |                | SARS-CoV NSP3 (Proteinase)                                  |
| 1877- | IKPVTYYLDGV          | 45%            | Rift valley river virus glycoprotein                        |
| 1887  |                     |                | Rift valley river virus glycoprotein                        |
| 1909- | EQPIDLVNPQP          | 91%            | SARS-CoV NSP3 (Replicase)                                   |
| 1919  | EQPIDLVPTQP          |                | SARS-CoV NSP3 (Replicase)                                   |
| 1914- | LVPNQYPYNA           | 80%            | SARS-CoV NSP3 (Replicase)                                   |
| 1923  |                     |                | SARS-CoV NSP3 (Replicase)                                   |
| 1968- | VAIKYKHYP            | 70%            | SARS-CoV NSP3 (Replicase)                                   |
| 1977  |                     |                | SARS-CoV NSP3 (Replicase)                                   |
| 2041- | CEDLKPVSEEV          |                |                                                              |
| 2051  |                     |                |                                                              |
| 2046- | KPVSEEEVVENPT        |                |                                                              |
| 2483- | PTDQSSYIVD           | 70%            | Influenza A(H6N1) hemagglutinin                             |
| 2492  |                     |                | Influenza A(H6N1) hemagglutinin                             |
| 2514- | ERHLSHFVN            | 50%            | Human calcivirus capsid                                     |
| 2523  | *****SHFVN           |                | Human calcivirus capsid                                     |
| 2551- | ESSAKSASVYYSQL       |                |                                                              |
| 2564  |                     |                |                                                              |
| 2642- | VDSDVETKADV          | 55%            | Feline calcivirus viral protein1 capsid                     |
| 2652  | VNRDQPSDV*           |                | Feline calcivirus viral protein1 capsid                     |
| 2655- | LKLQHQSdiev          |                |                                                              |
| 2665  |                     |                |                                                              |
| NSP4  | PVHVMKHTD            | 50%            | FMD virus RNA polymerase                                    |
|       | *VHVMRK****          |                | FMD virus RNA polymerase                                    |
| 2792- | SYTNDKACPL           | 50%            | Influenza(H1N1) hemagglutinin                               |
| 2801  | SYTNKK****           |                | Influenza(H1N1) hemagglutinin                               |
| 2844- | SGKPVPCYDTN          | 50%            | Sacbrood virus viral protein 3                              |
| 2853  | *****VPYIYDT*        |                | Sacbrood virus viral protein 3                              |
| 2926- | SVAYESELRPD          | 50%            | Human herpes virus glycoprotein                             |
| 2951  | **AYDSLRR**          |                | Human herpes virus glycoprotein                             |
| 2974- | RVVTTDFSEY           | 70%            | DENV2 NS3 (RNA helicase)                                   |
| 2983  | *VVTT*DISE*          |                | DENV2 NS3 (RNA helicase)                                   |
| 3244- | SGSDVLYQPQPTS        | 62%            | DENV3 NS5 (Polymerase)                                     |
| 3256  | SGKDFVFYLPE**        |                | DENV3 NS5 (Polymerase)                                     |
| 3249- | LYQPPQTSIT           | 70%            | Murine coronavirus RNA polymerase                           |
| 3258  | LYQPPQMTVT           |                | Murine coronavirus RNA polymerase                           |
| NSP5  | YEPLTQDHDVI          | 100%           | SARS-CoV Proteinase(Main)                                   |
|       | YEPLTQHDVI           |                | SARS-CoV Proteinase(Main)                                   |
| 3502- |                     |                | SARS-CoV Proteinase(Main)                                   |
| 3512  |                     |                | SARS-CoV Proteinase(Main)                                   |
| NSP6  | YKYYGNALDQ           | 64%            | H-1 parvovirus capsid                                       |
| 3719- | YK**LYGPNSLDQ        |                | H-1 parvovirus capsid                                       |
| 3729  |                     |                | H-1 parvovirus capsid                                       |
| 3810- | VYDLYVSTQE           | 50%            | Bombix mori cytopovirus1 RNA polymerase                    |
| 3819  | *YDLYV****           |                | Bombix mori cytopovirus1 RNA polymerase                    |
|       |                     |                | Bombix mori cytopovirus1 RNA polymerase                    |

Table 2. Peptide Sequence of Continuous Epitopes of SARS-CoV-2 Structural and orf Proteins and Similar Proteins of Viral Origins as Identified by blastp Analysis
| Protein Location in SARS-CoV-2 | Start SARS-CoV-2 Sequence | Epitope Sequence of Subject Peptide | % Homology | Organism Source of Sequence |
|-------------------------------|---------------------------|----------------------------------|------------|---------------------------|
| Surface Glycoprotein          | 33-42 TRGYYYPDKV          | *RGVYYYPD**                      | 70%        | SARS-CoV spike glycoprotein |
|                               | 37-47 YYPDVKFRSSV         | YYPDIEFIRS**                     | 64%        | SARS-CoV spike glycoprotein |
|                               | 44-53 RSSVLHSTOD          | -                                | -          | -                         |
|                               | 295-305 PLSETKCTLKS       | PLAELKCSVKS                      | 64%        | SARS-CoV spike glycoprotein |
|                               | 319-328 RVQPTESIVR        | -                                | -          | -                         |
|                               | 321-330 QPTESIVRF         | -                                | -          | -                         |
|                               | 364-373 DYSVLYNAS         | DYSVLYN**                       | 80%        | SARS-CoV spike glycoprotein |
|                               | 491-500 PLQSYGFQPT        | ***SYGFQ**                       | 50%        | Hepatitis C virus RNA      |
|                               | 501-510 NGVGYPYPRV        | *GIGYOPYRV                      | 80%        | SARS-CoV spike glycoprotein |
|                               | 575-584 AVRDPQTLIE        | *VRDPKTSIE                      | 70%        | SARS-CoV spike glycoprotein |
|                               | 776-787 KNTQELFQAQVKQ     | *NTREVFQAQVKQ                   | 91%        | SARS-CoV spike glycoprotein |
|                               | 783-793 AOVKQIYKTPP       | AOVKQIYKTP*                     | 82%        | SARS-CoV spike glycoprotein |
|                               | 802-813 FSQILPDPSKS       | FSQILPDPSK*                     | 91%        | SARS-CoV spike glycoprotein |
|                               | 913-922 QNVLYENQK         | QNVLYENQK*                      | 90%        | SARS-CoV spike glycoprotein |
|                               | 947-957 KLQDVOKNQAQ       | KLQDVOKNQAQ                     | 100%       | SARS-CoV spike glycoprotein |
|                               | 982-992 SRLDKVEAEVQ       | SRLDKVEAEQ                      | 100%       | SARS-CoV spike glycoprotein |
|                               | 1020-1038 R2125     | R2125                             | 60%        | Human gammaherpesvirus R   |
|                               | 1045-1054 QSLQTYVTQQ      | QSLQTYVTQQ                      | 100%       | SARS-CoV spike glycoprotein |
|                               | 1074-1088 HVTYVPAQEN     | HVTYVPAQEN                      | 82%        | SARS-CoV spike glycoprotein |
|                               | 1133-1142 VNNTVYDPLQ      | VNNTVYDPLQ                      | 90%        | SARS-CoV spike glycoprotein |
|                               | 1172-1180YYKLGASQRV       | YYKLGASQRV*                     | 55%        | Hepacivirus E2 glycoprotein |
|                               | 1200-1209 DSEPVLKGVK      | DSEPVLKGVK                      | 100%       | SARS-CoV spike glycoprotein |
| orf3                          | 132-141 KCRSKNPLL   | KCRSKNPLL                        | 50%        | Human betaherpesvirus      |
|                               | 202-212 VLHSYFTSDYY       | VLHSYFTSDYY                     | 45%        | Cowpox virus serine        |
|                               | 210-221 DYYQLYSTLST      | DYYQLYSTLST                      | 42%        | Chikungunya virus glycoprotein |
| Envelop protein               | 50-60 SLVKPSFYVYS        | SLVKPSFYVYS                      | 91%        | SARS-CoV small envelop protein |
| Membrane Glycoprotein         | 10-19 VEELKILLEQ         | *EQLAKLLEQ                      | 70%        | Human alphaherpesvirus glycerase |
|                               | 129-139 LTRPLESELV       | **RPLMEPEL*                     | 45%        | Human alphaherpesvirus glycerase |
|                               | 170-179 VATSRTLSSY       | *****TLSYY                       | 50%        | Escherichia virus2 C glycoprotein |
|                               | 175-185 TLSYKLGASQ       | *LYTYYL*                        | 50%        | Human orthopneumovirus glycoprotein |
|                               | 178-188 YYYKLGSQRVA      | *****GASQRV*                    | 55%        | Hepacivirus E2 glycoprotein |
| orf7a                         | 16-25 ELHYHQECVR         | ELHYHQECVR                      | 100%       | SARS-CoV orf 7a accession protein |
|                               | 68-78 PDGVKHYQYLR        | PDGVKHYQYLR                     | 55%        | SARS-CoV orf 7a accession protein |
|                               | 71-80 VKHVYQRAR         | **HTYQRAR**                      | 70%        | SARS-CoV orf 7a accession protein |
|                               | 73-83 HVYQRLARSV       | HTYQRLARSV                      | 91%        | SARS-CoV orf 7a accession protein |
| orf8                          | 22-32 LQSCTQHQPYV        | LQSCT*****                      | 45%        | Influenza A (H17N10) protein R |
|                               | 25-35 CTQHQPYVVD         | CNQSTPPYYVVD                    | 60%        | Human gammaherpesvirus protein R |
|                               | 27-38 QHQPYYVVDPC        | **HPYVLDD***                    | 42%        | Enteroctyivirus D68 virus Protein R |
|                               | 35-44 DPCIHFSKY          | *****HDAVRIILD                  | 40%        | Delta coronavirus spike protein R |
| Nucleocapsid phosphoprotein   | 79-88 SPDDQIGYXR         | *PDDQIGYXR                      | 90%        | SARS-CoV nucleocapsid protein R |
|                               | 237-246 KGQQQQGQTV       | **QQQGQ**                       | 50%        | Bourbon virus envelope protein R |
|                               | 239-250 QQQGQTVKKS       | QQQGQ**                         | 42%        | Bourbon virus envelope protein R |
|                               | 375-384 KADETQALPQ       | *ADETQAL**                      | 60%        | Betacoronavirus English protein R |
|                               | 378-394 TL               | DDF**QLQQ*                      | 35%        | Sindbis virus coat protein R |
|                               | 401-410 DDFSQKQLQSD      | DDF**QLQQ*                      | 70%        | Norovirus Hu VP1 protein R |
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

SARS-CoV-2 Epitope Mapping_Guevarra&Ulanday_Supplementary.docx