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

Leonardo A. Guevarra Jr. (laguevarra@ust.edu.ph)
University of Santo Tomas
Gianne Eduard L. Ulanday
National Institutes of Health - University of the Philippines Manila https://orcid.org/0000-0002-1093-9287

Research Article

Keywords: coronavirus, SARS-CoV-2, novel coronavirus, 2019-nCoV, epitope map, COVID-19, severe acute respiratory syndrome coronavirus 2

Posted Date: March 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-18689/v1

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,* o*rf8,* 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 of were also found at positions 488—497 (VETVKGLDYK), 555—564 (AQNSVRVLQKA), 713—725 (SKGLYRKCVKSRE), 1006—1015 (VEVQPQLEME), 1045—1054 (IVEEAKKVKP), 1048—1057 (EAKKVKPTVV), 1227—1236 (QDDKKIKACV), 2041—2051 (CEDLKPVSEEV), 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 (QPTESIVRFP) 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/S0140–6736(20)30183–5 (2020).
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 treport 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., Padkjærb, 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., & Villalaın, 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., & Buchmeierc, 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 determininants 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

ɔsition	SARS-CoV-2 Epitope Sequence	Sequence of Subject Peptide	% Homology	Organism Source and Protein of Related Sequence
9-19	NEKTHVQLSLP	****HVQLSLP	64%	SARS-CoV NSP1
35-44	VEEVLSEARQ	VEEALSEARE	80%	SARS-CoV NSP1
38-47	VLSEARQHLK	LSEAREHLK	80%	SARS-CoV NSP1
57-67	EKGVLPQLEQP	EKGVLPQLEQP	100%	SARS-CoV NSP1
66-75	QPYVFIKRSD	QPYVFIKRSD	100%	SARS-CoV NSP1
)1-311	RSVYPVASPNE	RSVYP******	45%	Human alphaherpesvirus envelop glycoprotein
38-497	VETVKGLDYK	-	-	-
)5-504	DYKAFKQIVE	**KAFMQVVE	60%	Escherichia phage T7 RNA polymerase
55-564	AQNSVRVLQKA	-	-	-
18-627	GTVYEKLKPV	*TVYDRLK**	50%	Sindbis virus envelop polyprotein
11-720	THSKGLYRKC	THGKGHYR**	60%	Rotavirus NSP2
13-725	SKGLYRKCVKSRE	-	-	-
55-774	PLEQPTSEAV	PIEQPT****	50%	Human alphaherpesvirus capsid
)4-803	KDTEKYCALA	***EKYCVL*	50%	Influenza A(H5N1) virus RNA Polymerase
20-929	MYCSFYPPDED	MYCSFYPPDE*	91%	SARS-CoV NSP3
1006-1015	VEVQPQLEME	-	-	-
1045-1054	IVEEAKKVKP	-	-	-
1048-1057	EAKKVKPTVV	-	-	-
1180-1189	KNLYDKLVSS	KNLYDK****	60%	Influenza A(H5N2) virus hemagglutinin
1198-1207	QVEQKIAEIP	QIEDKIEEI*	60%	SARS-CoV spike glycoprotein
1213-1222	PFITESKPSV	PFITDS**SV	70%	Rotavirus A capsid protein
1227-1236	QDDKKIKACV	-	-	-
1338-1352	TVEEAKTVLKKCKSA	TLEEAKTALKKCKSA	87%	SARS-CoV unique domain
1398-1407	IVSTIQRKYK	IMATIQRKYK	80%	SARS-CoV NSP3 (Replicase)
1423-	YFYTSKTTVA	FFYTSKEPVA	70%	SARS-CoV NSP3

Putative protein in SARS-CoV-2	Position	SARS-CoV-2 Epitope Sequence	Sequence of Subject Peptide	% Homology	Organism Source and Protein of Related Sequence
NSP6/7	3854-3863	KVATVQSKMS	KVSTVQ****	50%	HIV glycoprotein gp160
NSP7	3876-3885	LQQLRVESSS	LQQLRVESSS	100%	SARS-CoV NSP7 (Replicase light chain)
	3878-3887	QLRVESSSKL	QLRVESSSKL	100%	SARS-CoV NSP7 (Replicase light chain)
NSP8	3945-3955	ASEFSSLPSYA	ASEFSSLPSYA	100%	SARS-CoV NSP8 (Replicase heavy chain)
	3960-3969	AQEAYEQAVA	AQEAYEQAVA	100%	SARS-CoV NSP8 (Replicase heavy chain)
	3972-3983	DSEVVLKKLKKS	DSEVVLKKLKKS	100%	SARS-CoV NSP8 (Replicase heavy chain)
	3978-3989	KKLKKSLNVAKS	KKLKKSLNVAKS	100%	SARS-CoV NSP8 (Replicase heavy chain)
NSP9	4142-4151	NELSPVALRQ	NELSPVALRQ	100%	SARS-CoV NSP9 (Replicase)
	4205-4214	IYTELEPPCR	IYTELEPPCR	100%	SARS-CoV NSP9 (Replicase)
	4212-4221	PCRFVTDTPK	PCRFVTDTPK	100%	SARS-CoV NSP9 (Replicase)
	4219-4228	TPKGPKVKYL	TPKGPKVKYL	100%	SARS-CoV NSP9 (Replicase)
NSP10	4274-4283	VDAAKAYKDY	VDPAKAYKDY	90%	SARS-CoV NSP10
	4276-4286	AAKAYKDYLAS	*AKAYKDYLAS	91%	SARS-CoV NSP10
	4344-4353	DLKGKYVQIP	DLKGKYVQIP	100%	SARS-CoV NSP10
NSP12	4459-4468	DSYFVVKRHT	DSYFVVKRHT	100%	SARS-CoV NSP12
	4463-4472	VVKRHTFSNY	VVKRHTMSNY	90%	SARS-CoV NSP12
	4648-4658	HVDTDLTKPYI	HMDADLAKPLI	64%	SARS-CoV NSP12
	4683-4692	DQTYHPNCVN	DQTYHPNCIN	90%	SARS-CoV NSP12
	4749-4758	QDVNLHSSRL	QDVNLHSSRL	100%	SARS-CoV NSP12
	4841-4850	AISDYDYYRY	AISDYDYYRY	100%	SARS-CoV NSP12
	4994-5003	LKTVYSDVEN	LKTVYSDVE*	90%	SARS-CoV NSP12
	4996-5005	TVYSDVENPH	TVYSDVETPH	90%	SARS-CoV NSP12
	5107-5116	IADKYVRNLQ	IADKYVRNLQ	100%	SARS-CoV NSP12
	5113-5122	RNLQHRLYEC	RNLQHRLYEC	100%	SARS-CoV NSP12
	5119-	LYECLYRNRDV	LYECLYRNRDV	100%	SARS-CoV NSP12

1432				(Replicase)
1425-1434	YTSKTTVASL	YTSKEPVAS	70%	SARS-CoV NSP3 (Replicase)
1528-1537	LKRGDKSVYY	LKRGDKIVY*	80%	SARS-CoV NSP3 (Replicase)
1628-1638	RVEAFEYYHTT	RSEAFEYYHT*	82%	SARS-CoV NSP3 (Replicase)
1691-1702	NPPALQDAYYRA	NAPALQEAYYRA	83%	SARS-CoV NSP3 (Replicase)
1774-1783	LSYEQFKKGV	**FEEFKKG*	50%	Crimean-Congo Hemorrhagic Fever virus nucleoprotein
1788-1797	TCGKQATKYL	*CGRDATQYL	60%	SARS-CoV NSP3 (Proteinase)
1790-1803	GKQATKYLVQQESP	GRDATQYLVQQES*	71%	SARS-CoV NSP3 (Proteinase)
1808-1818	SAPPAQYELKH	SAPPAEYKL**	64%	SARS-CoV NSP3 (Proteinase)
1827-1836	YTGNYQCGHY	YTGNYQCGHY	100%	SARS-CoV NSP3 (Proteinase)
1873-1882	YTTTIKPVTY	YTTTIK****	60%	SARS-CoV NSP3 (Proteinase)
1877-1887	IKPVTYKLDGV	*KPVTY*****	45%	Rift valley river virus glycoprotein
1909-1919	EQPIDLVPNQP	EQPIDLVPTQP	91%	SARS-CoV NSP3 (Replicase)
1914-1923	LVPNQPYPNA	LVPTQPLPNA	80%	SARS-CoV NSP3 (Replicase)
1968-1977	VAIDYKHYTP	VAIDYRHY**	70%	SARS-CoV NSP3 (Replicase)
2041-2051	CEDLKPVSEEV	-	-	-
2045-2056	KPVSEEVVENPT	-	-	-
2483-2492	PTDQSSYIVD	**DQSWSYIVE	70%	Influenza A(H6N1) hemagglutinin
2514-2523	ERHSLSHFVN	*****SHFVN	50%	Human calcivirus capsid
2551-2564	ESSAKSASVYYSQL	-	-	-
2642-2652	VDSDVETKDVV	VNRDVQTSDV*	55%	Feline calcivirus viral protein1 capsid
2655-2665	LKLSHQSDIEV	-	-	-
2792-2801	PVHVMSKHTD	*VHVMRK***	50%	FMD virus RNA polymerase
2844-2853	SYTNDKACPL	SYTNNK****	50%	Influenza(H1N1) hemagglutinin
2926-2937	SGKPVPYCYDTN	****VPYIYDT*	50%	Sacbrood virus viral protein 3
2942-2951	SVAYESLRPD	**AYDSLR**	50%	Human herpes virus glycoprotein
2974-2983	RVVTTFDSEY	*VVTT*DISE*	70%	DENV2 NS3 (RNA helicase)
3244-3256	SGSDVLYQPPQTS	SGKDVFYLPPE**	62%	DENV3 NS5 (Polymerase)
3249-3258	LYQPPQTSIT	LYQPPTASVT	70%	Murine coronavirus RNA polymerase
3502-3512	YEPLTQDHVDI	YEPLTQDHVDI	100%	SARS-CoV Proteinase(Main)

	5129			
	5175-5184	KSVLYYQNNV	KAVLYYQNNV	90%	SARS-CoV NSP12
	5199-5208	KGPHEFCSQH	KGPHEFCSQH	100%	SARS-CoV NSP12
	5211-5225	LVKQGDDYVYLPYPD	LVKQGDDYVYLPYPD	100%	SARS-CoV NSP12
NSP13	5355-5364	YDHVISTSHK	YDHVISTSHK	100%	SARS-CoV Helicase
	5392-5404	MSYYCKSHKPPIS	*SYYCKSHKPPIS	92%	SARS-CoV Helicase
	5476-5485	ATVREVLSDR	ATVREVLSDR	100%	SARS-CoV Helicase
	5478-5487	VREVLSDREL	VREVLSDREL	100%	SARS-CoV Helicase
	5533-5542	VVYRGTTTYK	VVYRGTTTYK	100%	SARS-CoV Helicase
	5560-5570	SAPTLVPQEHY	SAPTLVPQEHY	100%	SARS-CoV Helicase
	5587-5596	SSNVANYQKV	SSNVANYQKV	100%	SARS-CoV Helicase
	5647-5656	KYLPIDKCSR	KYLPIDKCSR	100%	SARS-CoV Helicase
	5671-5680	KVNSTLEQYV	KVNSTLEQYV	100%	SARS-CoV Helicase
	5711-5721	VNARLRAKHYV	VNARLRAKHYV	100%	SARS-CoV Helicase
	5716-5725	RAKHYVYIGD	RAKHYVYIGD	100%	SARS-CoV Helicase
	5742-5751	EPEYFNSVCR	EPEYFNSVCR	100%	SARS-CoV Helicase
	5777-5786	SALVYDNKLK	SALVYDNKLK	100%	SARS-CoV Helicase
	5787-5796	AHKDKSAQCF	AHKDKSAQCF	100%	SARS-CoV Helicase
	5852-5861	LPTQTVDSSQ	LPTQTVDSSQ	100%	SARS-CoV Helicase
	5860-5869	SQGSEYDYVI	SQGSEYDYVI	100%	SARS-CoV Helicase
NSP14	5944-5958	HPTQAPTHLSVDTKF	HPTQAPTHLSVDIKF	100	SARS_Guanine-n7 methyltransferase
	6058-6067	FSRVSAKPPP	FTRVNAKPPP	80%	SARS_Guanine-n7 methyltransferase
	6274-6289	KFYDAQPCSDKAYKIE	KFYDAQPCSDKAYKIE	100%	SARS_Guanine-n7 methyltransferase
	6345-6354	YVNKHAFHTP	YVNKHAFHTP	100%	SARS_Guanine-n7 methyltransferase
	6370-6380	FYYSDSPCESH	FYYSDSPCESH	100%	SARS_Guanine-n7 methyltransferase
	6386-6395	SDIDYVPLKS	SDIDYVPLKS	100%	SARS_Guanine-n7 methyltransferase
	6408-6418	VCRHHANEYRL	VCRHHANEYRL	100%	SARS_Guanine-n7 methyltransferase
NSP15	6602-6612	GLQPSVGPKQA	*LQPSV*****	45%	Human rhinovirus RNA polymerase
	6628-6640	FNYYKKVDGVVQQ	FNYFKKVDGIIQQ	77%	SARS-CoV NSP15
	6759-6768	KSQDLSVVSK	KSQDLSVISK	90%	SARS-CoV NSP15

3719-3729	YKVYYGNALDQ	YK**YLGPGNSLDQ	64%	H-1 parvovirus capsid
3810-3819	VYDYLVSTQE	*YDYLV****	50%	Bombix mori cypovirus1 RNA polymerase
NSP15/16	6792-6802	TFYPKLQSSQA	TFYPKLQ****	64%	SARS-CoV NSP15

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 Epitope Sequence	Sequence of Subject Peptide	% Homology	Organism Source and Protein of Related Sequence
Surface Glycoprotein	33-42	TRGVYYPDKV	*RGVYYPD**	70%	SARS-CoV spike glycoprotein
	37-47	YYPDKVFRSSV	YYPDEIFRS**	64%	SARS-CoV spike glycoprotein
	44-53	RSSVLHSTQD	-	-	-
	295-305	PLSETKCTLKS	PLAELKCSVKS	64%	SARS-CoV spike glycoprotein
	319-328	RVQPTESIVR	-	-	-
	321-330	QPTESIVRFP	-	-	-
	364-373	DYSVLYNSAS	DYSVLYNS**	80%	SARS-CoV spike glycoprotein
	491-500	PLQSYGFQPT	***SYGFQ**	50%	Hepatitis C virus RNA polymerase
	501-510	NGVGYQPYRV	*GIGYQPYRV	80%	SARS-CoV spike glycoprotein
	575-584	AVRDPQTLEI	*VRDPKTSEI	70%	SARS-CoV spike glycoprotein
	776-787	KNTQEVFAQVKQ	*NTREVFAQVKQ	91%	SARS-CoV spike glycoprotein
	783-793	AQVKQIYKTPP	AQVKQMYKTP*	82%	SARS-CoV spike glycoprotein
	802-813	FSQILPDPSKPS	FSQILPDPLKP*	91%	SARS-CoV spike glycoprotein
	911-920	VTQNVLYENQ	VTQNVLYENQ	100%	SARS-CoV spike glycoprotein
	913-922	QNVLYENQKL	QNVLYENQK*	90%	SARS-CoV spike glycoprotein
	947-957	KLQDVVNQNAQ	KLQDVVNQNAQ	100%	SARS-CoV spike glycoprotein
	982-992	SRLDKVEAEVQ	SRLDKVEAEVQ	100%	SARS-CoV spike glycoprotein
	1002-1011	QSLQTYVTQQ	QSLQTYVTQQ	100%	SARS-CoV spike glycoprotein
	1064-1074	HVTYVPAQEKN	HVTYVPSQERN	82%	SARS-CoV spike glycoprotein
	1133-1142	VNNTVYDPLQ	INNTVYDPLQ	90%	SARS-CoV spike glycoprotein
	1260-1269	DSEPVLKGVK	VYDPLQPELD	100%	SARS-CoV spike glycoprotein
orf3	132-141	KCRSKNPLLY	*****NPLLY	50%	Human betaherpesvirus 5 envelop protein
	202-212	VLHSYFTSDYY	*****FTSDY*	45%	Cowpox virus serine proteinase inhibitor
	210-221	DYYQLYSTQLST	*YYELYPT****	42%	Chikungunya virus glycoprotein E2
Envelop protein	50-60	SLVKPSFYVYS	SLVKPTVYVYS	91%	SARS-CoV small envelop protein
Membrane Glycoprotein	10-19	VEELKKLLEQ	*EQLAKLLEQ	70%	Infectious pancreatic necrosis virus RNA polymerase
	129-139	LTRPLLESELV	**RPLMEPEL*	45%	Human alphaherpesvirus 1 Uracil DNA glycosylase
	170-179	VATSRTLSYY	*****TLSYY	50%	Escherichia virus2 C protein
	175-185	TLSYYKLGASQ	*LTYYKL****	50%	Human orthopneumovirus RNA polymerase
	178-188	YYKLGASQRVA	****GASQRV*	55%	Hepacivirus E2 glycoprotein
orf7a	16-25	ELYHYQECVR	ELYHYQECVR	100%	SARS-CoV orf 7a accessory protein
	68-78	PDGVKHVYQLR	*DGTRHTYQLR	55%	SARS-CoV orf 7a accessory protein
	71-80	VKHVYQLRAR	**HTYQLRAR	70%	SARS-CoV orf 7a accessory protein
	73-83	HVYQLRARSVS	HTYQLRARSVS	91%	SARS-CoV orf 7a accessory protein
orf8	22-32	LQSCTQHQPYV	LQSCT******	45%	Influenza A (H17N10) virus Hemagglutinin
	25-35	CTQHQPYVVD	CNQSTPYYVVD	60%	Human gammaherpesvirus 4 early antigen protein R
	27-38	QHQPYVVDDPCP	**HPYVLDD***	42%	Enterovirus D68 viral protein 2
	35-44	DPCPIHFYSK	*****HFYSK	50%	Rabies virus SADB19 Large Structural Protein
	110-119	EYHDVRVVLD	**HDAVRIILD	40%	Delta coronavirus spike protein
Nucleocapsid phosphoprotein	79-88	SPDDQIGYYR	*PDDQIGYYR	90%	SARS-CoV nucleocapsid protein
	237-246	KGQQQQGQTV	**QQQQG***	50%	Bourbon virus envelop glycoprotein
	239-250	QQQQGQTVTKKS	QQQQG*******	42%	Bourbon virus envelop glycoprotein
	375-384	KADETQALPQ	*ADETKAL**	60%	Betacoronavirus England 1 NSP3
	378-394	ETQALPQRQKKQQTVTL	*****PPRQKKQ****	35%	Sindbis virus coat protein C
	401-410	DDFSKQLQQS	DDF**QLQQ*	70%	Norovirus Hu VP1

Supplementary Files

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

- SARSCoV2EpitopeMappingGuevarraUlandaySupplementary.docx