Immunoinformatic identification of B cell and T cell epitopes in the SARS-CoV-2 proteome

Stephen N. Crooke, Inna G. Ovsyannikova, Richard B. Kennedy, Gregory A. Poland*

Mayo Clinic Vaccine Research Group, Mayo Clinic, Rochester, MN USA

Correspondence:
Gregory A. Poland, M.D., Director – Vaccine Research Group, Mayo Clinic, Guggenheim Building 611C, 200 First Street SW, Rochester, MN 55905 USA

Phone: (507) 284-4968; E-mail: poland.gregory@mayo.edu

Abstract

A novel coronavirus (SARS-CoV-2) emerged from China in late 2019 and rapidly spread across the globe, infecting millions of people and generating societal disruption on a level not seen since the 1918 influenza pandemic. A safe and effective vaccine is desperately needed to prevent the continued spread of SARS-CoV-2; yet, rational vaccine design efforts are currently hampered by the lack of knowledge regarding viral epitopes targeted during an immune response, and the need for more in-depth knowledge on betacoronavirus immunology. To that end, we developed a computational workflow using a series of open-source algorithms and webtools to analyze the proteome of SARS-CoV-2 and identify putative T cell and B cell epitopes. Using increasingly stringent selection criteria to select peptides with significant HLA promiscuity and predicted antigenicity, we identified 41 potential T cell epitopes (5 HLA class I, 36 HLA class II) and 6 potential B cell epitopes, respectively. Docking analysis and binding predictions demonstrated enrichment for peptide binding to HLA-B (class I) and HLA-DRB1 (class II) molecules. Overlays of predicted B cell epitopes with the structure of the viral spike (S) glycoprotein revealed that 4 of 6 epitopes were located in the receptor-binding domain of the S protein. To our knowledge, this is the first study to comprehensively analyze all 10 (structural, non-structural and accessory) proteins from SARS-CoV-2 using predictive algorithms to identify potential targets for vaccine development.

Keywords: Coronavirus; immunoinformatics; T-cell epitope; B-cell epitope; HLA molecules, HLA class I, HLA class II, peptide

Significance Statement:

The novel coronavirus SARS-CoV-2 recently emerged from China, rapidly spreading and ushering in a global pandemic. Despite intensive research efforts, our knowledge of SARS-CoV-2 immunology and the proteins targeted by the immune response remains relatively limited, making it difficult to rationally design candidate vaccines. We employed a suite of bioinformatic tools, computational algorithms, and structural modeling to comprehensively analyze the entire SARS-CoV-2 proteome for potential T cell and B cell epitopes. Utilizing a set of stringent selection criteria to filter peptide epitopes, we identified 41 T cell epitopes (5 HLA class I, 36 HLA class II) and 6 B cell epitopes that could serve as promising targets for peptide-based vaccine development against this emerging global pathogen.

Introduction

In December 2019, public health officials in Wuhan, China, reported the first case of severe respiratory disease attributed to infection with the novel coronavirus SARS-CoV-2 (1). Since its emergence, SARS-CoV-2 has spread rapidly via human-to-human transmission (2), threatening to overwhelm healthcare systems around the world and resulting in the declaration of a pandemic by the World Health Organization (3). The disease caused by the virus (COVID-19) is characterized by fever, pneumonia, and other respiratory and inflammatory symptoms that can result in severe inflammation of lung tissue and ultimately death—particularly among older adults or individuals with underlying comorbidities (4-6). As of this writing, the SARS-CoV-2 pandemic has resulted in 4 million confirmed cases of COVID-19 and over 280,000 deaths worldwide (7).

SARS-CoV-2 is the third pathogenic coronavirus to cross the species barrier into humans in the past two decades, preceded by severe acute respiratory syndrome coronavirus (SARS-CoV) (8, 9) and Middle-East respiratory syndrome coronavirus (MERS-CoV) (10). All three of these viruses belong to the β-coronavirus genus and have either been confirmed (SARS-CoV) or suggested (MERS-CoV, SARS-CoV-2) to originate in bats, with transmission to humans occurring through intermediary animal hosts (11-14). While previous zoonotic spillovers of coronaviruses have been marked by high case fatality rates (~10% for SARS-CoV; ~34% for MERS-CoV), widespread transmission of disease has been relatively limited (8,098 cases of SARS; 2,494 cases of MERS) (15). In contrast, SARS-CoV-2 is estimated to have a lower case fatality rate (~2-4%) but is far more infectious and has achieved world-wide spread in a matter of months (16).

As the number of COVID-19 cases continues to grow, there is an urgent need for a safe and effective vaccine to combat the spread of SARS-CoV-2 and reduce the burden on hospitals and healthcare systems. No licensed vaccine or therapeutic is currently available for SARS-CoV-2, although there are over 100 vaccine candidates reportedly in development worldwide. Seven vaccine candidates have rapidly progressed into Phase I/II clinical trials: adenoviral vector-based vaccines (CanSino Biologics, ChiCTR2000030906; University of Oxford, NCT04324606), nucleic-acid based vaccines encoding for the viral spike (S) protein (Moderna, NCT04283461; Inovio Pharmaceuticals, NCT04336410; BioNTech/Pfizer, 2020-001038-36), and inactivated virus formulations (Sinopharm, ChiCTR2000031809; Sinovac (NCT04352608) (17). While the advancement of these vaccine candidates into clinical testing is promising, it is imperative they meet stringent endpoints for safety (18). Preclinical studies of multiple experimental SARS-CoV vaccines have reported a Th2-type immunopathology in the lungs of vaccinated mice following viral challenge, suggesting hypersensitization of the immune response against certain viral proteins (19-22). Similarly, a modified vaccinia virus Ankara vector expressing the SARS-CoV S protein induced significant hepatitis in immunized ferrets (23). These data suggest that

candidate coronavirus vaccines that limit the inclusion of whole viral proteins may have more beneficial safety profiles.

The SARS-CoV-2 genome encodes for 10 unique protein products: 4 structural proteins (surface glycoprotein (S), envelope (E), membrane (M), nucleocapsid (N)); 5 non-structural proteins (open reading frame (ORF)3a, ORF6, ORF7a, ORF8, ORF10); and 1 non-structural polyprotein (ORF1ab) (**Figure 1A, B**) (24). It is currently unknown which epitopes in the SARS-CoV-2 proteome are recognized by the human immune system, although studies of SARS-CoV immune responses suggest that both cellular and humoral responses against structural proteins mediate protection against disease (19, 22, 25-27). It is likely that cellular immune responses against non-structural viral proteins also play a key role in orchestrating protective antiviral immunity (28-30). In lieu of biological data, immunoinformatic algorithms can be employed to predict peptide epitopes based on amino acid properties and known human leukocyte antigen (HLA) binding profiles (31-33). These computational approaches represent a validated methodology for rapidly identifying potential T cell and B cell epitopes for exploratory peptide-based vaccine development and have been recently used to identify target epitopes for MERS-CoV (34) and SARS-CoV-2, although many of these reports focus solely on structural proteins (35-38).

Herein, we employed a comprehensive immunoinformatics approach to identify putative T cell and B cell epitopes across the entire SARS-CoV-2 proteome (**Figure 1C**). We independently identified peptides from each viral protein that were restricted to either HLA class I or HLA class II molecules across a subset of the most common HLA alleles in the global population. By filtering this list of peptides on the basis of predicted binding affinity, antigenicity, and promiscuity, we produced 5 HLA class I-restricted and 36 HLA class II-restricted peptides as leading candidates for further study. We also evaluated linear and structural B cell epitopes in the SARS-CoV-2 spike protein, with six antigenic regions identified as potential sites for antibody binding. These selected peptides may serve as initial candidates in the rational and accelerated design of a peptide-based vaccine against SARS-CoV-2.

Methods

Comparison of genome sequences from SARS-CoV-2 isolates

Genomic sequences for reported SARS-CoV-2 isolates were identified and retrieved from the Virus Pathogen Resource (ViPR) database on February 27, 2020 (https://www.viprbrc.org/brc/home.spg?decorator=corona_ncov). Sequences that did not cover the complete viral genome (~29,900 nucleotides) were excluded from further analysis. Remaining sequences were aligned using the Clustal Omega program (version 1.2.4) from the European Bioinformatics Institute (39) and compared against the first reported genome sequence for SARS-CoV-2 (Wuhan-Hu-1; taxonomy

ID: 2697049) (1). Sequences from Wuhan-Hu-1 viral proteins were determined to be representative of those from all viral isolates and were subsequently used for epitope prediction analyses.

Prediction of SARS-CoV-2 T cell epitopes

Prediction of HLA class I and class II peptide epitopes was carried out with the 10 protein sequences reported for the Wuhan-Hu-1 isolate: E (GenBank accession: QHD43418), M (QHD43419), N (QHD43423), S (QHD43416), ORF3a (QHD43417), ORF6 (QHD43420), ORF7a (QHD43421), ORF8 (QHD43422), ORF10 (QHI42199), ORF1ab (QHD43415).

For $CD8^+$ T cell epitope prediction, NetCTL 1.2 (Immune Epitope Database) was initially used to evaluate the binding of nonameric peptides derived from each viral protein to the most common HLA class I supertypes present among the human population (40, 41). HLA class I molecules preferentially bind 9-mer peptides, and most algorithm training datasets have been based on peptides of this length. The weight placed on C-terminal cleavage and antigen transport efficiency was 0.15 and 0.05, respectively. The antigenic score threshold was 0.75. Peptides with scores above this threshold were subsequently analyzed on the NetMHCpan 4.0 server (Technical University of Denmark) to predict binding affinity and percentile rank across representative alleles of each major HLA class I supertype (HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-B*07:02, HLA-B*08:01, HLA-B*27:05, HLA-B*40:01, HLA-B*58:01, HLA-B*15:01), which collectively cover the majority of class I alleles present in the human population (42-44). Thresholds for defining binding strength were set at 0.5% and 2.0% for strong and weak binders, respectively.

For $CD4^+$ T cell epitope prediction, NetMHCIIpan 3.2 server (Technical University of Denmark) was used for predicting the binding affinity and percentile rank of 15-mer peptides derived from each viral protein across a reference panel of 27 HLA class II molecules (33, 45). Thresholds for defining binding strength were set at 2% and 10% for strong and weak binders, respectively.

HLA class I and class II peptides with high predicted binding affinities ($\leq$ 500 nM), high percentile ranks ($\leq$ 0.5% for class I; $\leq$ 2% for class II), and broad HLA coverage ($\geq$ 3 alleles) were independently analyzed on the VaxiJen 2.0 server (Edward Jenner Institute) (46, 47) using a conservative score threshold (0.7) to predict antigenicity.

Molecular docking of HLA class I peptides

Docking simulations of 5 HLA class I-restricted SARS-CoV-2 peptides with high antigenicity scores and a commonly shared predicted HLA molecule (HLA-DRB1*15:01) were performed using the GalaxyPepDock server (Seoul National University Laboratory of Computational Biology) (48). The structure of HLA-DRB1*15:01 was accessed from the Protein Data Bank as a co-crystallized structure of the HLA molecule with a nonameric SARS-CoV peptide (PDB ID: 3C9N) (49). The bound nonamer

peptide was removed from the structure using Chimera 1.14 (University of California-San Francisco) (50) prior to running simulations. Ten models of each peptide-HLA complex were generated on the basis of minimized energy scores, and the top model for each complex was selected for comparative analysis.

Prediction and structural modeling of SARS-CoV-2 B cell epitopes

Linear B cell epitope predictions were performed on the three exposed SARS-CoV-2 structural proteins: S (GenBank accession: QHD43416), M (QHD43419), and E (QHD43418) using the BepiPred 1.0 algorithm (51). Epitope probability scores were calculated for each amino acid residue using a threshold of 0.35 (corresponding to > 0.75 specificity and sensitivity below 0.5), and only epitopes $\geq$ 5 amino acid residues in length were further analyzed. The structure of the SARS-CoV-2 S protein was accessed from the Protein Data Bank (PDB ID: 6VSB) (52). Discontinuous (i.e., structural) B cell epitope predictions for the S protein structure were carried out using DiscoTope 1.1 (53) with a score threshold greater than – 7.7 (corresponding to > 0.75 specificity and sensitivity below 0.5). The main protein structure was modeled in PyMOL (Schrödinger, LLC), with predicted B cell epitopes identified by both BepiPred 1.0 and DiscoTope 1.1 highlighted as spheres.

Results

Genetic similarity of SARS-CoV-2 isolates

The primary goal of our study was to identify peptide epitopes that would be broadly applicable in vaccine development efforts against SARS-CoV-2. We identified 64 point mutations and 4 deletions across the genomes of 44 clinical isolates, with all deletions and the majority of mutations (n=45) occurring in the ORF1ab polyprotein (**Supp. Figure S1**). Single-point mutations were also found in the S protein (n=5), N protein (n=5), ORF8 protein (n=3), ORF3a protein (n=2), ORF10 protein (n=2), E protein (n=1), and M protein (n=1). Despite the genetic diversity introduced by these events (**Figure 1D**), matrix analysis determined that > 99% sequence identity was maintained across all viral genomes. Based on these findings and for study feasibility, the genome from the original virus isolate (Wuhan-Hu-1; GenBank: MN908947) was selected as the consensus sequence for all further analyses.

Prediction of $CD8^+$ T cell epitopes in the SARS-CoV-2 proteome

We next identified potential $CD8^+$ T cell epitopes from all proteins in the SARS-CoV-2 proteome. Using the NetCTL 1.2 predictive algorithm, we analyzed the complete amino acid sequence of each viral protein to generate sets of 9-mer peptides predicted to be recognized across at least one of the major HLA class I supertypes (**Figure 2A, Supp. Figure S2**). This approach yielded a significant number of potential epitopes from each viral protein (ORF10: 9, ORF6: 17, ORF8: 23, E: 25, ORF7: 39, N: 80, M: 87, ORF3a: 87, S: 321, ORF1ab: 2814), with the number directly related to the size of the parent

protein. We used the NetMHCpan 4.0 server to further refine the list of potential $CD8^+$ T cell epitopes by predicting binding affinity across representative HLA class I alleles (see Methods) and assigning percentile scores to quantify binding propensity. Peptides with percentile rank scores $\leq$ 0.5% (i.e., strong binders) were filtered using a 500 nM threshold for binding affinity to further delineate 740 candidate HLA class I epitopes from the viral proteome (54). For feasibility reasons, we refined our selection to 83 candidate epitopes by excluding peptides predicted to bind only one HLA molecule (**Supp. Table S1**). The resultant peptides were enriched for predicted binders to HLA-B molecules (HLA-B*15:01=50; HLA-B*58:01=32; HLA-B*08:01=31) (**Figure 2B**). A final round of selection on the basis of HLA promiscuity (i.e., predicted binding to $\geq$ 3 HLA molecules) and predicted antigenicity scoring using the VaxiJen 2.0 server produced a subset of five candidate peptides (four ORF1ab, one S protein) as potential targets for vaccine development (**Table 1**) with the hypothesis that increased HLA binding promiscuity meant broader population base coverage by those peptides. These peptides were predicted to provide 74% global population coverage and had higher predicted binding affinities for HLA-B molecules (B*08:01=42.6 nM; B*15:01=67.7 nM; B*58:01=110.3 nM) compared to HLA-A molecules (A*01:01=238.6 nM; A*24:02=142.9 nM), with the exception of one ORF1ab-derived peptide (MMISAGFSL) that was predicted to bind HLA-A*02:01 with high affinity (IC_{50}= 6.9 nM) (**Figure 2C**).

Prediction of $CD4^+$ T cell epitopes in the SARS-CoV-2 proteome

We also sought to identify potential HLA class II peptides from SARS-CoV-2, as the stimulation of $CD4^+$ T-helper cells is critical for robust vaccine-induced adaptive immune responses. Using the NetMHCIIpan 3.2 server, we identified 801 candidate HLA class II peptides from the viral proteome predicted to have high binding affinity ($\leq$ 500 nM) and percentile rank scores $\leq$ 2% across a reference panel of HLA molecules covering > 97% of the population (33, 45). Similar to HLA class I epitope predictions, the number of class II epitopes identified for each viral protein (ORF10: 4, E protein: 7, ORF7: 8, ORF8: 10, ORF6: 14, N: 15, M: 29, ORF3a: 31, S: 96, ORF1ab: 587) was largely proportional to protein size. After excluding peptides predicted to bind to only a single HLA molecule in our panel, we refined our selection to 211 peptides (**Supp. Table S2**), which were enriched for binding to HLA-DRB1 molecules (n=142) (**Figure 2D**). Filtering on HLA promiscuity and predicted antigenicity scores yielded a subset of 36 peptides (24 ORF1ab, 5 S protein, 2 M protein, 2 ORF7, 1 ORF3a, 1 ORF6, 1 ORF8) as $CD4^+$ T cell epitopes for further study (**Table 1**). These peptides were predicted to collectively provide 99% population coverage and have significantly higher average binding affinities for HLA-DR alleles (DRB1=56.4 nM; DRB3=50.9 nM; DRB4=70.1 nM; DRB5=18 nM) compared to HLA-DP (155.9 nM) or HLA-DQ (238.6 nM) molecules (**Figure 2E**).

*Characterization of HLA class I peptide docking with HLA-B*15:01*

The five candidate HLA class I peptides identified by our computational approach were predicted to provide coverage across six HLA alleles (A*01:01, A*02:01, A*24:02, B*08:01, B*15:01, B*58:01). The peptide FAMQMAYRF was the only candidate predicted to bind to A*24:02 molecules, whereas MMISAGFSL was predicted to uniquely bind A*02:01 and B*08:01 molecules. Four of the five peptides were predicted to bind A*01:01 and B*58:01 molecules, but all were predicted to bind with relatively high affinity (average IC_{50} = 67.7 nM) to HLA-B*15:01. Therefore, we performed molecular docking studies of each peptide with the molecular structure of HLA-B*15:01 (PDB: 3C9N).

All peptides were predicted to bind within the peptide binding groove, forming hydrogen bond contacts with numerous amino acid side chains (**Figure 3A**). The binding motif for HLA-B*15:01 is highly selective for residues at the P2 and P9 anchor positions, with a preference for bulky hydrophobic amino acids at the C-terminus (**Figure 3B**) (55). All candidate peptides possessed terminal residues (Phe, Tyr, Leu) that fit into the hydrophobic binding pocket of the HLA groove, further supporting that these peptides should be strong binders of HLA-B*15:01 and promising candidates for vaccine development studies.

Prediction of B cell epitopes in SARS-CoV-2 proteins

An effective vaccine should stimulate both cellular and humoral immune responses against the target pathogen; therefore, we also sought to identify potential B cell epitopes from SARS-CoV-2 proteins. We limited our analysis to the primary structural proteins exposed on the virus capsid (S, N, M, and E), as these are the most accessible antigens for engaging B cell receptors. Using the Bepipred 1.0 algorithm, we identified 26 potential linear B cell epitopes in the S protein, 14 potential epitopes in the N protein, and 3 potential epitopes in the M protein (**Table 2**). No epitopes were identified in the E protein. Studies have previously shown the S protein to be the predominant target of neutralizing antibodies against coronaviruses (56, 57), and, as our findings indicate this to likely be the case for SARS-CoV-2, we focused all subsequent analyses on the S protein. While the N protein is also a major target of the antibody response (58), it is unlikely these antibodies have any neutralizing activity based on the viral structure. As epitope conformation can significantly influence recognition by antibodies, we also employed DiscoTope 1.1 to identify discontinuous B cell epitopes in the protein structure. Our analysis identified 14 potential structural epitopes in the S protein (7 in the S1 domain, 7 in the S2 domain), with six regions having significant overlap with our predicted linear epitopes (**Table 2**). Antigenic regions identified in both analyses were modeled using the recently published structure of the SARS-CoV-2 S protein (52) to examine their accessibility for antibody binding. Epitopes in the S2 domain (P792-D796; Y1138-D1146) were clustered near the base of the spike protein, whereas regions in the S1 domain (D405-D428; N440-N450; G496-P507; D568-T573) were exposed on the protein surface (**Figure 4**).

Discussion

In the face of the COVID-19 pandemic, it is imperative that safe and effective vaccines be rapidly developed in order to induce widespread herd immunity in the population and prevent the continued spread of SARS-CoV-2. Our study identified probable peptide targets of both cellular and humoral immune responses against SARS-CoV-2 using computational methodologies to investigate the entire viral proteome *a priori*. Studies such as these are paramount during the early stages of pandemic vaccine development given the relative scarcity of biological data available on the viral immune response, and we employed an approach that allowed us to systematically refine our predictions using increasingly stringent criteria to select a subset of the most promising epitopes for further study. The data we have curated could inform the design of a candidate peptide-based vaccine or diagnostic against SARS-CoV-2.

As selective pressures are known to introduce viral mutations that promote fitness and can lead to evasion of immune responses (59, 60), we first sought to investigate the genetic similarity of all reported SARS-CoV-2 clinical isolates and identify a consensus sequence for use in our epitope prediction studies. We identified 68 mutations/deletions across the 44 genomes of clinical isolates reported as of 27 February 2020. Despite these variations, the viral genomic identity was > 99% conserved across all isolates. As the protein coding sequences were largely conserved, the genome of the original virus isolate (Wuhan-Hu-1) was deemed a representative consensus sequence for analysis of the SARS-CoV-2 proteome.

$CD4^+$ and $CD8^+$ T cell responses will likely be directed against both structural and non-structural proteins during antiviral immune responses, as all viral proteins are accessible for processing and presentation on the HLA molecules of infected cells. Therefore, we sought to identify T cell epitopes across the entire viral proteome. Our analysis identified 83 potential $CD8^+$ T cell epitopes (**Supp. Table S1**) and 211 potential $CD4^+$ T cell epitopes (**Supp. Table S2**), with stringent filtering for more promiscuous peptides with high predicted antigenicity yielding a subset of 5 $CD8^+$ T cell epitopes and 36 $CD4^+$ T cell epitopes (**Table 1**) as potential targets for vaccine development. A single study by Grifoni and colleagues has recently reported the computational identification of 241 $CD4^+$ T cell epitopes from SARS-CoV-2 (35), and 22 peptides from our analysis shared sequence homology or were nested within peptides identified in their study. Moreover, seven peptides from this initial report were replicated in our final subset of HLA class II epitopes, supporting that these peptides may be promising vaccine targets.

An increasing number of studies have employed predictive algorithms to identify potential HLA class I epitopes for SARS-CoV-2, although relatively few have comprehensively analyzed the entire viral proteome. A report from Feng *et al.* recently outlined the identification of 499 potential class I epitopes in the main structural proteins from SARS-CoV-2 but did not consider any non-structural proteins (38). Grifoni and colleagues conducted a more rigorous analysis, identifying 628 unique $CD8^+$ T cell epitopes

across all SARS-CoV-2 proteins but focusing their analyses solely on peptides with sequence homology to known SARS-CoV epitopes (35). Our approach initially identified ~ 3,500 potential $CD8^+$ T cell epitopes across all viral proteins, which we refined to a subset of 5 peptides (**Table 1**). One peptide derived from ORF1ab (MMISAGFSL) was predicted to bind HLA-A*02:01 with high affinity (IC_{50}= 6.9 nM) (**Figure 2C**). Given the prevalence of this allele in the American and European populations (25-60% frequency) (61), MMISAGFSL may represent a promising epitope capable of providing broad vaccine population coverage.

We also observed a notable enrichment of epitopes predicted to bind HLA-B molecules–particularly HLA-B*15:01–as we imposed more stringent selection criteria (**Figure 2B**). All five peptides identified by our approach were predicted to be relatively strong binders for this allele (IC_{50} = 67.7 nM), with molecular docking simulations illustrating strong contacts with amino acid residues in the peptide binding groove (**Figure 3 A, B**). A recent computational study identified another HLA-B allele (B*15:03) as having a high capacity for presenting epitopes from SARS-CoV-2 that were conserved among other pathogenic coronaviruses (62). These data collectively suggest the HLA-B locus may be significantly associated with the immune response to SARS-CoV-2 (and potentially other coronaviruses), with further biological studies warranted to determine the true role of host genetics in SARS-CoV-2 immunology.

Lastly, we analyzed the primary structural proteins of SARS-CoV-2 (S, N, M, E proteins) for potential B cell epitopes, as an ideal vaccine would be designed to stimulate both cellular and humoral immunity. Our analysis identified potential linear B cell epitopes in all proteins except for the E protein (**Table 2**). The greatest number of epitopes were predicted in the surface-exposed S protein (n=26), but a significant number of epitopes were also predicted for the N protein (n=14). This is not surprising, as previous reports identified the N protein as a significant target of the humoral response to SARS-CoV (63, 64). As the S protein is the predominant surface protein and has been the primary target of neutralizing antibody responses against other coronaviruses (56, 57), we elected to focus our subsequent analyses solely on antigenic regions in the S protein. We identified 14 potential structural epitopes in the S protein structure and referenced against our linear epitope predictions to identify six regions that were independently identified by both analyses (**Table 2, Figure 4**). Feng *et al*. recently reported the computational identification of 19 surface epitopes in the S protein using Bepipred and the Kolaskar method (38), four of which had significant sequence overlap with the regions identified by our analyses.

To further evaluate the potential of these six antigenic regions as targets for antibody binding, we modeled their surface accessibility on the crystal structure of the SARS-Cov-2 spike protein (52). Four regions in the S1 domain (D405-D428; N440-N450; G496-P507; D568-T573) were solvent exposed (**Figure 4 A, B**), with minimal steric hindrance for antibody accessibility. The S1 domain contains the residues (N331-V524) important for virus binding to angiotensin converting enzyme 2 (ACE2) on the cell

surface (65), and studies have shown that antibodies with potent neutralizing activity against SARS-CoV target this domain (66-68). Indeed, three of the four S1 epitopes identified in our analyses are located in the ACE2-binding region, supporting their potential utility in vaccine development against SARS-CoV-2. Two regions were identified in the S2 "stalk" domain of the S protein (**Figure 4 A, C**). While Y1138-D1146 is located at the base of the S protein and likely inaccessible to antibodies, P792-D796 is on the outer face of the protein and has been previously identified as part of a larger B cell epitope that is conserved with SARS-CoV (35). As SARS-CoV S2-specific antibodies have previously been shown to possess antiviral activity (66), it is interesting to speculate whether a strategy similar to targeting the influenza hemagglutinin protein stalk could be employed for developing a broadly reactive coronavirus vaccine.

Our study possessed several strengths and limitations. Rather than restricting our analyses of HLA class I and class II epitopes to specific proteins based on prior studies of SARS-CoV immunology, we investigated the complete proteome of SARS-CoV-2 using an unbiased approach. Furthermore, we employed a multi-tiered strategy for identifying putative B cell and T cell epitopes from all viral proteins studied. Our initial analyses were performed with liberal thresholds for epitope identification, and at each additional step, we imposed more stringent selection criteria to filter these peptides to a subset of B cell and T cell epitopes for further study. Nevertheless, the results of this study are derived purely from computational methods, and it should be noted that computational algorithms can fail to capture a significant number of antigenic peptides (69). Experimental validation with biological samples will ultimately be needed.

During the early stages of a pandemic, access to sufficient biological samples may be extremely limited, so we must continue to utilize methodologies—such as computational predictive algorithms—that allow us to explore the epitope landscape for experimental vaccine development. Our approach in this study allowed us to identify and refine a manageable subset of T cell and B cell epitopes for further testing as components of a SARS-CoV-2 vaccine. Based on our results, our proposed SARS-CoV-2 vaccine formulation could contain the following: 1) one or more B cell peptide epitopes from the S protein to generate protective neutralizing antibodies; and 2) multiple HLA class I and class II-derived peptides from other viral proteins to stimulate robust $CD8^+$ and $CD4^+$ T cell responses. Based on global allele frequencies, these class I and class II peptides would be expected to collectively provide 74% and 99% population coverage, respectively. While such a vaccine could be readily formulated as a synthetic polypeptide or an adjuvanted peptide mixture, these strategies may not retain the epitope structural features necessary to induce a robust antibody response. Recombinant nanoparticles and assembly into VLPs represent promising alternative vaccine platforms, as they have been extensively used for the controlled display and delivery of peptide-based vaccine components (70-73). By omitting whole viral

proteins from the vaccine formulation, a peptide-based SARS-CoV-2 vaccine should have a well-tolerated safety profile and avoid the adverse events previously observed with experimental SARS-CoV vaccines (19-22).

In summary, we have identified 41 potential T cell epitopes (5 HLA class I, 36 HLA class II) and 6 potential B cell epitopes from across the SARS-CoV-2 proteome that are predicted to have broad population coverage and could serve as the basis for designing investigational peptide-based vaccines. Further study on the biological relevance and immunogenicity of these peptides is warranted in an effort to develop a safe and effective vaccine to combat the SARS-CoV-2 pandemic.

Acknowledgments

The authors would like to thank Caroline L. Vitse for editorial assistance with this manuscript. The research presented here was not supported by any specific funding source.

Conflicts of Interest

Dr. Poland is the chair of a Safety Evaluation Committee for novel investigational vaccine trials being conducted by Merck Research Laboratories. Dr. Poland offers consultative advice on vaccine development to Merck & Co. Inc., Avianax, Adjuvance, Valneva, Medicago, Sanofi Pasteur, GlaxoSmithKline, and Emergent Biosolutions. Drs. Poland and Ovsyannikova hold three patents related to measles and vaccinia peptide research. Dr. Kennedy holds a patent on vaccinia peptide research. Dr. Kennedy has received funding from Merck Research Laboratories to study waning immunity to measles and mumps after immunization with the MMR-II® vaccine. All other authors declare no competing financial interests. This research has been reviewed by the Mayo Clinic Conflict of Interest Review Board and was conducted in compliance with Mayo Clinic Conflict of Interest policies.

388 References

389

390 1. Wu F, *et al.* (2020) A new coronavirus associated with human respiratory disease in China.
391 *Nature* 579(7798):265-269.
392 2. Chan JF, *et al.* (2020) A familial cluster of pneumonia associated with the 2019 novel
393 coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*
394 395(10223):514-523.
395 3. Cucinotta D & Vanelli M (2020) WHO Declares COVID-19 a Pandemic. *Acta Biomed*
396 91(1):157-160.
397 4. Chen N, *et al.* (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel
398 coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 395(10223):507-513.
399 5. Wang D, *et al.* (2020) Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel
400 Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA*.[ePub ahead of print] doi:
401 10.1001/jama.2020.1585
402 6. Huang C, *et al.* (2020) Clinical features of patients infected with 2019 novel coronavirus in
403 Wuhan, China. *Lancet* 395(10223):497-506.
404 7. World Health Organization. (2020) Coronavirus disease (COVID-19) Situation Report - 113.
405 https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200512-covid-19-
406 sitrep-113.pdf?sfvrsn=feac3b6d_2. May 12, 2020.
407 8. Drosten C, *et al.* (2003) Identification of a novel coronavirus in patients with severe acute
408 respiratory syndrome. *N Engl J Med* 348(20):1967-1976.
409 9. Ksiazek TG, *et al.* (2003) A novel coronavirus associated with severe acute respiratory syndrome.
410 *N Engl J Med* 348(20):1953-1966.
411 10. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, & Fouchier RA (2012) Isolation of a
412 novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 367(19):1814-
413 1820.
414 11. Li W, *et al.* (2005) Bats are natural reservoirs of SARS-like coronaviruses. *Science*
415 310(5748):676-679.
416 12. Zhou P, *et al.* (2020) A pneumonia outbreak associated with a new coronavirus of probable bat
417 origin. *Nature* 579(7798):270-273.
418 13. Memish ZA, *et al.* (2013) Middle East respiratory syndrome coronavirus in bats, Saudi Arabia.
419 *Emerg Infect Dis* 19(11):1819-1823.
420 14. Haagmans BL, *et al.* (2014) Middle East respiratory syndrome coronavirus in dromedary camels:
421 an outbreak investigation. *Lancet Infect Dis* 14(2):140-145.
422 15. Walls AC, *et al.* (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike
423 Glycoprotein. *Cell* 181(2):281-292.
424 16. Weston S & Frieman MB (2020) COVID-19: Knowns, Unknowns, and Questions. *mSphere* 5(2).
425 17. World Health Organization (2020) Draft landscape of COVID-19 candidate vaccines.
426 https://www.who.int/who-documents-detail/draft-landscape-of-covid-19-candidate-vaccines. May
427 5, 2020.
428 18. Poland GA (2020) Tortoises, hares, and vaccines: A cautionary note for SARS-CoV-2 vaccine
429 development. *Vaccine* [ePub ahead of print]. doi: 10.1016/j.vaccine.2020.04.073
430 19. Tseng CT, *et al.* (2012) Immunization with SARS coronavirus vaccines leads to pulmonary
431 immunopathology on challenge with the SARS virus. *PLoS One* 7(4):e35421.
432 20. Deming D, *et al.* (2006) Vaccine efficacy in senescent mice challenged with recombinant SARS-
433 CoV bearing epidemic and zoonotic spike variants. *PLoS Med* 3(12):e525.
434 21. Yasui F, *et al.* (2008) Prior immunization with severe acute respiratory syndrome (SARS)-
435 associated coronavirus (SARS-CoV) nucleocapsid protein causes severe pneumonia in mice
436 infected with SARS-CoV. *J Immunol* 181(9):6337-6348.

22. Bolles M, *et al.* (2011) A double-inactivated severe acute respiratory syndrome coronavirus vaccine provides incomplete protection in mice and induces increased eosinophilic proinflammatory pulmonary response upon challenge. *J Virol* 85(23):12201-12215.
23. Weingartl H, *et al.* (2004) Immunization with modified vaccinia virus Ankara-based recombinant vaccine against severe acute respiratory syndrome is associated with enhanced hepatitis in ferrets. *J Virol* 78(22):12672-12676.
24. Lu R, *et al.* (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395(10224):565-574.
25. Channappanavar R, Fett C, Zhao J, Meyerholz DK, & Perlman S (2014) Virus-specific memory CD8 T cells provide substantial protection from lethal severe acute respiratory syndrome coronavirus infection. *J Virol* 88(19):11034-11044.
26. Ng OW, *et al.* (2016) Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. *Vaccine* 34(17):2008-2014.
27. Zhao J, *et al.* (2016) Airway Memory CD4(+) T Cells Mediate Protective Immunity against Emerging Respiratory Coronaviruses. *Immunity* 44(6):1379-1391.
28. Lorente E, *et al.* (2016) Structural and Nonstructural Viral Proteins Are Targets of T-Helper Immune Response against Human Respiratory Syncytial Virus. *Mol Cell Proteomics* 15(6):2141-2151.
29. Ip PP, *et al.* (2014) Alphavirus-based vaccines encoding nonstructural proteins of hepatitis C virus induce robust and protective T-cell responses. *Mol Ther* 22(4):881-890.
30. Henriques HR, *et al.* (2013) Targeting the non-structural protein 1 from dengue virus to a dendritic cell population confers protective immunity to lethal virus challenge. *PLoS Negl Trop Dis* 7(7):e2330.
31. Tomar N & De RK (2014) Immunoinformatics: a brief review. *Methods Mol Biol* 1184:23-55.
32. Backert L & Kohlbacher O (2015) Immunoinformatics and epitope prediction in the age of genomic medicine. *Genome Med* 7:119.
33. Jensen KK, *et al.* (2018) Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* 154(3):394-406.
34. Tahir Ul Qamar M, *et al.* (2019) Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. *J Transl Med* 17(1):362.
35. Grifoni A, *et al.* (2020) A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* 27(4):671-680 e672.
36. Fast E, Altman RB, & Chen B (2020) Potential T-cell and B-cell Epitopes of 2019-nCoV. *bioRxiv*:2020.2002.2019.955484.
37. Seema M (2020) T Cell Epitope-Based Vaccine Design for Pandemic Novel Coronavirus 2019-nCoV. doi: 10.26434/chemrxiv.12029523.v1 (April 3, 2020)
38. Feng Y, *et al.* (2020) Multi-epitope vaccine design using an immunoinformatics approach for 2019 novel coronavirus in China (SARS-CoV-2). *bioRxiv*:2020.2003.2003.962332.
39. Madeira F, *et al.* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47(W1):W636-W641.
40. Larsen MV, *et al.* (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 8:424.
41. Larsen MV, *et al.* (2005) An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur J Immunol* 35(8):2295-2303.
42. Hoof I, *et al.* (2009) NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 61(1):1-13.
43. Jurtz V, *et al.* (2017) NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* 199(9):3360-3368.

44. Nielsen M & Andreatta M (2016) NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med* 8(1):33.
45. Greenbaum J, *et al.* (2011) Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* 63(6):325-335.
46. Doytchinova IA & Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 8:4.
47. Doytchinova IA & Flower DR (2007) Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine* 25(5):856-866.
48. Ko J, Park H, Heo L, & Seok C (2012) GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res* 40(Web Server issue):W294-297.
49. Roder G, Kristensen O, Kastrup JS, Buus S, & Gajhede M (2008) Structure of a SARS coronavirus-derived peptide bound to the human major histocompatibility complex class I molecule HLA-B*1501. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64(Pt 6):459-462.
50. Pettersen EF, *et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605-1612.
51. Larsen JE, Lund O, & Nielsen M (2006) Improved method for predicting linear B-cell epitopes. *Immunome Res* 2:2.
52. Wrapp D, *et al.* (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367(6483):1260-1263.
53. Haste Andersen P, Nielsen M, & Lund O (2006) Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 15(11):2558-2567.
54. Sette A, *et al.* (1994) The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *Journal of Immunology* 153:5586-5592.
55. Roder G, *et al.* (2006) Crystal structures of two peptide-HLA-B*1501 complexes; structural characterization of the HLA-B62 supertype. *Acta Crystallogr D Biol Crystallogr* 62(Pt 11):1300-1310.
56. Okba NMA, *et al.* (2020) SARS-CoV-2 specific antibody responses in COVID-19 patients. *medRxiv*:2020.2003.2018.20038059.
57. Wang Q, *et al.* (2016) Immunodominant SARS Coronavirus Epitopes in Humans Elicited both Enhancing and Neutralizing Effects on Infection in Non-human Primates. *ACS Infect Dis* 2(5):361-376.
58. Zhang L, *et al.* (2020) Anti-SARS-CoV-2 virus antibody levels in convalescent plasma of six donors who have recovered from COVID-19. *Aging (Albany NY)* 12(8):6536-6542.
59. Doud MB, Hensley SE, & Bloom JD (2017) Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathog* 13(3):e1006271.
60. Keck ML, Wrensch F, Pierce BG, Baumert TF, & Foung SKH (2018) Mapping Determinants of Virus Neutralization and Viral Escape for Rational Design of a Hepatitis C Virus Vaccine. *Front Immunol* 9:1194.
61. Ellis JM, *et al.* (2000) Frequencies of HLA-A2 alleles in five U.S. population groups. Predominance of A*02011 and identification of HLA-A*0231. *Human Immunology* 61(3):334-340.
62. Nguyen A, *et al.* (2020) Human leukocyte antigen susceptibility map for SARS-CoV-2. *medRxiv*:2020.2003.2022.20040600.
63. Huang LR, *et al.* (2004) Evaluation of antibody responses against SARS coronaviral nucleocapsid or spike proteins by immunoblotting or ELISA. *J Med Virol* 73(3):338-346.
64. Qiu M, *et al.* (2005) Antibody responses to individual proteins of SARS coronavirus and their neutralization activities. *Microbes Infect* 7(5-6):882-889.

65. Tai W, *et al.* (2020) Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol*. [ePub ahead of print] doi: 10.1038/s41423-020-0400-4.
66. Zeng F, *et al.* (2006) Quantitative comparison of the efficiency of antibodies against S1 and S2 subunit of SARS coronavirus spike protein in virus neutralization and blocking of receptor binding: implications for the functional roles of S2 subunit. *FEBS Lett* 580(24):5612-5620.
67. Berry JD, *et al.* (2010) Neutralizing epitopes of the SARS-CoV S-protein cluster independent of repertoire, antigen structure or mAb technology. *MAbs* 2(1):53-66.
68. He Y, *et al.* (2006) Identification and characterization of novel neutralizing epitopes in the receptor-binding domain of SARS-CoV spike protein: revealing the critical antigenic determinants in inactivated SARS-CoV vaccine. *Vaccine* 24(26):5498-5508.
69. Johnson KL, Ovsyannikova IG, Mason CJ, Bergen HR, III, & Poland GA (2009) Discovery of naturally processed and HLA-presented class I peptides from vaccinia virus infection using mass spectrometry for vaccine development. *Vaccine* 28(1):38-47.
70. Zhang L, *et al.* (2017) Development of Autologous C5 Vaccine Nanoparticles to Reduce Intravascular Hemolysis in Vivo. *ACS Chem Biol* 12(2):539-547.
71. Brune KD, *et al.* (2016) Plug-and-Display: decoration of Virus-Like Particles via isopeptide bonds for modular immunization. *Sci Rep* 6:19234.
72. Zhai L, *et al.* (2017) A novel candidate HPV vaccine: MS2 phage VLP displaying a tandem HPV L2 peptide offers similar protection in mice to Gardasil-9. *Antiviral Res* 147:116-123.
73. McCarthy DP, Hunter ZN, Chackerian B, Shea LD, & Miller SD (2014) Targeted immunomodulation using antigen-conjugated nanoparticles. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* 6(3):298-315.
74. Zhang Y & Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. *Proteins* 57(4):702-710.
75. Lee H, Heo L, Lee MS, & Seok C (2015) GalaxyPepDock: a protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Res* 43(W1):W431-435.

576 Figure legends

577 Figure 1. (A) Diagram of SARS-CoV-2 virion structure with the major structural proteins (S, M, N, and
578 E) highlighted. (B) Cartoon representation of the SARS-CoV-2 genome with the 10 major protein-coding
579 regions annotated. The box diagrams are proportional to the protein size. (C) Diagram of peptide
580 identification workflow illustrating the algorithms used (33, 40-43, 45-47, 51, 53) and filtering criterion
581 applied to refine peptide selection. (D) Cladogram illustrating the genetic relationship of SARS-CoV-2
582 isolates. The original viral isolate and consensus sequence (Wuhan-Hu-1) is highlighted in red.

583

584 Figure 2. Immunogenicity scoring of peptides in the SARS-CoV-2 proteome with predicted HLA class I
585 and II coverage and binding affinities. (A) Plots illustrating the NetCTL score for each sequential peptide
586 across the entire amino acid sequence for each SARS-CoV-2 protein. Scores presented are the highest
587 score identified across all HLA class I supertypes for each peptide. (B) Total number of predicted peptide
588 epitopes distributed across HLA class I alleles. (C) Average predicted binding affinities by HLA allele for
589 the top candidate class I peptides listed in Table 1. (D) Total number of predicted peptide epitopes
590 distributed across HLA class II alleles. (E) Average predicted binding affinities by HLA allele for the top
591 candidate class II peptides listed in Table 1.
592

593

594 Figure 3. Docking of top predicted HLA class I peptides with a shared HLA molecule. (A) Structural
595 docking model for each indicated peptide with the molecular structure of HLA-B*15:01 (PDB: 3C9N).
596 Individual panels represent top-down views of the peptide binding groove. (B) Binding motif for HLA-
597 B*15:01. (C) Template Modeling and Interaction Similarity scores for the selected peptide docking
598 models shown in panel A. (74, 75)

599

600

601 Figure 4. Modeling of predicted B cell epitopes on the crystal structure of the S glycoprotein. Predicted
602 structural epitopes in the S1 domain (A) and S2 domain (B) highlighted on the structure of the S
603 glycoprotein monomer (PDB: 6VSB). (C) Top predicted B cell epitopes identified by both Bepipred and
604 DiscoTope prediction algorithms highlighted on the trimeric structure of the S glycoprotein. Inset panels
605 show the S1 domain (upper) and S2 domain (lower). Predicted epitopes are highlighted as colored atoms
606 (green, blue, red) on the surface of the S protein (salmon).

607

608

609

610

611 Table 1. Top predicted HLA class I and class II T cell epitopes.

Protein	Peptide	Antigenicity Score	Predicted Alleles	Binding Affinity (nM)
		Class I		
S	FAMQMAYRF	1.0278	A*24:02	142.9
			B*15:01	123.9
			B*58:01	23.4
ORF1ab	LSFKELLVY	0.7234	A*01:01	371.8
			B*15:01	42.6
			B*58:01	35.7
ORF1ab	MMISAGFSL	1.0248	A*02:01	6.9
			B*08:01	367.6
			B*15:01	16.2
ORF1ab	MSNLGMPSY	0.9272	A*01:01	184.2
			B*15:01	74.1
			B*58:01	87.6
ORF1ab	STNVTIATY	0.7143	A*01:01	241.1
			B*15:01	81.9
			B*58:01	294.5
		Class II		
M	ASFRLFARTRSMWSF	0.7304	DRB1*01:01	19.2
			DRB1*07:01	30.9
			DRB1*08:02	53.5
			DRB1*09:01	49.9
			DRB1*11:01	12.2
			DRB5*01:01	16.3
			DPA1*02:01/DPB1*05:01	256.2
			DPA1*02:01 DPB1*14:01	387.3
M	LLQFAYANRNRFLYI	0.7387	DRB1*03:01	179.8
			DRB1*07:01	58.2
			DRB1*08:02	225.6
			DRB1*11:01	36.2
			DRB1*13:02	27.8
			DRB3*02:02	46.6
			DRB5*01:01	26.3
S	AAEIRASANLAATKM	0.7125	DRB1*08:02	101.3
			DRB1*13:02	23.0
			DRB3*02:02	52.7
			DQA1*01:02/DQB1*06:02	141.5
			DPA1*02:01/DPB1*14:01	327.4
S	ALQIPFAMQMAYRFN	1.0112	DRB1*09:01	52.9
			DRB1*12:01	159.5
			DRB1*15:01	50.3
S	PYRVVVLSFELLHAP	0.8161	DPA1*02:01/DPB1*01:01	79.6
			DPA1*01:03/DPB1*02:01	53.3
			DPA1*01:03/DPB1*04:01	77.1
			DPA1*03:01/DPB1*04:02	92.9
S	QPYRVVVLSFELLHA	0.9109	DPA1*02:01/DPB1*01:01	73.2
			DPA1*01:03/DPB1*02:01	50.2
			DPA1*01:03/DPB1*04:01	71.4
			DPA1*03:01/DPB1*04:02	90.1
			DPA1*02:01/DPB1*05:01	211.1

S	YQPYRVVVLSFELLH	0.9711	DPA1*02:01/DPB1*01:01	102.2
			DPA1*01:03/DPB1*04:01	93.0
			DPA1*03:01/DPB1*04:02	127.5
			DPA1*02:01/DPB1*05:01	299.3
ORF1ab	ANYIFWRNTNPIQLS	1.0311	DRB1*04:05	89.9
			DRB1*07:01	35.2
			DRB1*13:02	13.5
ORF1ab	FKWDLTAFGLVAEWF	0.8059	DQA1*05:01/DQB1*02:01	178.3
			DQA1*03:01/DQB1*03:02	425.3
			DQA1*04:01/DQB1*04:02	349.3
ORF1ab	HIQWMVMFTPLVPFW	0.7238	DQA1*01:01/DQB1*05:01	293.1
			DPA1*02:01/DPB1*01:01	116.3
			DPA1*01:03/DPB1*04:01	84.6
			DPA1*03:01/DPB1*04:02	135.4
ORF1ab	IINLVQMAPISAMVR	0.7682	DRB1*01:01	12.8
			DRB1*08:02	118.8
			DRB4*01:01	54.7
ORF1ab	INLVQMAPISAMVRM	0.9037	DRB1*12:01	176.9
			DRB4*01:01	57.1
			DQA1*01:02/DQB1*06:02	116.5
			DPA1*02:01/DPB1*14:01	398.6
ORF1ab	IVFMCVEYCPIFFIT	1.0267	DPA1*02:01/DPB1*01:01	116.2
			DPA1*01:03/DPB1*02:01	53.9
			DPA1*01:03/DPB1*04:01	70.9
			DPA1*03:01/DPB1*04:02	144.9
ORF1ab	IVTALRANSAVKLQN	0.7692	DRB1*08:02	115.9
			DRB1*13:02	9.4
			DRB3*02:02	19.5
			DPA1*02:01/DPB1*14:01	408.7
ORF1ab	KGRLIIRENNRVVIS	0.7821	DRB1*12:01	170.9
			DRB1*13:02	9.5
			DRB1*15:01	48.2
			DRB4*01:01	58.8
ORF1ab	KSAFYILPSIISNEK	0.7169	DRB1*01:01	9.3
			DRB1*04:01	49.3
			DRB1*04:05	47.5
			DRB1*08:02	96.3
ORF1ab	LIVTALRANSAVKLQ	0.7473	DRB1*01:01	8.8
			DRB1*07:01	39.2
			DRB4*01:01	78.6
			DQA1*01:02/DQB1*06:02	142.5
			DPA1*02:01/DPB1*14:01	368.3
ORF1ab	NLPFKLTCATTRQVV	1.1632	DRB1*07:01	35.9
			DRB1*09:01	58.6
			DRB5*01:01	23.9
ORF1ab	PASRELKVTFFPDLN	1.0155	DPA1*02:01/DPB1*01:01	76.9
			DPA1*01:03/DPB1*02:01	48.9
			DPA1*01:03/DPB1*04:01	64.3
			DPA1*03:01/DPB1*04:02	149.5
ORF1ab	PFAMGIIAMSAFAMM	0.9834	DRB1*01:01	12.3
			DRB1*09:01	57.6
			DQA1*05:01/DQB1*03:01	45.6
ORF1ab	QMNLKYAISAKNRAR	1.5044	DRB1*01:01	14.9

			DRB1*04:01	56.9
			DRB1*08:02	49.1
			DRB1*09:01	45.2
			DRB1*11:01	22.1
			DRB3*02:02	84.9
			DPA1*02:01/DPB1*14:01	158.3
ORF1ab	QQKLALGGSVAIKIT	1.2533	DRB1*01:01	12.6
			DRB1*07:01	23.4
			DRB1*09:01	32.3
			DQA1*05:01/DQB1*03:01	42.9
ORF1ab	RFKESPFELEDFIPM	1.2101	DPA1*02:01/DPB1*01:01	74.0
			DPA1*01:03/DPB1*02:01	65.9
			DPA1*01:03/DPB1*04:01	81.9
			DPA1*03:01/DPB1*04:02	130.6
ORF1ab	SAFAMMFVKHKHAFL	0.7305	DRB1*08:02	110.4
			DRB1*11:01	18.3
			DRB1*15:01	50.9
			DRB4*01:01	79.2
			DRB5*01:01	15.1
ORF1ab	SFLAHIQWMVMFTPL	0.8215	DPA1*02:01/DPB1*01:01	103.9
			DPA1*01:03/DPB1*02:01	47.8
			DPA1*01:03/DPB1*04:01	70.7
			DPA1*03:01/DPB1*04:02	140.6
ORF1ab	SIGFDYVYNPFMIDV	1.0823	DPA1*02:01/DPB1*01:01	108.9
			DPA1*01:03/DPB1*02:01	47.1
			DPA1*01:03/DPB1*04:01	81.9
			DPA1*03:01/DPB1*04:02	137.6
ORF1ab	TEETFKLSYGIATVR	0.8859	DRB1*01:01	8.7
			DRB1*07:01	21.8
			DRB1*09:01	25.9
ORF1ab	VLVQSTQWSLFFFLY	0.7309	DPA1*02:01/DPB1*01:01	77.0
			DPA1*01:03/DPB1*02:01	35.3
			DPA1*01:03/DPB1*04:01	42.3
			DPA1*03:01/DPB1*04:02	93.1
ORF1ab	VQSTQWSLFFFLYEN	0.7509	DPA1*02:01/DPB1*01:01	107.1
			DPA1*01:03/DPB1*02:01	49.9
			DPA1*03:01/DPB1*04:02	129.8
ORF1ab	WLIINLVQMAPISAM	0.9389	DRB1*12:01	130.6
			DRB4*01:01	65.9
			DQA1*01:02/DQB1*06:02	139.6
ORF1ab	YFNMVYMPASWVMRI	0.7244	DRB1*01:01	8.3
			DRB1*04:05	80.2
			DRB1*07:01	38.2
			DRB1*09:01	37.4
			DRB1*12:01	184.5
			DRB1*15:01	30.1
ORF3	KKRWQLALSKGVHFV	0.8172	DRB1*01:01	9.2
			DRB1*07:01	11.6
			DRB1*08:02	200.3
			DRB1*09:01	17.9
			DRB1*11:01	43.1
			DRB1*12:01	119.6
			DRB1*13:02	30.0

			DRB1*15:01	34.2
			DRB4*01:01	79.8
			DRB5*01:01	18.4
ORF6	MFHLVDFQVTIAEIL	1.0366	DQA1*05:01/DQB1*02:01	192.0
			DQA1*01:01/DQB1*05:01	292.1
			DPA1*02:01/DPB1*01:01	108.3
			DPA1*01:03/DPB1*04:01	100.7
ORF7	VKHVYQLRARSVSPK	1.0865	DRB1*01:01	14.3
			DRB1*08:02	150.6
			DRB1*11:01	38.3
			DRB4*01:01	86.6
ORF7	NKFALTCFSTQFAFA	1.1728	DPA1*02:01/DPB1*01:01	50.9
			DPA1*01:03/DPB1*02:01	29.1
			DPA1*01:03/DPB1*04:01	35.9
			DPA1*03:01/DPB1*04:02	80.2
			DPA1*02:01/DPB1*05:01	273.4
ORF8	SKWYIRVGARKSAPL	0.8829	DRB1*01:01	13.7
			DRB1*08:02	87.8
			DRB1*09:01	50.7
			DRB1*11:01	15.3
			DRB5*01:01	8.8

Table 2. Top predicted B cell epitopes.

Peptide	Residues	Bepipred Score[a]	DiscoTope Score[a]
DEVRNIAPGNTGKIADTNTKLPDD	405-428	0.715	-5.71
NLDSKVGGSYN	440-450	0.577	-5.77
GFNPTVGYNP	496-507	1.01	-5.73
DIADTT	568-573	0.853	-5.55
PPIKD	792-796	0.936	-3.28
VYDPLQPELDSF	1138-1149	0.747	-4.12

[a]Reported scores represent the average calculated across all amino acids.

Figure 1.

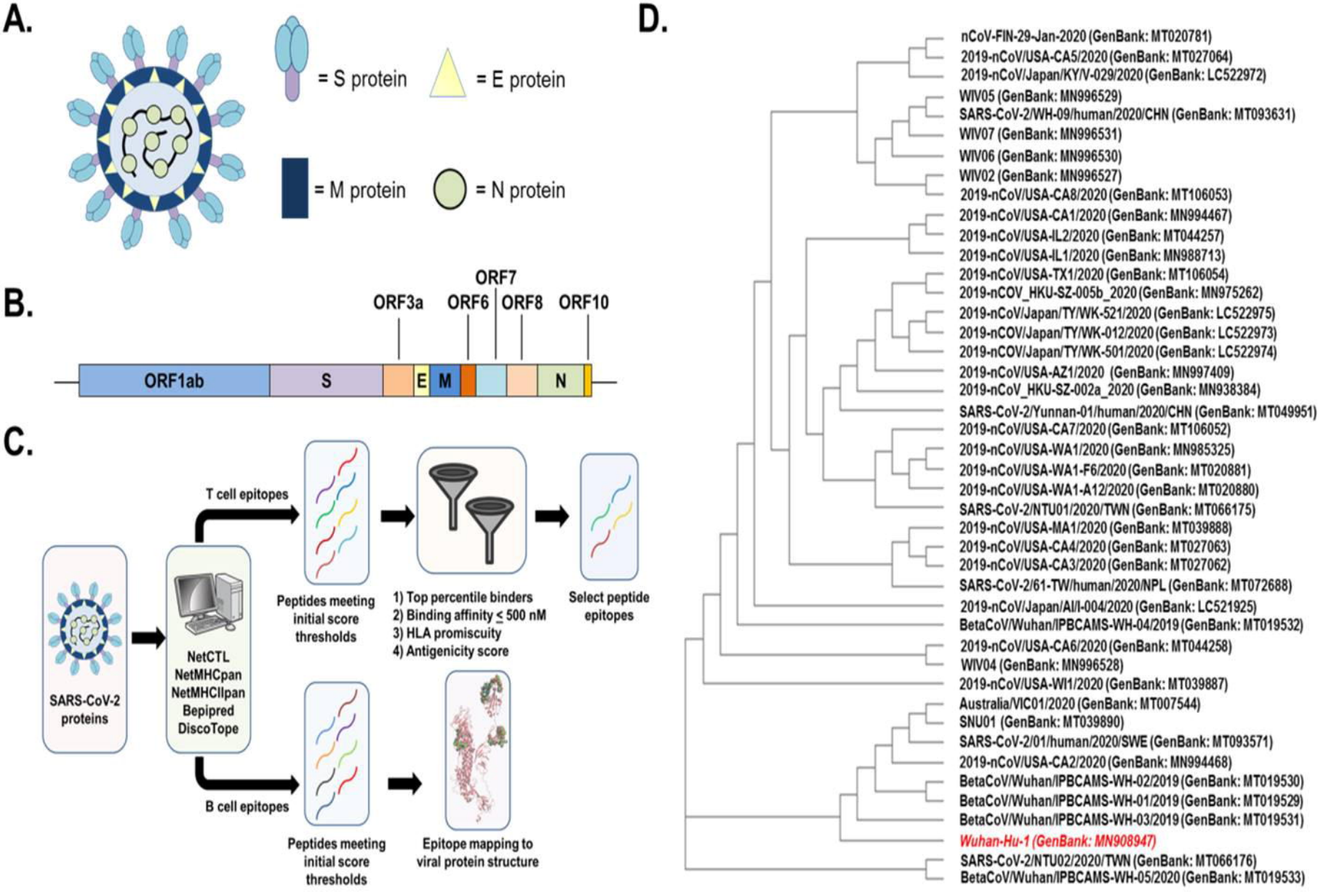

A.
= S protein
= E protein
= M protein
= N protein
B.
ORF7
ORF3a
ORF6
ORF8
ORF10
ORF1ab
S
E
M
N
C.
T cell epitopes
SARS-CoV-2 proteins
NetCTL
NetMHCpan
NetMHCIIpan
Bepipred
DiscoTope
Peptides meeting initial score thresholds
1) Top percentile binders
2) Binding affinity ≤ 500 nM
3) HLA promiscuity
4) Antigenicity score
Select peptide epitopes
B cell epitopes
Peptides meeting initial score thresholds
Epitope mapping to viral protein structure
D.
nCoV-FIN-29-Jan-2020 (GenBank: MT020781)
2019-nCoV/USA-CA5/2020 (GenBank: MT027064)
2019-nCoV/Japan/KY/V-029/2020 (GenBank: LC522972)
WIV05 (GenBank: MN996529)
SARS-CoV-2/WH-09/human/2020/CHN (GenBank: MT093631)
WIV07 (GenBank: MN996531)
WIV06 (GenBank: MN996530)
WIV02 (GenBank: MN996527)
2019-nCoV/USA-CA8/2020 (GenBank: MT106053)
2019-nCoV/USA-CA1/2020 (GenBank: MN994467)
2019-nCoV/USA-IL2/2020 (GenBank: MT044257)
2019-nCoV/USA-IL1/2020 (GenBank: MN988713)
2019-nCoV/USA-TX1/2020 (GenBank: MT106054)
2019-nCOV_HKU-SZ-005b_2020 (GenBank: MN975262)
2019-nCoV/Japan/TY/WK-521/2020 (GenBank: LC522975)
2019-nCOV/Japan/TY/WK-012/2020 (GenBank: LC522973)
2019-nCOV/Japan/TY/WK-501/2020 (GenBank: LC522974)
2019-nCoV/USA-AZ1/2020 (GenBank: MN997409)
2019-nCoV_HKU-SZ-002a_2020 (GenBank: MN938384)
SARS-CoV-2/Yunnan-01/human/2020/CHN (GenBank: MT049951)
2019-nCoV/USA-CA7/2020 (GenBank: MT106052)
2019-nCoV/USA-WA1/2020 (GenBank: MN985325)
2019-nCoV/USA-WA1-F6/2020 (GenBank: MT020881)
2019-nCoV/USA-WA1-A12/2020 (GenBank: MT020880)
SARS-CoV-2/NTU01/2020/TWN (GenBank: MT066175)
2019-nCoV/USA-MA1/2020 (GenBank: MT039888)
2019-nCoV/USA-CA4/2020 (GenBank: MT027063)
2019-nCoV/USA-CA3/2020 (GenBank: MT027062)
SARS-CoV-2/61-TW/human/2020/NPL (GenBank: MT072688)
2019-nCoV/Japan/AI/I-004/2020 (GenBank: LC521925)
BetaCoV/Wuhan/IPBCAMS-WH-04/2019 (GenBank: MT019532)
2019-nCoV/USA-CA6/2020 (GenBank: MT044258)
WIV04 (GenBank: MN996528)
2019-nCoV/USA-WI1/2020 (GenBank: MT039887)
Australia/VIC01/2020 (GenBank: MT007544)
SNU01 (GenBank: MT039890)
SARS-CoV-2/01/human/2020/SWE (GenBank: MT093571)
2019-nCoV/USA-CA2/2020 (GenBank: MN994468)
BetaCoV/Wuhan/IPBCAMS-WH-02/2019 (GenBank: MT019530)
BetaCoV/Wuhan/IPBCAMS-WH-01/2019 (GenBank: MT019529)
BetaCoV/Wuhan/IPBCAMS-WH-03/2019 (GenBank: MT019531)
Wuhan-Hu-1 (GenBank: MN908947)
SARS-CoV-2/NTU02/2020/TWN (GenBank: MT066176)
BetaCoV/Wuhan/IPBCAMS-WH-05/2020 (GenBank: MT019533)

Figure 2.

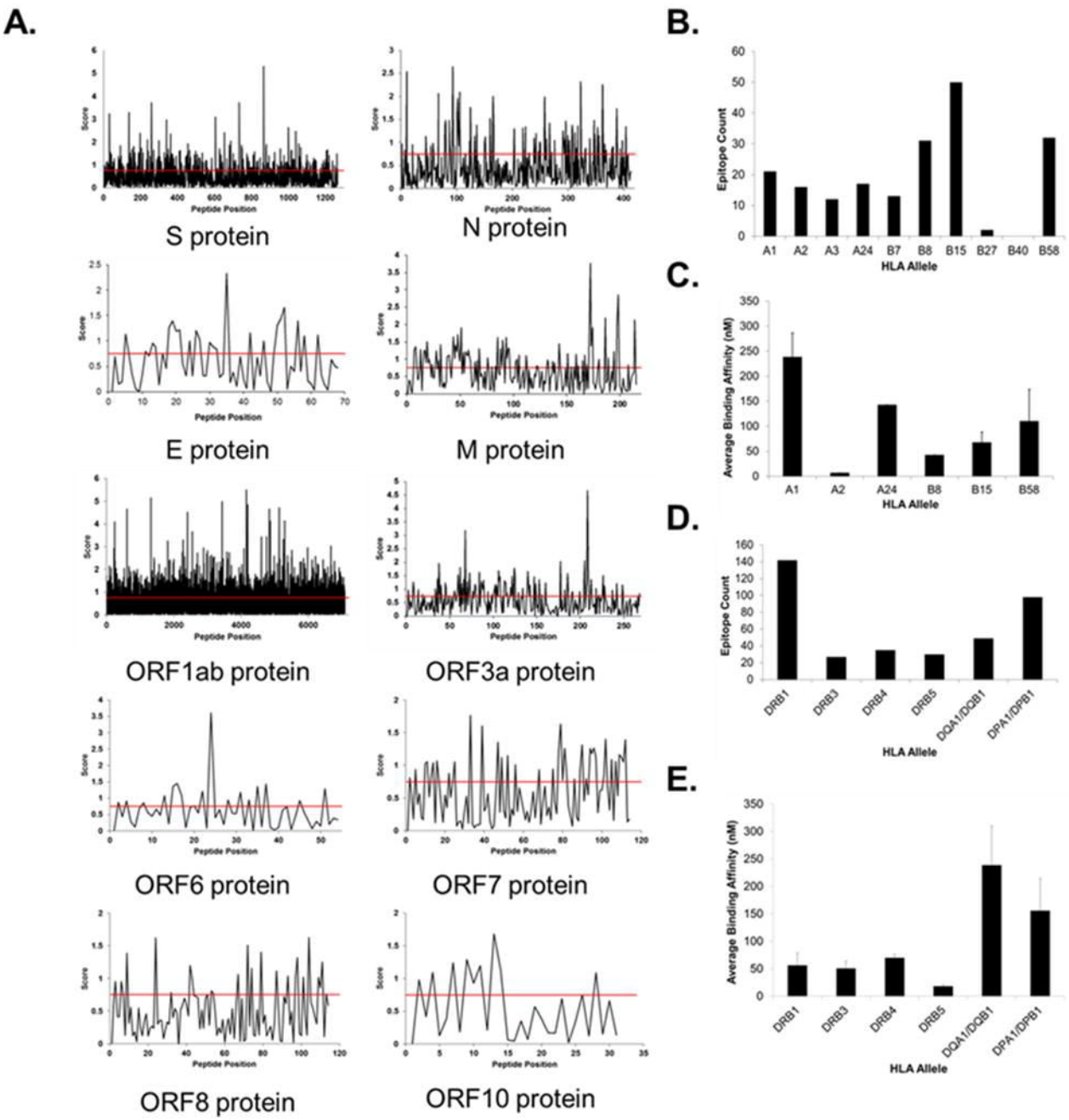
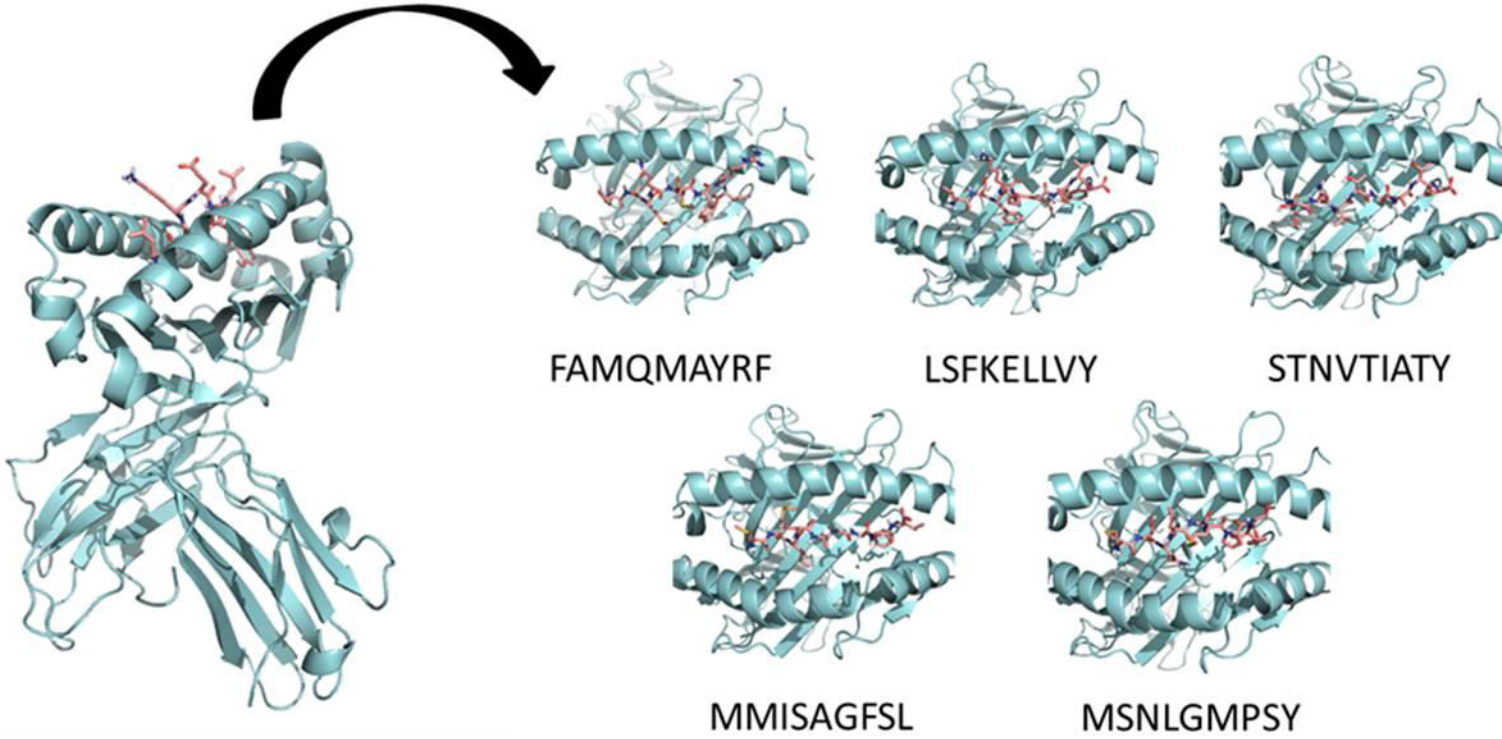
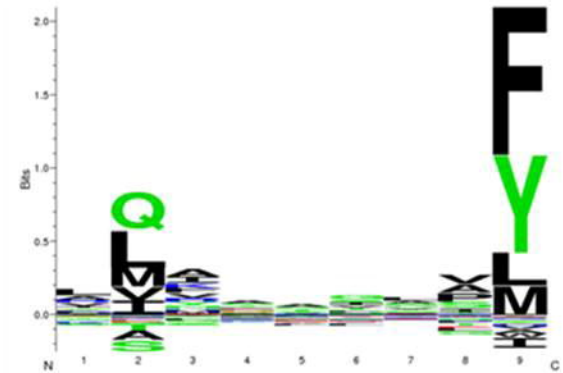

A.
S protein
N protein
E protein
M protein
ORF1ab protein
ORF3a protein
ORF6 protein
ORF7 protein
ORF8 protein
ORF10 protein
Score
Peptide Position
B.
Epitope Count
A1 A2 A3 A24 B7 B8 B15 B27 B40 B58
HLA Allele
C.
Average Binding Affinity (nM)
A1 A2 A24 B8 B15 B58
HLA Allele
D.
Epitope Count
DRB1 DRB3 DRB4 DRB5 DQA1/DQB1 DPA1/DPB1
HLA Allele
E.
Average Binding Affinity (nM)
DRB1 DRB3 DRB4 DRB5 DQA1/DQB1 DPA1/DPB1
HLA Allele

Figure 3.

A.



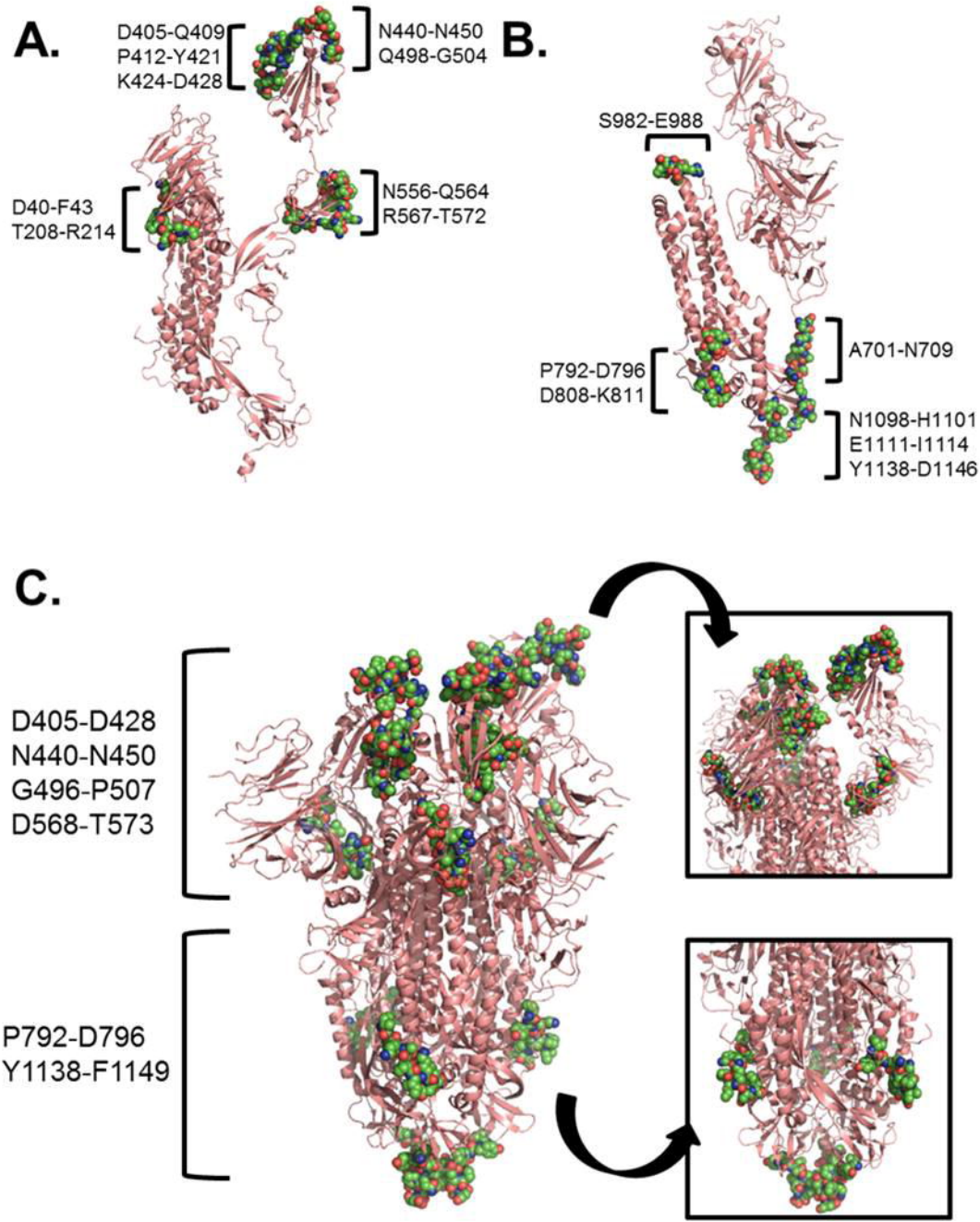
B.



C.

Peptide Sequence	Template Modeling Score	Interaction Similarity Score
FAMQMAYRF	0.992	202.0
LSFKELLVY	0.963	238.0
STNVTIATY	0.994	200.0
MMISAGFSL	0.984	225.0
MSNLGMPSY	0.963	234.0

Figure 4.


A.
D405-Q409
P412-Y421
K424-D428
N440-N450
Q498-G504
D40-F43
T208-R214
N556-Q564
R567-T572
B.
S982-E988
A701-N709
P792-D796
D808-K811
N1098-H1101
E1111-I1114
Y1138-D1146
C.
D405-D428
N440-N450
G496-P507
D568-T573
P792-D796
Y1138-F1149